

**Search for Direct Stop Production Using the Razor Variables with the CMS
Experiment at the CERN LHC**

BY

LUCIE GAUTHIER

B.Sc., Université Paris XI, 2007

M.Sc., Université Paris XI, 2009

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Chicago, 2015

Chicago, Illinois

Defense Committee :

Richard Cavanaugh, UIC Department of Physics/Fermilab, Chair and Advisor

Cecilia Gerber, Department of Physics

Wai Yee Keung, Department of Physics

Steve Nahn, Fermilab

Nikos Varelas, Department of Physics

ACKNOWLEDGMENTS

Writing acknowledgments feels like an embarrassing exercise. Generally speaking, I am afraid I am not good at explicitly saying thank you, I just try my best to give back. Besides, I am not sure I am proud enough of my work to name people who helped me. Was it really worth it? And, of course, would somebody feel forgotten? Given the size of the project, the list could be quite long...

Yet, I do feel grateful towards a certain number of persons. To begin with, I am grateful to whomever I may have met during these six years, anyone who supported me despite the varying moodiness induced by the stress and (self) pressure that have been the most faithful companions while carrying out this work.

I am particularly thankful to Roberto Rossin, for guiding me through the trigger world, and pushing me towards responsibilities not usually endorsed by graduate students (and who still taught me about the detector when he came back for a visit with his students!). Working within the trigger group has been highly stimulating and a rich experience from a human point of view. I would also like to thank the susy PAG conveners for their trust.

The razor analyses are not particularly easy to understand, and I appreciate that the razor group welcomed me while they were already at a quite advanced stage, with all my naive questions and doubts. Of course, I would like to thank my advisor, Richard Cavanaugh, for his patience and kindness, even when demotivation or irritation gained me.

ACKNOWLEDGMENTS (Continued)

I definitely enjoyed much more online operations activities, being at P5. I more often went there as a ‘parasite guide’ rather than a worker: yet, after hundreds of these visits, I remain admiring of all the work being done over there. Some thoughts as well go for all the visitors who had the curiosity to come and visit us.

TABLE OF CONTENTS

<u>CHAPTER</u>		<u>PAGE</u>
1	[BEYOND] THE STANDARD MODEL OF PARTICLE PHYSICS	1
1.1	The Standard Model of Particle Physics	1
1.1.1	Particle Physics?	1
1.1.2	Formalism	6
1.1.3	Known issues	11
1.2	Introducing supersymmetry	12
1.2.1	Of bosons and fermions	12
1.2.2	Supersymmetry: a solution to regularize the Higgs boson's mass	13
1.2.3	R-parity	14
1.2.4	Supersymmetry breaking and naturalness	15
1.2.5	Simplified models	17
2	THE LHC AND THE CMS DETECTOR	20
2.1	The LHC	20
2.2	The CMS detector	21
2.2.1	Silicon tracker	23
2.2.2	Calorimeters	24
2.2.2.1	Electromagnetic calorimeter	24
2.2.2.2	Preshower	25
2.2.2.3	Hadronic calorimeter	26
2.2.3	Energy absorption in the CMS detector	27
2.2.4	The magnet	28
2.2.5	Muon system	28
2.2.5.1	Drift tubes	29
2.2.5.2	Cathode strips chambers	32
2.2.5.3	Resistive plate chambers	34
2.3	Event reconstruction	34
2.3.1	Particle Flow, or Global Event Description, algorithm	35
2.3.2	Lepton identification and isolated track veto	37
2.3.2.1	Electron and muon identification	37
2.3.2.2	Isolated track veto	39
2.3.3	Jets	40
2.3.3.1	b-jets	42
2.3.4	Missing transverse momentum	43
2.4	Event simulation	44
2.4.1	Event generation	44
2.5	Detector simulation	45

TABLE OF CONTENTS (Continued)

<u>CHAPTER</u>		<u>PAGE</u>
3	TRIGGERING AT CMS	47
3.1	Setting the scene: collisions at the LHC, why do we need a trigger system	48
3.2	Data taking at CMS	49
3.2.1	Level 1	51
3.2.1.1	Muon system	52
3.2.1.2	Calorimetric system	53
3.2.1.3	Global trigger	53
3.2.2	Data acquisition system	54
3.2.3	High level trigger	55
3.2.4	The Grid, or WLCG: Worldwide LHC Computing Grid.	56
3.3	The trigger Studies Group: organizing the data taking for hundreds of analyses	58
3.3.1	Trigger menus	60
3.3.1.1	Foreword: data organization	60
3.3.1.2	L1 menu	60
3.3.1.3	HLT menu	61
3.3.1.4	Menu development: rate & timing	65
3.3.1.4.1	Rate evaluation	65
3.3.1.4.2	Controlling the timing	66
3.3.2	Trigger paths for supersymmetry searches	69
3.3.2.1	Addressing the increasing number of pile-up interactions at HLT	72
3.3.2.2	Towards better jets	78
3.3.2.3	Addressing pile-up interactions at L1	81
3.3.3	Selected tales of trigger activities	83
4	SEARCH FOR THIRD GENERATION SPARTICLES WITH THE RAZOR VARIABLES	88
4.1	Introduction	88
4.2	The razor variables	89
4.2.1	Resonances, supersymmetry event	89
4.2.1.1	<i>Bump</i> search	89
4.2.1.2	Supersymmetry events: an underconstrained system and no bump	92
4.2.2	Introducing the razor variables	93
4.2.2.1	In the squark's $S_{i=1,2}$ rest frames	93
4.2.2.2	From the squarks' rest frames to the razor frame	94
4.3	Analysis	98
4.3.1	A shape analysis	99
4.3.2	Splitting into boxes and analysis selection	102
4.3.2.1	Standard Model backgrounds	103
4.3.2.2	Leptonic boxes analysis selection	105

TABLE OF CONTENTS (Continued)

<u>CHAPTER</u>		<u>PAGE</u>
4.3.2.3	Hadronic box analysis selection	108
4.3.2.3.1	Rejection of the QCD background against ttbar events using a BDT	109
4.3.2.3.2	Top tagging variables	112
4.3.2.3.3	Quark-gluon discrimination variables	116
4.3.2.4	Data samples	119
4.3.2.5	Comparison of data versus simulation	119
4.3.3	Establishing the background modeling	120
4.3.3.1	Leptonic boxes	120
4.3.3.2	Hadronic boxes	122
4.3.4	Probing our sensitivity to signal: expected limits and signal injection tests	126
4.3.4.1	Limit setting - LHC CLs	128
4.3.4.1.1	Building a test statistics, CL_s definition	128
4.3.4.1.2	Nuisance parameters treatment	130
4.3.4.1.3	Observed and expected limits	135
4.3.4.1.4	Limits from our background only fits to simulated events . . .	135
4.3.4.2	Signal injection tests	137
4.3.4.2.1	BJetHP box	137
4.3.4.2.2	Leptonic boxes	141
4.3.5	Data fits and results	145
4.3.5.1	Limits on T2tt with our samples and analysis	147
4.3.5.2	Comparison with existing results	152
5	CONCLUSIONS	155
5.1	Summary	155
5.2	A human adventure - final word	155
	APPENDICES	158
	Appendix A	159
	CITED LITERATURE	193
	VITA	198

LIST OF TABLES

<u>TABLE</u>		<u>PAGE</u>
I	SIMULATED (MC) SAMPLES, CROSS SECTIONS, NUMBER OF EVENTS GENERATED, RESULTING WEIGHT TO MATCH 19.3FB^{-1} OF DATA.	106
II	MC YIELDS FOR BACKGROUND PROCESSES, ELE BOX, AFTER EACH CUT. THE NUMBER OF EVENTS CORRESPONDS TO 19.3FB^{-1}	108
III	MC YIELDS FOR BACKGROUND PROCESSES, MU BOX, AFTER EACH CUT. THE NUMBER OF EVENTS CORRESPONDS TO 19.3FB^{-1}	109
IV	MC YIELDS FOR BACKGROUND PROCESSES, BJETHP BOX, AFTER EACH CUT. AGAIN, THE FINAL KINEMATIC CUTS REMOVE EVENTS WITH SMALL MISSING MOMENTUM (WJETS, SINGLE TOP), AND DI-TOP EVENTS WITH LOW M_R VALUES. THE FINAL BDT CUT REJECTS FURTHER THE QCD BACKGROUND. THE NUMBER OF EVENTS CORRESPONDS TO 19.3FB^{-1}	118
V	FREE PARAMETERS VALUES AFTER FITTING THE SIMULATED TTBAR EVENTS SAMPLE, ELE BOX. TTJ1B AND TTJ2B REFER TO THE COMPONENT: ONE FOR ONE B-TAGGED JET EVENTS, ONE FOR TWO OR MORE B-TAGGED JETS EVENTS.	126
VI	FREE PARAMETERS VALUES AFTER FITTING THE SIMULATED TTBAR EVENTS SAMPLE, MU BOX. TTJ1B AND TTJ2B REFER TO THE COMPONENT: ONE FOR ONE B-TAGGED JET EVENTS, ONE FOR TWO OR MORE B-TAGGED JETS EVENTS.	128
VII	FREE PARAMETERS VALUES AFTER FITTING THE SIMULATED MC COCKTAIL EVENTS SAMPLE, BJETHP BOX. TTJ1B AND TTJ2B REFER TO THE COMPONENT: ONE FOR ONE B-TAGGED JET EVENTS, ONE FOR TWO OR MORE B-TAGGED JETS EVENTS.	129

LIST OF TABLES (Continued)

<u>TABLE</u>		<u>PAGE</u>
VIII	SIGNAL INJECTION TESTS, LEPTONIC BOXES. AFTER BUILDING A BACKGROUND PLUS SIGNAL, WITH A TRIAL CROSS-SECTION, SAMPLE, WE FIT IT WITH THE SIGNAL PLUS BACKGROUND PROBABILITY DENSITY FUNCTION, ALLOWING THE CROSS SECTION TO FLOAT. THE FIT RESULT RETURNS BACK THE INJECTION CROSS-SECTION QUITE WELL FOR VALUES FOR WHICH WE HAVE SENSITIVITY, THAT IS, FOR WHICH WE ARE ABLE TO DISTINGUISH SIGNAL FROM BACKGROUND.	145

LIST OF FIGURES

<u>FIGURE</u>		<u>PAGE</u>
1	Mendeleev table	2
2	A Standard Model chart	5
3	Higgs mass one loop corrections, illustrated for top and stops	17
4	Supersymmetric cross sections at the LHC, as a function of sparticles' masses	18
5	T2tt simplified models	19
6	The CMS detector	22
7	An electromagnetic shower	25
8	Map of the magnetic field magnitude in the CMS detector	29
9	Location of muon chambers in CMS	30
10	Muon system: drift tubes	31
11	Muon system: CSC	33
12	Example of a Feynman diagram, representing a full process, from production to final state.	36
13	From parton to jets	41
14	Hadronization models	46
15	First step into Trigger World	47
16	Old fashioned data selection	49
17	CMS Data AcQuisition (DAQ) system	50
18	Overview of the Level 1 trigger	51

LIST OF FIGURES (Continued)

<u>FIGURE</u>		<u>PAGE</u>
19	A glimpse of the underlying organisation for trigger activities at CMS (Run I)	59
20	L1 menu as visualized in WBM (<i>Web Based Monitoring tool, which gathers information about online operations</i>)	62
21	Luminosity decrease during an LHC fill	63
22	Visualizing an HLT menu using the database browser	64
23	Structure of an HLT path	68
24	A possible supersymmetry event, decay chain	70
25	Sharing of the bandwidth requested by supersymmetric searches in 2011	71
26	Effect of pile-up interactions on jets -1	73
27	Effect of pile-up interactions on jets -2	74
28	A pathologic trigger path cross section versus luminosity	75
29	Effect on pile-up corrections on the cross section of paths relying on H_T and α_T	77
30	A <i>turn-on curve</i> or efficiency plot for a path cutting at 400GeV on online H_T , HLT_HT400	78
31	Turn-on curves for different set of online jets requiring an online $p_t > 30\text{GeV}$, as a function of offline jet p_t , obtained on $Z \rightarrow \mu\mu$ events.	81
32	A L1 jet in the $\eta - \phi$ plane	82
33	Cross section of L1 seeds cutting on H_T , as a function of luminosity	83
34	Efficiency (or turn-on) curves for L1_HTT150, for different GCT settings, in high and low pile-up conditions	84
35	Selected tales - trigger world	86
36	Selected tales - trigger world	87

LIST OF FIGURES (Continued)

<u>FIGURE</u>		<u>PAGE</u>
37	The T2tt simplified model	89
38	Resonance production and decay diagram	90
39	Di-muon invariant mass spectrum at the LHC	91
40	A <i>canonical</i> supersymmetry event	93
41	Frames used in the M_R derivation	95
42	Distribution of M_R , for different stop masses	96
43	Kinematic cuts applied on the razor variables for the inclusive and mul- tijet razor analyses.	99
44	Distribution of the number of events per toy in a representative R^2 , M_R cell	101
45	Building <i>French Flag</i> plots	102
46	M_R distribution for background and signal events, before selection . . .	107
47	Efficiency of the trigger path HLT_QuadJet50, used for the hadronic BJet box	110
48	Sketch of a (Boosted) Decision Tree	111
49	BDT performance	112
50	Deriving top tagging variables: casting what we see into what we expect	113
51	Helicity angle	114
52	Top-tagging variables	115
53	Quark - gluon discrimination variables	117
54	Comparison of data with simulated events for R^2 and M_R distributions, leptonic boxes	121
55	Comparison of data with simulated events for R^2 and M_R distributions, hadronic BJetHP box	122

LIST OF FIGURES (Continued)

<u>FIGURE</u>		<u>PAGE</u>
56	Efficiency map after the analysis selection, for the T2tt simplified model	123
57	Background modeling test using ttbar+jets for the Mu box	124
58	Background modeling test using ttbar+jets for the Ele box	125
59	Background modeling test using simulated Standard Model processes for the BJetHP box	127
60	Computing CL_s	131
61	Expected (black dotted with uncertainty bands) and observed (red) limits using simulated data	136
62	Expected (black dotted with uncertainty bands) and observed (red) limits using simulated data	138
63	T2tt signal used for signal injection tests of the BJetHP box	139
64	Signal injection test for BJetHP box - Background only fit to a background+signal sample	140
65	Signal injection test - Finding back the injection cross section	141
66	Signal injection test - Ele box	143
67	Signal injection test - Mu box	144
68	Fitting our data sample after the BJetHP box selection with the background only model	146
69	Fitting our data sample after the Ele box selection with the background only model	148
70	Fitting our data sample after the Mu box selection with the background only model	149
71	Exclusion limits for mLSP=25GeV, T2tt, BJetHP box	150
72	Exclusion limits for mLSP=25GeV, T2tt. Mu and Ele boxes are shown separately and combined	151

LIST OF FIGURES (Continued)

<u>FIGURE</u>		<u>PAGE</u>
73	Exclusion limits for $m_{\text{LSP}}=25\text{GeV}$, T2tt, BJetHP, Mu and Ele boxes combined	152
74	Combined Exclusion Limits in the 2D stop mass - LSP mass plane for all boxes	153
75	Existing CMS and ATLAS results with 8TeV data	154
76	Refuge du Couvercle	156
77	Recycling & developing: the LHC accelerator complex	161
78	Where is the LHC?	164
79	A cut view of the CMS detector (Image Michael Hoch/Maximilien Brice)	166
80	A model of the CMS detector.	167
81	A schematic view of the CMS detector.	168
82	Schematic cut view of CMS	177

LIST OF ABBREVIATIONS

CMS	Compact Muon Solenoid
CSC	Cathode Strip Chamber
CSV	Combined Secondary Vertices
DAQ	Data Acquisition
DPG	Detector Physics Group
DT	Drift Tube
ECAL	Electromagnetic CALorimeter
EG	EGamma: electron or photon-like object
GCT	Global Calorimeter Trigger
GMT	Global Muon Trigger
GT	Global Trigger
HCAL	Hadronic CALorimeter
LEP	Large Electron Positron Collider
LHC	Large Hadron Collider
MC	Monte Carlo
MSSM	Minimal SuperSymmetric Model
PAG	Physics AnalysisGroup

LIST OF ABBREVIATIONS (Continued)

POG	Physics ObjectGroup
QCD	Quantum ChromoDynamics
QED	Quantum ElectroDynamics
RPC	Resistive Plate Chamber
SM	Standard Model
SUSY	SUper SYmmetry
TPG	Trigger Primitive Generator
TSG	Trigger Studies Group
TTC	Timing Trigger and Control
TTS	Trigger Throttling System

SUMMARY

After decades of development, the Large Hadron Collider first entered into operations in September 2008. During its first run, which ended in February 2013, the 27km long circular accelerator collided protons at a center of mass energy up to 8TeV. This unprecedented energy has been desired to further unveil physics phenomena at their most elementary scale, hopefully solving some of the known issues of the current Standard Model of Particle Physics. Indeed, while this theory encountered many successes during the twentieth century, questions remain. The discovery of the long sought Higgs boson in 2012 is probably the main highlight of the first LHC run. Another key topic lies in the search for new physics beyond the Standard Model, such as supersymmetry.

The present work addresses a search for supersymmetric partners of the Standard Model top quarks, using the razor variables. The razor variables have been designed so as to exhibit a characteristic shape for such signals, to be distinguished from the falling spectrum they display for the Standard Model processes. The resulting shape analysis is developed in chapter 4. Stop masses are excluded from 200GeV up to 700GeV for an LSP mass of 25GeV.

The huge amount of collisions, several millions per second, provided by the LHC requires a sound *sorting* machinery to quickly select the interesting events, corresponding to the rare processes we are interested in out of the non so interesting ones. The third chapter of this thesis is dedicated to the *trigger* system of the CMS experiment.

SUMMARY (Continued)

The CMS experiment itself, one of the *eyes* of the LHC, is detailed in chapter 2, from a physicist point of view, and by a CMS guide in the appendix. As part of a public research site, the outreach activities around the CMS detector are indeed well developed and deserve some advertizing.

CHAPTER 1

[BEYOND] THE STANDARD MODEL OF PARTICLE PHYSICS

1.1 The Standard Model of Particle Physics

1.1.1 Particle Physics?

What is physics? In a nutshell, physics consists of describing the world around us, in terms how things interact, putting aside everything pertaining to life. Yet, there is a vast variety of fields within physics, from semiconductor physics to nanophysics, cosmology, medical physics, fluid mechanics: particle physics is just one of these many specialities.

Once upon a time...

Greek philosophers developed two main views about what the world could be made of. One line of thought considered that everything arose from four elements: water, air, fire, earth. Another view favored an *atomic* model, which described the tiniest matter components as *unbreakable bricks*. This ancient concept, forgotten over centuries, jumped into modernity in the nineteenth century when Mendeleev proposed a classification for these building blocks (Figure 1): there were several kinds of bricks, or elements.

However, these atoms soon appeared to be way too numerous: we had dozens of them. Too many to be the smallest constituents of matter, too many to be *elementary*. They had to be composite, made of more fundamental components. Thomson's experiments, at the end of the nineteenth century, concluded several decades of research involving electricity and officially

ОПЫТЪ СИСТЕМЫ ЭЛЕМЕНТОВЪ.

ОСНОВАННОЙ НА ИХЪ АТОМНОМЪ ВѢСѢ И ХИМИЧЕСКОМЪ СХОДСТВѢ.

			Ti = 50	Zr = 90	? = 180.
			V = 51	Nb = 94	Ta = 182.
			Cr = 52	Mo = 96	W = 186.
			Mn = 55	Rh = 104,4	Pt = 197,1.
			Fe = 56	Ru = 104,4	Ir = 198.
			Ni = 59	Pd = 106,8	Os = 199.
			Cu = 63,4	Ag = 108	Hg = 200.
H = 1	Be = 9,4	Mg = 24	Zn = 65,2	Cd = 112	
	B = 11	Al = 27,1	? = 68	Ur = 116	Au = 197?
	C = 12	Si = 28	? = 70	Sn = 118	
	N = 14	P = 31	As = 75	Sb = 122	Bi = 210?
	O = 16	S = 32	Se = 79,4	Te = 128?	
	F = 19	Cl = 35,5	Br = 80	I = 127	
Li = 7	Na = 23	K = 39	Rb = 85,4	Cs = 133	Tl = 204.
		Ca = 40	Sr = 87,6	Ba = 137	Pb = 207.
			? = 45	Ce = 92	
			?Er = 56	La = 94	
			?Yt = 60	Di = 95	
			?In = 75,6	Th = 118?	

Д. Менделѣевъ

Figure 1. **Mendeleev table** - During the nineteenth century, many sorts of atoms were identified. Mendeleev proposed a classification for them, based on their chemical properties. These properties are actually the result of the inner structure of atoms. Figure uploaded on en.wikipedia.org, reproducing the table as published by Mendeleev in 1869. Public domain.

acknowledged the first elementary particle – the electron. Rutherford, in 1909, shot α particles towards a sheet of gold and found evidence for the presence of a nucleus within the atom. After the discovery of the (positively charged) nucleus, Rutherford also showed that all nuclei derived from the simplest of them, the hydrogen’s nucleus, which he named *proton*. Neutrons were discovered later in the 1930s.

In the 1950s, many particles had been discovered, often through cosmic rays studies, and physicists were running out of greek letters to name them. Again, these particles were far too numerous to be elementary. This was to be solved in the 1960’s, when quarks were first theorized and then discovered as components of protons, neutrons, and many of the recently discovered particles¹.

We would have been happy with describing just the world around us. Yet, particles physics experiments brought us more than what we wanted. The muon was discovered in the 1930s, a particle very similar to the electron but two hundred times heavier. The discovery of the up and down quarks, components of protons and neutrons, were later followed by discoveries of the charm, strange, and bottom quarks in the 1970s, and finally the top quark was discovered in the 1990s². The tau lepton, an even heavier electron (1.8 GeV), showed up in the 1970s as well. For the ordinary matter, up and down quarks, electrons with their neutrino, are enough. Nevertheless, in particle physics experiments, we see these two additional families, very similar

¹The generic name of *hadrons* is given to all particles made of quarks. Two sub-categories make a further distinction: *mesons* are made of two quarks, while *baryons* carry three quarks.

²The top quark was theoretically predicted before being observed at the Tevatron in 1994(1).

to the first one, except for their weight. As a result, the particles of the heavier two families are unstable and decay into the particles of the first family.

Describing the matter content is not enough, we want to know how these particles interact. In particle physics, interactions, or forces, are carried by other particles, the force carriers:

- the photon, which carries the electromagnetic force;
- the gluons, which carry the strong interaction between quarks;
- the W and Z bosons, which carry the weak interaction;
- the hypothetical graviton, which would carry gravitational interactions between massive particles.¹

We could write that, just like we, human beings, exchange words, particles exchange particles to talk to each other. And, just like we may be using different (sets of) words corresponding to different languages, particles can interact through four forces. However, not all particles speak all languages: leptons are only sensitive to the weak and gravitational interaction and, if they are charged, to the electromagnetic force as well. Even the force carriers themselves may be sensitive to some of the forces: W bosons are charged, hence they may feel an electromagnetic field².

¹The graviton has not yet been observed and neither has it properly been included in the theory.

²What is a field? In physics, we could say that it is a quantity defined in a given region of space. We could consider a temperature field: the values the temperature takes over a region of space. An electric field is something that we would notice because charged particles do not follow straight lines in its presence. We observe the field thanks to the force it induces. From a mathematical point of view, a field is a function, possibly vectorial.

We now have the table of the known Standard Model of Particle Physics particles, summarized in Figure 2. Figure 2 shows that the set of *matter* particles come in three families, ordered by mass. The quarks, shown in yellow, are sensitive to all interactions, while the charged leptons, shown in green, charged leptons are sensitive to all interactions except the strong one. Neutrinos, also in green, are only sensitive to the weak interaction and gravity. In orange, the fourth column, we have the force carriers. We will come back to the Higgs boson in the next section.

mass →	$\approx 2.3 \text{ MeV}/c^2$	$\approx 1.275 \text{ GeV}/c^2$	$\approx 173.07 \text{ GeV}/c^2$	0	$\approx 126 \text{ GeV}/c^2$
charge →	$2/3$	$2/3$	$2/3$	0	0
spin →	$1/2$	$1/2$	$1/2$	1	0
	u up	c charm	t top	g gluon	H Higgs boson
QUARKS	$\approx 4.8 \text{ MeV}/c^2$	$\approx 95 \text{ MeV}/c^2$	$\approx 4.18 \text{ GeV}/c^2$	0	
	$-1/3$	$-1/3$	$-1/3$	0	
	$1/2$	$1/2$	$1/2$	1	
	d down	s strange	b bottom	γ photon	
	$0.511 \text{ MeV}/c^2$	$105.7 \text{ MeV}/c^2$	$1.777 \text{ GeV}/c^2$	$91.2 \text{ GeV}/c^2$	
	-1	-1	-1	0	
	$1/2$	$1/2$	$1/2$	1	
	e electron	μ muon	τ tau	Z Z boson	
LEPTONS	$< 2.2 \text{ eV}/c^2$	$< 0.17 \text{ MeV}/c^2$	$< 15.5 \text{ MeV}/c^2$	$80.4 \text{ GeV}/c^2$	
	0	0	0	± 1	
	$1/2$	$1/2$	$1/2$	1	
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	W W boson	
					GAUGE BOSONS

Figure 2. **A Standard Model chart** - The set of *matter* particles comes in three families, ordered by mass (first three columns, starting from the left). In yellow, quarks are sensitive to all interactions. In green, charged leptons are sensitive to all interactions except the strong one. Neutrinos, also in green, are only sensitive to the weak interaction and gravity. In orange, fourth column, we have the force carriers. We will come back to the Higgs boson later.

1.1.2 Formalism

Physics is not just words, physics is both mathematics and words in order to predict and quantify systems' evolutions. Let us detail a bit the mathematics behind Figure 2 (see, for instance, (2)).

Particle physics was born in the twentieth century and benefited from the mathematical development from the nineteenth: an elegant formulation of physics laws stems from minimizing the *action*. In other words, Nature always chooses the easiest way to go from spatial point A to spatial point B between a time t_0 and a later time t_1 . The *action* is defined as:

$$S = \int_{t_0}^{t_1} L(q, \dot{q}, t) dt. \quad (1.1)$$

We integrate the *Lagrangian* over time, where the Lagrangian is defined by the kinetic and potential energies that describe our system. The quantities, q and $\dot{q}$ are generalized coordinates that parameterize the phase space we use to study our system. Minimizing the action gives us the equations of motion, or path, followed by the system described by the Lagrangian in the corresponding phase space. This procedure is quite generic, from which the well-known Euler-Lagrange equations follow:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0. \quad (1.2)$$

For example, in classical mechanics, one can use these equations to derive Newton's laws of motion.

In particle physics, we start with building the Lagrangian for a free particle (that is, we want to predict a particle's motion when no force affects it):

$$L_{free} = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi. \quad (1.3)$$

ψ is a mathematical view of our fermionic¹ particles: this four component Dirac spinor contains a particle along with its anti-particle. γ^μ are the Dirac matrices, required for a relativistic quantum mechanics theory, and m is the mass of the particle described by ψ . By definition, the repetition of the μ index indicates that we sum over all possible values for μ , which correspond to the four space-time coordinates. By minimizing the action of this Lagrangian, one obtains the well-known Newtonian rule: a free particle follows a straight and uniform path.

To make things more interesting, we introduce forces, through their fields or force carriers. We begin with the photon. To do so, we require *local $U(1)$ gauge invariance*. That is, our Lagrangian has to be invariant under a local $U(1)$ gauge transformation²:

$$\psi \rightarrow e^{i\alpha(x)} \psi. \quad (1.4)$$

¹Matter particles carry an intrinsic angular momentum, called spin, in half integer quantized units and are known as fermions. See 1.2.1.

²In the 19th century, mathematicians, and later physicists, became quite interested in symmetries. In physics, a symmetry is of particular importance since it corresponds to a conserved quantity, as proved by Noether's theorem (3). The local $U(1)$ gauge invariance is the symmetry ensuring conservation of the electric charge.

At this point, we need to introduce the *covariant derivative*:

$$D_\mu = \partial_\mu + ieA_\mu(x); \quad (1.5)$$

where the field A_μ transforms as:

$$A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu\alpha. \quad (1.6)$$

Now, to preserve invariance, our *quantum electrodynamics* (QED) Lagrangian takes the form:

$$L_{QED} = \bar{\psi} (\gamma^\mu (i\partial_\mu + eA_\mu) - m) \psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (1.7)$$

The vector field A_μ introduces the mathematical face of the photon, along with the electric charge e . The field strength tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ corresponds to the kinetic energy term for the electromagnetic field, allowing the photon to propagate.

Then, quantum electrodynamics and weak interactions combine within the electroweak force (EWK). Requiring the Lagrangian to be invariant under an $SU(2) \times U(1)$ transformation yields:

$$\begin{aligned} L_{EWK} = & \bar{\psi}_L \gamma^\mu \left(i\partial_\mu - \frac{g'}{2} Y B_\mu - \frac{g}{2} \boldsymbol{\tau} \cdot \mathbf{W}_\mu \right) \psi_L + \\ & \bar{\psi}_R \gamma^\mu \left(i\partial_\mu - \frac{g'}{2} Y B_\mu \right) \psi_R - \frac{1}{4} \mathbf{W}_{\mu\nu} \mathbf{W}^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}. \end{aligned} \quad (1.8)$$

A key feature arises in this Lagrangian: *chirality*. For weak interactions, *left-handed* and *right-handed* states appear, as denoted by the subscripts L and R :

$$\psi_L = \frac{1}{2} (1 - \gamma^5) \psi. \quad (1.9)$$

$$\psi_R = \frac{1}{2} (1 + \gamma^5) \psi. \quad (1.10)$$

Left-handed states are charged under $SU(2)$ and so contain the following doublets: (u_L, d_L) , (c_L, s_L) , (t_L, b_L) , $(e_L, \nu_{e,L})$, $(\mu, \nu_{\mu,L})$, $(\tau, \nu_{\tau,L})$. They carry a $1/2$ weak isospin and transform under $SU(2)$ in the following way: one of the elements of the doublet can be turned into the other by an $SU(2)$ transformation through the emission of a gauge boson, W^+ or W^- . Right-handed states are not charged under $SU(2)$ and so are singlets, that is their weak isospin is equal to zero and they are not affected by $SU(2)$ transformations.

The gauge fields B and $\mathbf{W}$ are not the one we actually observe, because the electroweak symmetry is broken. Indeed, with the electroweak Lagrangian given by Equation 1.8, we cannot explain particles' masses, because the different treatment of left-handed and right-handed states now prohibits mass terms such as the one we had for Quantum Electrodynamics. A solution to this issue consists of introducing a new complex scalar field ϕ , whose contribution to the Lagrangian is:

$$L_{\text{Higgs}} = \left| \left(i\partial_\mu - \frac{g'}{2} Y B_\mu - \frac{g}{2} \boldsymbol{\tau} \mathbf{W}_\mu \right) \phi \right|^2 - V(\phi). \quad (1.11)$$

This term leads to a spontaneous symmetry breaking, meaning that the Lagrangian remains invariant under $SU(2) \times U(1)$, but the vacuum state is not. Upon breaking the electroweak symmetry, we obtain the physical W and Z bosons, and the photon. Thanks to the first modulus, the W and Z bosons become massive. $V(\phi)$ is the Higgs potential:

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda^2 |\phi^\dagger \phi|^2. \quad (1.12)$$

We can now add Yukawa interaction (4) terms that allow fermions to get massive through their interaction with the Higgs field:

$$L_{\text{Yukawa}} = -\lambda_f [\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \bar{\phi} \psi_L]. \quad (1.13)$$

Thus, the Higgs field, and therefore its corresponding particle the Higgs boson, is of paramount importance: without it, all the particles of the Standard Model are massless, which is quite in disagreement with experiments. It was a key success of the first LHC run to report its observation, more than forty years after the theoretical initial proposal ((5), (6), (7) and (8)).

Finally, to account for the strong interaction (9), we require the $SU(3)$ gauge invariance (corresponding to the *color charge* conservation¹), yielding the Quantum ChromoDynamics term of the Standard Model Lagrangian:

$$L_{\text{QCD}} = -g (\bar{\psi} \gamma^\mu \mathbf{T}_a \psi) G_\mu^a - \frac{1}{4} \mathbf{G}_{\mu\nu}^a \mathbf{G}^{\mu\nu,a} \quad (1.14)$$

We obtain the gluons, gauge bosons of the strong interaction, through G . The eight gluons allow for *color* charge exchange between colored quarks, but also between themselves. Hence, unlike photons in electrodynamics, gluons in QCD may interact with each others. Because gluons do not carry electric charge, they do not interact with photons. Since the $SU(3)$ symmetry is an exact symmetry, and not broken, gluons are massless.

1.1.3 Known issues

Despite all the successes encountered over the past half century, the Standard Model of Particle Physics still faces thorny issues. For instance, neutrinos do not have a mass in the Standard Model. Yet, neutrino oscillations (they can change flavor) require them to be massive (10).

¹Of course, the color charge does not mean that strongly interacting particles are actually colored like objects in our daily lives. Yet, color at the quantum level comes in three flavor, and the analogy with light color yields a convenient way to describe color neutral, or white, observed states. Indeed, quarks and gluons, which carry color charges, are never observed on their own, they always combine within color neutral hadrons. If we consider that the three flavors are red, blue, and green, a color neutral state can be obtained with a combination of a color charge and the corresponding anti-color charge, or by mixing red, blue and green.

We also know that the Standard Model does not describe all the matter content of the Universe. In particular, it can not account for *dark matter*. Indeed, while cosmological studies indicate that about 25% of the Universe content should be dark matter, particle physics does not yet provide us with the type of particles necessary to describe it (11).

Another embarrassing aspect of the current Standard Model of particle physics lies in the divergence of corrections to the Higgs boson mass. Indeed, while the Higgs boson allows other particles to have mass, its mass itself depends on the masses of these other particles. Unfortunately, when computing the corrections to the Higgs mass, highly divergent infinities appear.

1.2 Introducing supersymmetry

1.2.1 Of bosons and fermions

Fermionic states can be occupied by one, and only one, fermion. On the contrary, there is always room for more bosons in bosonic states. Thanks to the former statement, the periodic chart of atoms hold: electrons have to be arranged in several *layers* because they are fermions, thus they cannot all rest in the lowest energy state. It is because photons are bosons that we can make lasers, accumulating photons in the very same state.

Mathematically, bosons carry an integer spin while fermions carry a half integer spin. In terms of operators, the spin is similar to a rotation...but particles do not rotate on themselves. It can thus be thought of as a property that particles carry around, like, say, an umbrella. For instance, electrons with a spin $+1/2$ carry an umbrella pointing upwards, while electrons with a spin $-1/2$ carry an umbrella pointing downwards. At least, it shouldn't be too hard to remember that an electron does not actually carry an umbrella, whereas it might be trickier

to keep in mind that spin is like a rotation which is not rotating¹. The practical point about the spin is that it comes in to explain how many particles we can put in a given state: one or plenty.

1.2.2 Supersymmetry: a solution to regularize the Higgs boson's mass

In the Standard Model of particles physics, the mass of the Higgs boson receives corrections proportional to a high energy cutoff, Λ_{UV}^2 which could be significantly bigger than the Higgs boson mass, experimentally observed at 126 GeV. In other words, these corrections may be highly divergent, which is not acceptable for a realistic physics quantity. Interestingly, while bosonic particles terms contribute with a plus sign, fermions contribute with a minus sign:

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{UV}^2 + \frac{\lambda_S^2}{16\pi^2}\Lambda_{UV}^2 \quad (1.15)$$

While the coupling constants λ do not allow for cancellation between known SM particles, we could imagine that for each standard model fermion, there exist two new bosons, and vice-versa, so that the terms may cancel each other. This is the basic principle of supersymmetry ((12), (13)). Each new particle is called super partner of its standard model counterpart and follows the symmetry rule:

$$Q|\text{Boson}\rangle = |\text{Fermion}\rangle \quad (1.16)$$

¹I always keep in mind this drawing, showing electrons carrying umbrellas pointing downwards or upwards, in *Alice in Quantumland*, by Robert Gilmore.

$$Q|\text{Fermion}\rangle = |\text{Boson}\rangle. \quad (1.17)$$

The names of bosonic superpartners (assumed to be scalars) to fermionic particles are built by prepending an *s* to the name of the Standard Model particle: we obtain *squarks* (*stop*, *sbottom*, *etc...*) and *sleptons* (*selectron*, *stau*, *etc...*). The names of the fermionic superpartners to the Standard Model gauge bosons and Higgs particles are formed by using the suffix *-ino*: *Bino*, *Winos*, *Zino*, *gluinos*, *Higgsinos*. Supersymmetric particles (sparticles) differ from their standard model partner only by their spin, with a 1/2 integer difference.

To represent these particles with mathematical objects, we build *supermultiplets*. Within a supermultiplet, superpartners must have the same mass, electric charge, weak isospin and color degrees of freedom. *Matter*, or *chiral* supermultiplets contains the standard model fermions and their scalar superpartners. The Standard Model vector gauge bosons are arranged together with their fermionic superpartners within *gauge* supermultiplets. To preserve gauge anomaly cancellations, but also to allow the proper Yukawa couplings that give mass to up and down types quarks, we need two chiral supermultiplets for the Higgs sector. The Standard Model Higgs boson appears as a linear combination of two new fields, the neutral parts of these up and down Higgs supermultiplets.

1.2.3 R-parity

A supersymmetric Lagrangian can accommodate terms which, although they preserve gauge invariance, we usually chose not to allow. These terms would violate lepton or baryon number conservation, and, as a consequence would allow proton decay, unless the coupling constants

associated to these extra terms are extremely small. Rather than making this quite weird assumption, we define R-parity:

$$P_R = (-1)^{3(B-L)+2s} \quad (1.18)$$

and require an additional symmetry principle such that the products of R-parities for all particles involved in a given interaction equals +1. Standard Model particles have an R-parity equal to +1, while sparticles have an R-parity equal to -1. Besides preserving baryon and lepton numbers conservation, remarkable sides effects are:

- sparticles are produced in pairs, and vertices involving sparticles always contain an even number of them;
- the lightest supersymmetric particle (LSP) is stable;
- in collider experiments, supersymmetric decay chains always end with a state containing an odd number of LSP(s).

The Lightest Supersymmetric Particle (LSP), is of special importance: if it is electrically neutral, this weakly interacting particle provides an interesting candidate for dark matter since it must be massive. These will escape the detector unseen, yielding a key feature of supersymmetric events: missing energy (see 2.3.4).

1.2.4 Supersymmetry breaking and naturalness

Introducing supersymmetry, all the terms in Equation 1.15 should cancel thanks to the superpartners. Yet, since we have not observed any super partner so far, supersymmetry has to be a broken symmetry, with sparticles heavier than their Standard Model version.

Just like the Z and photon, which arise as mixture of W and B fields in the electroweak symmetry breaking process, mass eigenstates of a supersymmetric theory may be different from the sparticles described in 1.2.2. The supersymmetry breaking terms affect gauginos mass eigenstates: we expect four neutralinos and two charginos, mixing with the original gauginos. It also affects sfermions. In practice, the first and second families are barely modified, due to their relatively small Yukawa interactions. Hence, because the mixing is proportional to the Standard Model fermion masses, it is much more important for the third generation sfermions.

Supersymmetry breaking may be done with respect to the masses (sparticles and particles have different masses) or their coupling constants. Breaking supersymmetry in terms of the coupling constants prevents the resulting model from having the sought for cancellations in Equation 1.15. Thus, the Supersymmetry breaking mechanism usually affects mass terms, and the coupling constants from the Standard Model are preserved. As a result, since the couplings to the Higgs boson(s) are large for the (heavy) third generation fermions, the Higgs mass will depend strongly on massive particles (see Figure 3). Hence, the masses of the stau, sbottom and stop are expected to be not so far from their Standard Model partners, to preserve a natural cancellation of the divergent terms in Equation 1.15. This is known as *naturalness* and suggests that the third generation sparticles should be rather light, compared with other sparticles.

An easy way to obtain supersymmetric vertices and build diagrams representing processes involving supersymmetric interactions consists in replacing two of the SM particles with their superpartners in SM processes. At hadron colliders, the main production modes derive from QCD interactions. At the LHC, the dominant production modes are: squark-gluino, a pair

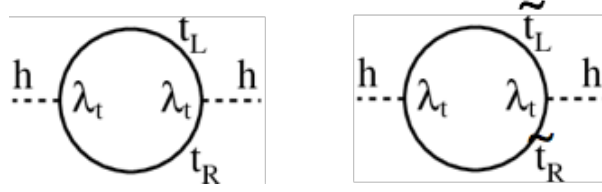


Figure 3. **Higgs mass one loop corrections, illustrated for top and stops** - The Higgs boson mass receives one loop corrections from top and stops with the same coupling constant.

Since the top mass is large, this λ_t constant is large as well, and, consequently, the stop cannot be too heavy.

of squarks, or two gluinos. Figure 4 shows the cross sections for various production modes. Inspired by the potential for supersymmetry to explain the lightness of the Higgs boson mass, this work will focus on searches for directly produced stop quark pairs, indicated by the blue curve in Figure 4.

1.2.5 Simplified models

A phenomenologically viable supersymmetric Lagrangian requires more than a hundred new free parameters, the search area is vast! To make searches for new physics interpretations easier, one can reduce the number of free parameters by making quite restrictive assumptions about the way supersymmetry is broken. For example, within the CMSSM (constrained minimal supersymmetric model, see for instance (12), (13)) :

- all the scalar particles at a grand unified scale have the same mass, m_0 ;
- all the gauginos (gauge bosons partners) at a grand unified scale have the same mass, $m_{1/2}$;

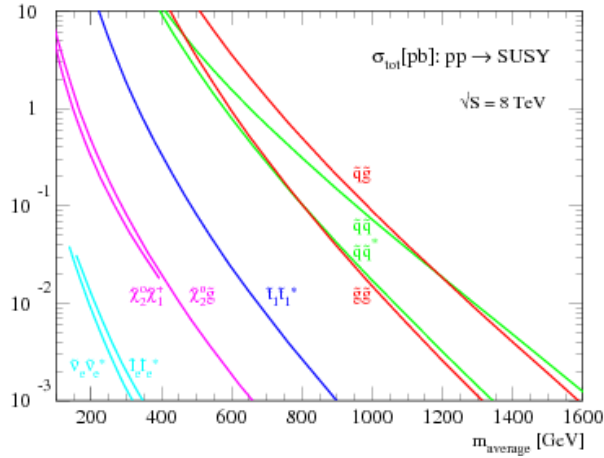


Figure 4. **Supersymmetric cross sections at the LHC, as a function of particles' masses** - QCD production dominates at hadron-hadron colliders, hence yielding squarks and gluinos whereas electroweak production is suppressed. Except for stops, shown separately, squarks are assumed to be degenerate, that is, have all the same mass. Plot available at http://www.thphys.uni-heidelberg.de/~plehn/includes/prospino/prospino_lhc8.eps

- all the trilinear couplings at a grand unified scale are the same, A_0 .

Two additional free parameters remain: $\tan \beta$, which is the ratio of the vacuum expectation values of the Higgs doublets and the sign of the higgsino mass term. This model is quite restrictive and does not allow to test a wide range of supersymmetric models.

In order to be more generic and efficient in our searches for new physics (including, but not only, supersymmetry), *simplified models* (14) have been designed, making the most of phenomenological similarities across various theoretical propositions. A simplified model consists of a small subset of hypothetical particles, their production mode and decay along with branching ratios, which are generic across a wide range of supersymmetric models. Simplified

models can thus be used to interpret searches for new physics pertaining to supersymmetric processes, maintaining a small number of model parameters but without making too restrictive assumptions. Further details on simplified models may be found in (15).

Many simplified models have been developed. The analysis developed in chapter 4 focused on the model known as $T2tt$, shown on Figure 5. It corresponds to direct stop production, each decaying to a top quark plus an LSP, often denoted as $\tilde{\chi}_1^0$ or $\tilde{\chi}$. The two LSPs escape the detector unseen. The top quarks decay through Standard Model processes: a weak decay towards a b quark and a W, the latter itself decaying either leptonically to a lepton plus a neutrino or hadronically to two quarks. The quarks fragment and hadronize into jets (see 2.3.3).

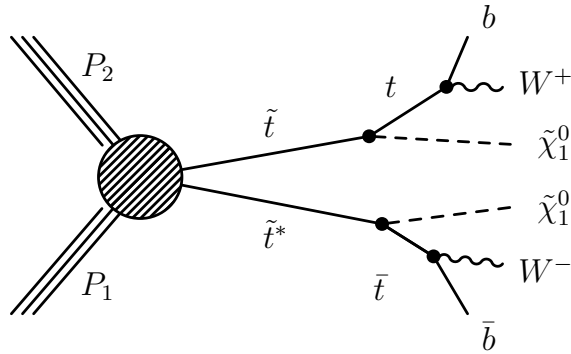


Figure 5. **T2tt simplified models** - The T2tt model refers to stop pair production, with each stop decaying to top plus an LSP.

CHAPTER 2

THE LHC AND THE CMS DETECTOR

2.1 The LHC

The Large Hadron Collider, or LHC, is a 27 km long circular collider, sitting at the French-Swiss border near Geneva, Switzerland. It is located in the LEP¹ tunnel between 50 and 170 meters below ground. After decades of development, the most powerful accelerator in the world became operational in September 2008, outclassing the Tevatron which accelerated protons and anti-protons up to about 1 TeV until 2011. Relying on superconducting technology, be it for radio-frequency (RF) accelerating cavities or the deflecting and focusing magnets, the LHC may accelerate protons as well as heavy ions, such as lead nuclei.

The LHC has been designed to push further the energy frontier and extend the reach towards hypothetical particles unseen so far. During the first Run, between 2009 and 2013, protons were accelerated up to 4TeV, giving 8TeV collisions in the center of mass frame of reference. It allowed to discover the long sought Higgs boson (see 1.1.2) in 2012. This thesis relies on the 8TeV data collected in 2012.

¹Large Electron Positron, collider dismantled in 2000.

Along the circular ring of LHC, four detectors have been placed to observe the collisions. Two of these, ATLAS¹ and CMS², are multi-purpose detectors, studying both heavy ions and proton-proton collisions to shed light on a variety of new physics theories. The other two are more focused on specific topics: LHCb concentrates on B-hadrons (hadrons involving b quarks) physics, while ALICE³ targets heavy ions collisions.

The LHC works with cycles, or proton fills. The protons are initially extracted from a small bottle of hydrogen and accelerated in several stages (Figure 77): first they are fed into a linear accelerator, then into the Proton Synchrotron (PS), followed by the Super-Proton-Synchrotron (SPS), and finally the LHC. Thus, the LHC is only the last step of several stages of acceleration that bring protons from 450 GeV up to 4 TeV. When beams are injected into the LHC, they usually collide for about ten hours, during which the luminosity, hence the probability for observing something interesting, decreases and the beams are dumped.

2.2 The CMS detector

This section summarizes the parts of the CMS detector that are relevant to this thesis and is based on (16) and (17), where more extensive descriptions of the CMS detector can be found.

The CMS detector is a multi-purpose detector, whose central feature is a 3.8 T superconducting solenoid magnet (Figure 6). Several sub-detectors are arranged within and around this

¹A Toroidal LHC ApparatuS

²Compact Muon Solenoid

³A Lead Ion Collider Experiment

magnet, to reconstruct the collisions provided by the LHC. The innermost layers of the detector are meant to reconstruct charged particles tracks for momentum measurement. They are followed by calorimeters measuring their energy. Last, after the solenoid, the outermost layers provide a second momentum measurement for muons.

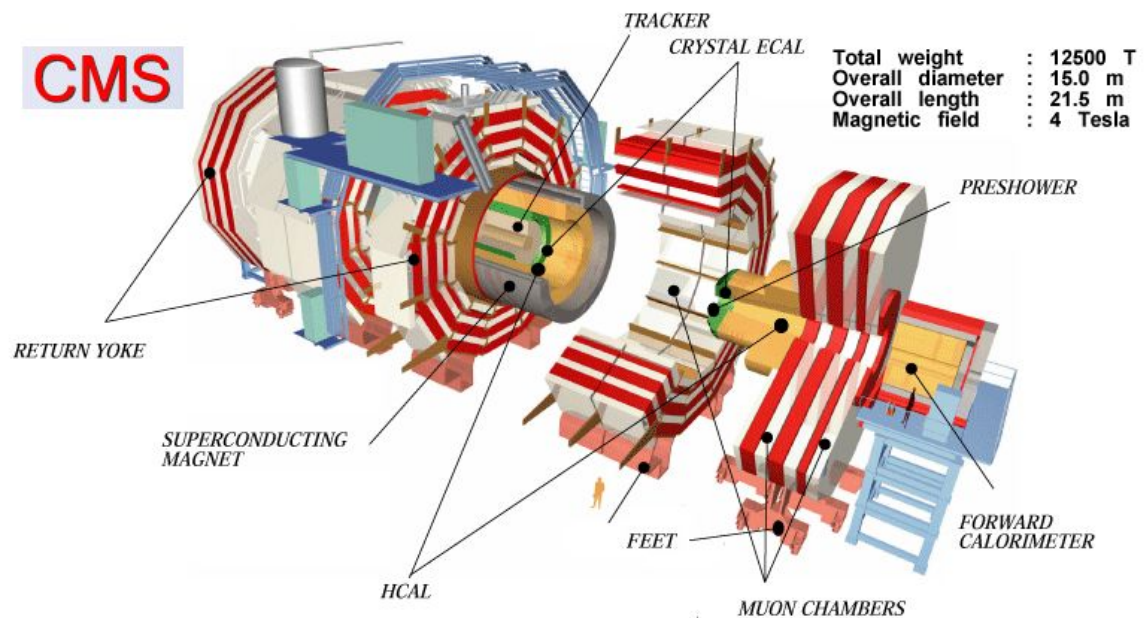


Figure 6. **The CMS detector** - A schematic view of the CMS detector ©Copyright CERN (2008-2014) for the benefit of the CMS Collaboration.

A right-handed coordinate system is used in CMS, which has a z -axis along the cylinder, pointing towards the counter-clockwise direction, has an x -axis pointing upwards, and has a

y -axis pointing towards the LHC center. We call ϕ the azimuthal angle in the transverse plane as measured from the x -axis, and we call θ the polar angle as measured with respect to the z -axis. In practice, the *pseudo-rapidity* is preferred to the latter angle, θ . The pseudo-rapidity is defined as $\eta = -\ln[\tan(\theta/2)]$, and is an approximation for rapidity $\frac{1}{2} \ln[(E + p_z)/(E - p_z)]$, which presents the interesting feature that the density of particles that is emitted within a (pseudo-)rapidity interval is constant.

2.2.1 Silicon tracker

The tracking system of the CMS detector relies on silicon sensors, either pixel or strips. These sensors are operated as reverse bias pn diodes. When an ionizing particle issued in a collision goes through a sensor, it creates electron-hole pairs, whose drift in the electric field induces a current. This current is a signature of the particle's passage. Multiple layers of sensors are used, so as to gather several points along the particle's path (or track), and hence allow to reconstruct of the track.

The tracking system of the CMS detector begins with a *pixel* subdetector. This rather small part of the detector presents the finest granularity. This is required by the fact that, the closer we are to the interaction point, the higher the density of particles to detect. The nearly 66 millions pixels of silicon, $100 \mu\text{m}$ by $150 \mu\text{m}$ in size, are arranged in three cylinders at radii 4.4 cm, 7.3 cm, 10.2 cm, completed with two endcap disks on each side. This setup allows a $10 \mu\text{m}$ resolution in $r - \phi$ (with r being the radial distance measured from the z -axis), and a $20 \mu\text{m}$ resolution in z , for pseudo-rapidity up to $|\eta| < 2.5$.

As we move away from the interaction point, the flux of particles decreases, allowing larger detector cells. The next layers of the CMS tracker are made with strips of silicon. Ten layers are used, four of them with two-sided modules whose sensors have a stereo angle of 100 mrad, to provide a two-dimensionnal information. About 10 millions strips are arranged to provide a resolution of about $50\text{ }\mu\text{m}$ in $r - \phi$ and z .

2.2.2 Calorimeters

The CMS detector is equipped with two calorimeters, which fit within the solenoid (2.2.4): the ECAL, for Electromagnetic Calorimeter, and the HCAL, for Hadronic Calorimeter.

2.2.2.1 Electromagnetic calorimeter

The electromagnetic calorimeter of CMS is designed to measure the energy of photons, electrons and positrons. It consists of more than 60000 lead tungstate (PbWO_4) truncated-pyramid shape crystals in the barrel part ($|\eta| < 1.479$), complemented by about 8000 crystals in the endcaps (up to $|\eta| < 3.0$). The crystal size varies to provide a hermetic detector. In the barrel part, they are 23 cm long, while their cross section increases from $2.2 \times 2.2\text{ cm}^2$ to $3.3 \times 3.3\text{ cm}^2$. In the endcaps, they are 22 cm long with a cross section of $2.8 \times 2.8\text{ cm}^2$ for the front face and $3.0 \times 3.0\text{ cm}^2$ in the rear face. Electromagnetic particles lose most of their energy in the crystals through radiation: high energy photons turn into electron-positron pairs, while electrons and positrons radiate photons through bremsstrahlung radiation¹. These two processes give rise to *electromagnetic showers* (Figure 7). When the energy of the emitted

¹Any charged particle following a bent path radiates, and loses energy through photon emission. This radiation may also be referred to as *synchrotron* radiation.

photons is no longer sufficient to create pairs, photons propagate in the transparent material and cause it to scintillate (that is, atoms from the crystals are excited by these photons, which they absorb before re-emitting energy through new photons, carrying a blue-green light). The resulting light is collected by photodetectors, either avalanche photodiodes, APDs, or vacuum photo-triodes, VPTs, depending on the magnetic field.

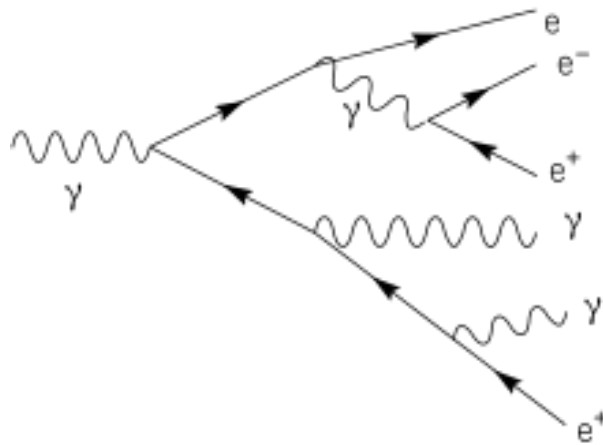


Figure 7. **An electromagnetic shower** - Highly energetic photons create electron-positron pairs, which emit photons through bremsstrahlung radiation because of the magnetic field.

2.2.2.2 Preshower

As we can see from Figure 6, in the endcaps, the ECAL is preceded with a preshower. This sampling calorimeter is made of a lead layer within which particles initiate their electromagnetic shower, and silicon strips 2mm wide measuring the energy deposit and the shower shape. The

much finer granularity of this extra layer enables one to detect two nearby photons, hence more easily identify the presence of high momentum neutral pions.

2.2.2.3 Hadronic calorimeter

The Hadronic calorimeter is designed to measure the energy of the light (hence quasi-stable) hadrons emitted in a collision - the heaviest ones are too short lived to be seen directly. The HCAL is a sampling calorimeter, consisting of alternating layers of brass (Cu, Zn) and plastic scintillator tiles. Particles lose energy in the brass absorber, interacting with the dense nuclear material and causing the emission of secondary nuclear particles. Both initial and secondary particles cause the active plastic material to scintillate, producing light. Thus, we may observe hadronic showers in the HCAL, just like we see electromagnetic showers in the ECAL. The scintillator layers are equipped with wavelength shifting fibers, spliced to optical fibers for information transmission. They are segmented in cells with a size of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. Across the full thickness of the calorimeter, these cells define towers. The energy of the initial particle may be reconstructed from the full tower, or, possibly, several adjacent towers. The HCAL is made of a cylindrical barrel and endcap plugs, which together cover pseudo-rapidity up to $|\eta| < 3.0$. A forward hadronic calorimeter (HF), based on longitudinal quartz fibers embedded within iron, extends coverage up to $|\eta| < 5.0$, so as to be as hermetic as possible, for missing transverse momentum estimation (see 2.3.4).

2.2.3 Energy absorption in the CMS detector

Two quantities are used to describe the energy absorption in calorimeters: the radiation length X_0 , for electromagnetic energy absorption, and the nuclear interaction length λ_0 , more suitable to describe particles interaction with hadronic calorimeters.

The radiation length X_0 definition encompasses two aspects: for electrons, it is the mean distance over which they lose most of their energy (only $1/e$ left) through bremsstrahlung radiation; for photons, it is $7/9$ of the mean free path before pair production of electrons.

The nuclear interaction length λ_0 is the distance after which the number of relativistic charged particles has been reduced by $1/e$, while they penetrate into matter.

Calorimeters are designed so as to absorb most of the energy to measure it. To preserve this energy measurement in the calorimeters, the tracking system should absorb as little energy as possible. For the CMS detector, the material budget is essentially¹ divided as follows:

- a CMS tracking system that has at about 1.6 radiation lengths and 0.4 interaction lengths;
- an ECAL that corresponds to about 26 radiation lengths for electrons, photons, and about 1 nuclear interaction length for hadrons;
- a preshower with about 3 interaction lengths;
- an HCAL whose active depth is at least 5.82 interaction lengths.

1.6 radiation length for the tracking system is not so small: tracking gaseous detectors such as time projection chambers (used by ALICE for instance) absorb just a few percent of

¹The values vary depending on η .

a radiation length. Yet, the CMS detector has to be able to take data at a 40MHz frequency, thus needs not only light but fast detectors. This is why a silicon based tracking system has been chosen, despite the rather high amount of material it represent.

2.2.4 The magnet

The magnet allows one to determine the sign and the momentum of the charged particles emitted in a collision, thanks to the bending of their trajectories resulting from the Lorentz force. The higher the magnetic field, the higher the bending power: this, along with feasibility considerations, decides of the size of the solenoid.

The CMS magnet is a superconducting solenoid, with a four winding-layers of Niobium-Titanium cables. This solenoid, which is 13 meters long with an internal diameter of 6 meters, receives a current of 18000 Amperes, and produces a 3.8 T magnetic field. The field is mostly homogeneous within the solenoid as can be seen on Figure 8. Around the solenoid, four layers of iron constitute the iron yoke, intended to close the field lines.

2.2.5 Muon system

The Muon System is extensively described in (19), from which the pictures of this paragraph come. As can be seen on Figure 9, the muon system comprises three kinds of detectors.

In the barrel part, the background rate and the magnetic field are low enough to allow for the use of gaseous drift tubes (DTs). In the endcap parts, both the background rate and the magnetic field are higher, and hence gaseous cathode strip chambers (CSCs) are used. Finally, resistive plate chambers (RPCs) are used everywhere for triggering purposes, thanks to their fast detector response.

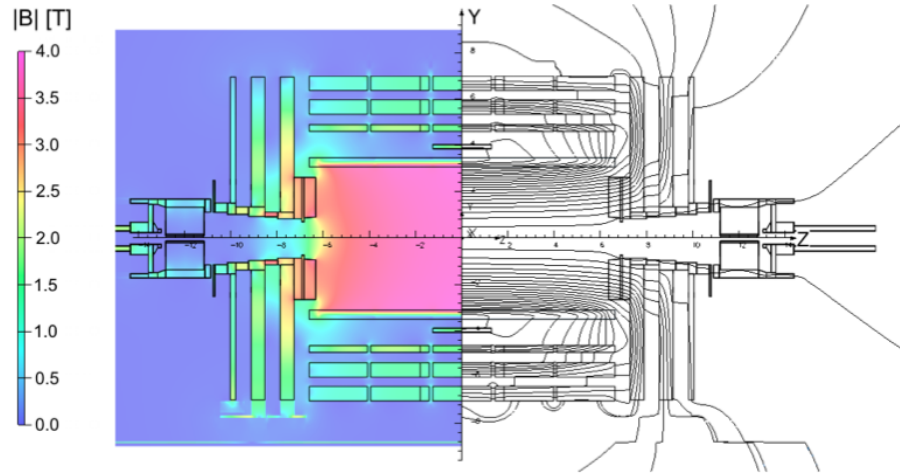


Figure 8. **Map of the magnetic field magnitude in the CMS detector** - Within the solenoid, hence the tracking and calorimetric systems, the field is homogeneous, and reaches nearly 4 T. The iron yoke concentrates the field and returns it so that the field outside of the detector quickly goes down to 0T. Within the muon chambers outside of the solenoid, the field also changes sign, which causes the direction of bending from muon tracks to change (18).

2.2.5.1 Drift tubes

The primary muon barrel system consists of 250 drift tube chambers, located within the iron return yoke of CMS. They are longitudinally split to fit within the five barrel wheels, and each wheel contains four layers of chambers. These concentric layers - labelled MB1, 2, 3, 4 as we move further away from the axis - contain 12 drift tube chambers each, except for the last concentric layer, MB4, which contains 14 chambers. Each chamber consists of three superlayers: a superlayer is itself made of four planes of drift tubes. Within a chamber, the first and third superlayers have tubes parallel to the beamline to provide measurements in the $r - \phi$ plane, whereas the second layer is orthogonal to provide a longitudinal measurement.

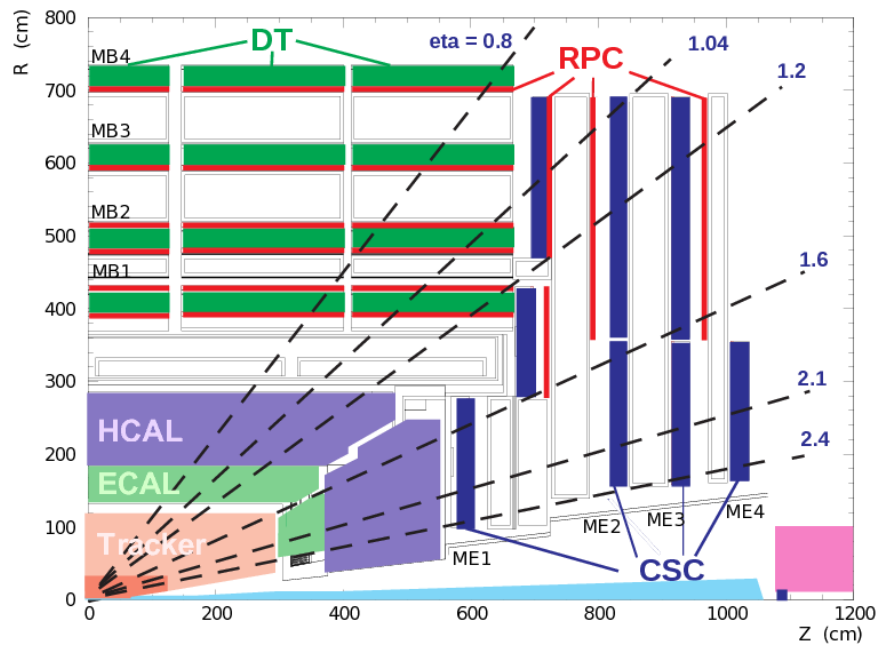


Figure 9. **Location of muon chambers in CMS** - Drift tubes in the barrel part, CSC in the endcaps, with the addition of RPCs in both (16).

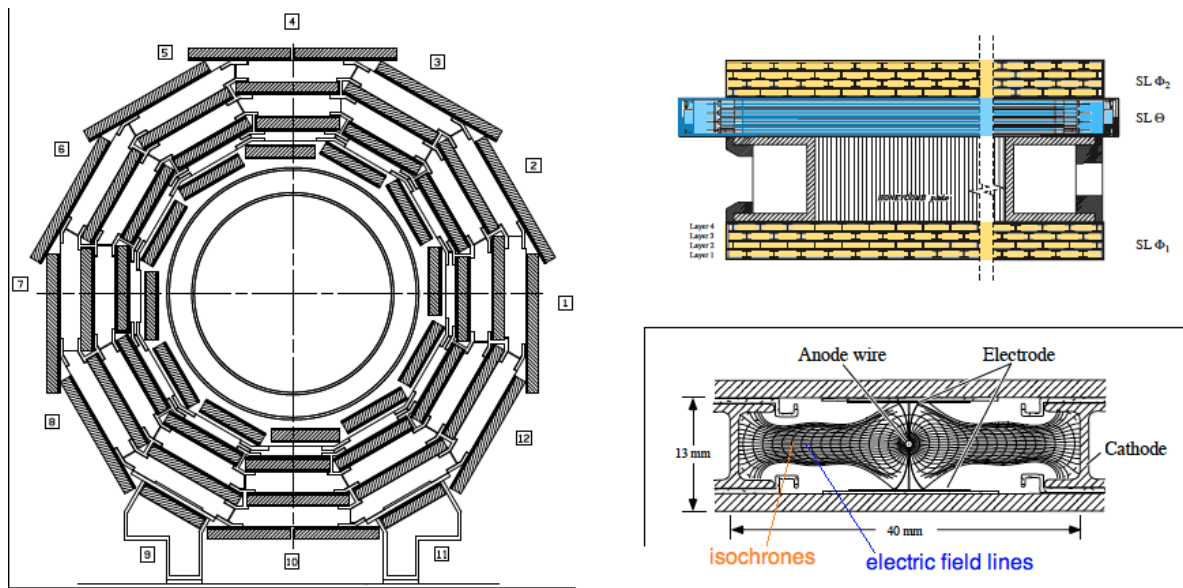


Figure 10. **Muon system: drift tubes** - (left) Drift tubes location within a barrel wheel; (top right) DT superlayer: in yellow shades, the two r - ϕ layers, in blue shades, the z layer; (bottom right) a single drift tube - cross section. The lines indicate the electric field lines and the isochrones (electrons emitted along a given isochrone will have the same drift time, that is, the time needed to reach the anode).

When a muon enters a drift tube, it ionizes the gas (Ar mostly, plus 10 to 20% of CO₂ at atmospheric pressure). An electric potential difference between the drift tube walls and an anode creates an electric field that drives ionization electrons towards the anode, with a maximum drift time of 400 ns¹. A single drift tube is 4 cm-wide and at most 2.5 m long. The spatial resolution of a drift tube chamber is a few 100 μm .

2.2.5.2 Cathode strips chambers

In the endcap regions, which extend forward up to $|\eta| < 2.4$, muon chambers have to accommodate a “not so friendly” environment. On the one hand, the magnetic field can be as high as 3 T and may be significantly inhomogeneous across a disk, and on the other hand, the background rate is much higher than in the barrel region. In this case, *background* refers to any particle not originating from the main interaction, which may leave a signal in a muon detector. Such particles are typically soft and at high pseudo-rapidity. Such backgrounds may arise from different sources:

- non-collision muons that arise from the halo of the beam, which are due to interactions of the beam with the beam pipe or stray gas molecules far upstream of the collision point.

Such muons can occur with a rate of up to 1 kHz/cm² for the innermost layers;

¹400ns: this is 8 times the duration between two collisions. To avoid mixing consecutive collisions, the granularity of the detector has to be adjusted so that it is highly unlikely that two - or eight - consecutive collisions will send particles in the very same cells. On average, only one event out of several hundred will send a muon to a muon system cell. We also have to think about possible sources of ‘background’ (something, not a muon from the main interaction, able to leave a signal in a muon system cell). For the innermost wheel, the background rate is of the order of 10Hz/cm²: a particle every 2 million collisions per cm², every two thousand collisions per drift tube at most.

- punch-throughs from the calorimeters, which represent electromagnetic or hadronic particles that, despite the absorption properties of the heavy calorimeters, managed to produce showers that traverse them completely;
- random hits induced by neutrons/gammas originating from hadronic cascade, accelerator components;
- electromagnetic debris from very high energy muons passing through matter (muon bremsstrahlung).

To provide muon detection that is able to cope with this background in the forward endcaps, close to 500 cathode strip chambers are arranged within four disks, labelled ME1, ME2, ME3, ME4 (ME1 is located closest to the barrel, ME4 closest to the beam). ME2, ME3, and ME4 (ME1) carry two (three) concentric rings of chambers. The layout may be seen on Figure 9.

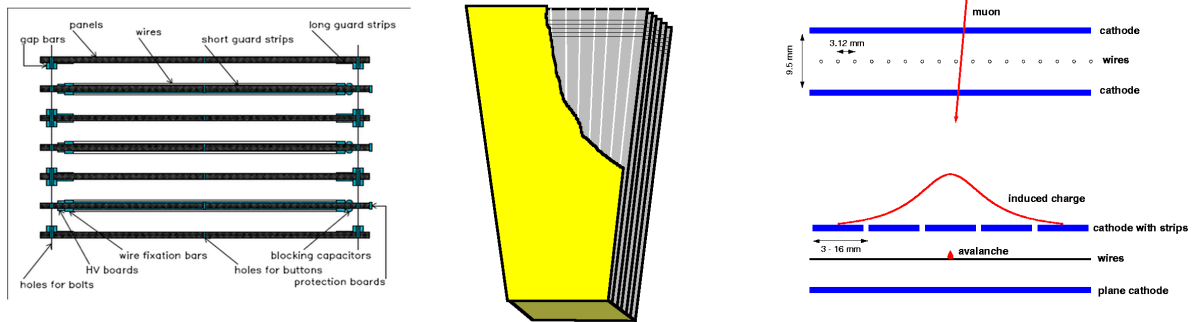


Figure 11. **Muon system: CSC** - (left) CSC principle; (center) a single Cathode Strip Chamber ; (right) an exploded view of a CSC chamber.

A cathode strip chamber consists of seven copper clad planes, six of which carry radial cathode strips, facing orthogonal groups of anode wires. The gap in between the planes is filled with a gas mixture (AR-CO₂-CF₄). When a muon hits a chamber, it ionizes the gas, leading to an avalanche production of electrons, collected by the wires. The signal collected by the wires is fast enough to be used for trigger purposes, and provides a rough measure of the radial coordinate (a few millimeters precision). The charge on the wire induces an image charge on the facing strips, allowing a precise (a few hundreds of microns) azimuthal coordinate measurement. The CSCs resolution is a few mm.

2.2.5.3 Resistive plate chambers

Resistive Plate Chambers (RPCs) are also gaseous muon detectors. They complement the two muon detectors described previously for trigger purposes, thanks to their fast readout time: they can provide the time and position of muon hits in 3 ns, which is much less than the nominal bunch crossing spacing of 25 ns.

RPCs consists of two high resistivity planes coated with graphite, so as to bring them to high voltages. This causes one plate to be an anode and the other a cathode. The gap between those planes is filled with gas, which gets ionized by particles going through, yielding avalanches of electrons. These drift in the gas towards the cathode, which they traverse to reach strips of aluminum, where the electrical current they induce signals the passage of a muon.

2.3 Event reconstruction

This section presents some objects or concepts that are used elsewhere in the thesis, but does not aim at giving a full and detailed description of the event reconstruction.

2.3.1 Particle Flow, or Global Event Description, algorithm

Our detector is good enough to provide a generator-level-like picture...

When we describe a production process and its subsequent decay chain, we often draw Feynman diagrams such as Figure 12. A detector view of the same event is quite different. First, in the diagram we see all of the interaction steps, while in our detector, we only have access to the last stage that ends with stable particles (or particles with a long enough lifetime¹ so as to be visible inside the detector). Second, the information we collect in the detector are merely electrical signals. These have to be interpreted as energy deposits and combined to derive high-level objects that are suitable for data analyses, such as identified particles with their four momenta.

The particle-flow algorithm ((20), (21)), also known as the Global Event Description, aims at using the information from the whole detector to reconstruct particle-like objects, just as we would see them in the final state of simulated events, before processing them through a detector simulation. It provides us with a list of particles: electrons, muons, photons, light hadrons (pions mostly). The muons are first obtained using deposits in muon chambers matched to tracks in the tracking system. After removing the latter from the list of available tracks, tracks that can be matched to energy clusters in the ECAL and HCAL give the charged hadrons. The remaining tracks may be matched with energy clusters in the ECAL only, so as to reconstruct possible electron candidates. However, special care is needed, since electrons usually radiate

¹Muons decay to electrons within $2\mu\text{s}$, positrons would eventually annihilate with electrons.

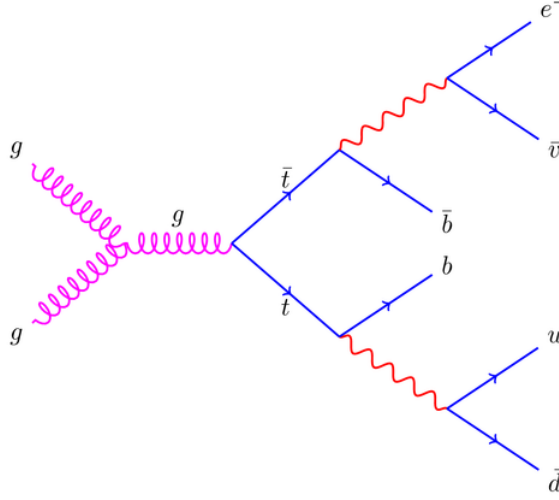


Figure 12. **Example of a Feynman diagram, representing a full process, from production to final state.** - All the particles, at each step, can be seen. In our detector, we only have access to the final state, through electrical signals.

bremsstrahlung photons as they travel through the tracker material, and the ECAL energy clusters from the bremsstrahlung photons need to be connected to the electron, and not identified as individual prompt photons. Once all of the tracks have been used, the remaining energy clusters in the ECAL and HCAL provide neutral hadrons, and the last energy clusters in the ECAL only (i.e. not connected with any HCAL energy deposits) furnish photon candidates. The list of objects finally returned by the particle flow algorithm is made of what we call *particle flow candidates*.

Such a reconstruction is made feasible by the strong magnetic field, which provides a good separation between the trajectories of charged and neutral particles. The granularity of each

sub-detector is also a key requirement. The complementarity between the energy and momentum measurements in the different sub-detectors result in a better determination of the particles' four momenta. These particles may then be used to compute other physics analyses objects, such as jets and missing transverse momentum.

2.3.2 Lepton identification and isolated track veto

2.3.2.1 Electron and muon identification

Particles used in analyses are reconstructed out of the electrical signals gathered from the detector. The particle identification process is never fully accurate: it may have a varying *efficiency*, coming along with a given *purity*. Typically, when provided a list of electrons for a given event, we only have access to a certain percentage of the complete set of electrons that were produced (this is the efficiency). And, that list may contain particles identified as electrons, even though they are actually not (in which case, the purity of the list is below 100%). In practice, both quantities vary against each other: the more efficient the reconstruction is, the less pure is the final set of particles. The stricter, or *tighter*, our reconstruction criteria are, the better the purity of the sample – but we miss some particles when doing so.

When designing analyses, and establishing a selection that targets specific final state signatures, or channels, we want to ensure that we have a sound understanding of the objects we use. For muons (22), we use the following criteria to improve the purity of the particles initially labelled as muons:

- to differentiate prompt muons (that arise from the primary interaction) from secondary muons (that arise from the decay of other particles in flight) and to reject muons from

cosmic rays, we require a minimum number of tracker and pixel hits (to ensure a good p_t measurement), along with small longitudinal ($d_z < 5$ mm) and transverse ($d_{xy} < 2$ mm) distances between the track and the primary vertex;

- to avoid *punch-through*, that is particles from showers not fully absorbed in the calorimeters, we require at least two hits in the muon chambers and at least one hit in the global fit between the muon system and the tracker.

The resulting set of muons is composed of what we call *tight* muons, while the initial set of muons are *loose* muons. The efficiency and purity of the resulting set is evaluated using well-known samples, both simulated events and real data. In data, we usually make use of $Z \rightarrow \mu\mu$ events. Tight muons are used when we want to clearly select a muon sample. Loose muons are used when we want to veto muons. Doing so, we can obtain well-defined samples, matching the desired topologies and channels.

For electrons, similar principles apply, with different selection criteria (23):

- the sum of the energies of any ECAL clusters associated with the electron energy must match the track momentum;
- there must be a spatial match between the ECAL clusters associated with the electron and its track;
- the spatial width of the ECAL clusters should be limited in η ;
- any hadronic leakage, determined by H/E (the ratio of hadronic energy deposit energy divided by electromagnetic energy deposit), must be small for prompt electrons;

- to reject electrons from photon conversion, we rely on the transverse impact parameter (which should be larger for electrons arising from conversions), we use the number of missing hits in the tracker (which should again be larger for electrons arising from conversions), and we look for partner tracks that are consistent with conversions.

The final electron identification algorithm makes use of all or part of these variables, and is used to define different working points: *tight* electrons correspond to the working point with the best purity and *loose* electrons correspond to the working point with highest efficiency.

Moreover, to further ensure that our selected leptons actually originate from the primary interaction (*prompt* leptons), we apply isolation criteria, to avoid leptons from hadronic decays within jets. The definition of isolation derives from the sum of the hadronic and electromagnetic energy deposits in a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ around our lepton candidate. As the number of pile-up interactions (3.3) increased during the first LHC run, these isolation requirements had to be adjusted so as to discard pile-up particles surrounding prompt leptons. The relative isolation is then defined by dividing the energy sum by the lepton p_t . We require this relative isolation to be smaller than 0.15.

2.3.2.2 Isolated track veto

In order to further reject leptons that would fail even the loose identification, we apply an isolated track veto at particle flow candidate level (24). Isolated tracks are those particle-flow candidates coming from the primary vertex that are charged, have $p_t > 10\text{GeV}$ and have relative isolation (as defined above) of less than 0.1. In the leptonic final states, the isolated track veto is only applied if the charge of the track is opposite to that of the tight lepton.

2.3.3 Jets

Jets appear in experiments as collimated and energetic bunches of particles. Quarks or gluons, initially emitted in the collision, radiate gluons that can then split into pairs of quarks or radiate even more gluons. This process continues in what is known as a *parton shower*. At some point in the shower, these partons combine into hadrons, a process known as *hadronization*. Some of the hadrons may also subsequently decay, via the strong, electromagnetic, or weak interactions, producing light hadrons, photons, or leptons within jets. Usually, about two third of a jet content is made of hadrons and the remainder photons. The data we collect contains the final particles emitted during the hadronization process, which we then have to cluster to build the jets as shown on Figure 13. Doing so, we can assess the energy and kinematic properties of the initial quark or gluon parton that gave rise to the jet. While the first jet algorithms merely assembled particles or calorimeter cells that could fit within a cone, LHC experiments now rely on more complex principles, such as k_t and *anti- k_t* algorithms ((25), (26)). These sequential algorithms compute the following distance between all possible inputs i and j (reconstructed particles or calorimeters cells) and the beam B :

$$d_{i,j} = \min \left(p_{t,i}^{2k}, p_{t,j}^{2k} \right) \frac{\Delta_{i,j}^2}{R^2} \quad (2.1)$$

$$d_{i,B} = p_{t,i}^{2k}, \quad (2.2)$$

where $p_{t,...}$ is the momentum of input i or j , k is an integer between -1 and 1, and R is a parameter describing the radius of the jet. When $d_{i,j}$ reaches its smallest value, i and j are

combined and the next round starts. If $k > 0$, we have what is called the k_t algorithm. For $k = 0$, we name it the Cambridge-Aachen algorithm, and when $k < 0$, the algorithm is referred to as *anti- k_t* .

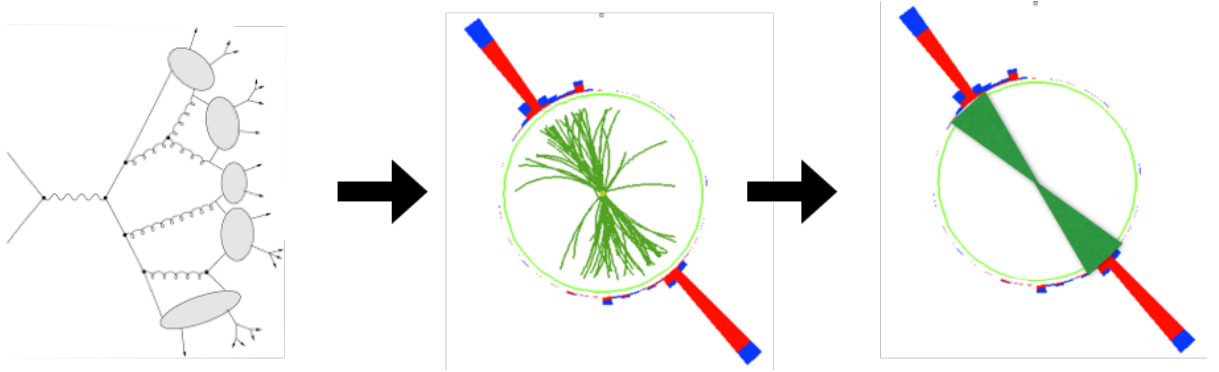


Figure 13. **From parton to jets** - Quarks or gluons emitted at tree level radiate gluons & quarks, making *parton showers*, which recombine within hadrons. Combining these hadrons signatures in the detector within *jets*, we may get an estimation of the initial partons properties.

At the CMS experiment, we (mostly) used *anti- k_t* jets, with a radius of 0.5 for the analysis of the first Run (Run I) of the LHC data. It has a several of good properties related to jet reconstruction stability. The probability for a quark or a gluon to radiate a gluon diverges at low energy and at small angles (0 or π , with respect to the flight direction of the quark). The *anti- k_t* algorithm is said to be both *infrared safe* and *collinear safe*: that is, the jet algorithm is stable with respect to extra radiation (hence more objects to cluster), be it at low energy

or at small angle. Cambridge-Aachen and k_t jets also have these properties. This was not the case with the first cone algorithms (25).

2.3.3.1 b-jets

Hadrons containing a b quark (known as B hadrons) generate jets with characteristic features that help to distinguish them from jets generated by other, light quarks or gluons (27). These features stem from the rather long lifetime of B hadrons, typically a few ps, or 10^{-12} s. This is very long compared to strong interaction processes, typically of the order of 10^{-24} s: B hadrons decay through weak interaction, driven by the $|V_{cb}|$ parameter of the CKM matrix (28). The amplitude of this process is proportional to the square of $|V_{cb}| = (41.1 \pm 1.3)10^{-3}$, which is quite small, and hence gives rise to the long time needed for b quarks to decay to c quarks.

This rather long lifetime allows B hadrons to travel about $500 \mu\text{m}$ before decaying, although too short to reach the calorimeters or even the tracker. Thus, b-jets appear as originating from secondary vertices that are displaced with respect to the interaction point. Further, the impact parameters of tracks associated to a candidate b-jet tend to be larger than for prompt tracks. Thanks to the large mass of b-hadrons compared to their daughter particles, the latter also tend to possess a hard momentum spectrum. Leptons coming from semi-leptonic decays of b quarks, within the jet, may also be used in the identification algorithm. Several algorithms have been developed, with different levels of complexity, using variables quantifying these different properties. The Combined Secondary Vertex (29) algorithm, which combines several variables to build a single discriminant value, is used by most CMS analyses and also by this work, as described in Chapter 4.

Two key figures of merit characterize a tagging algorithm: the efficiency and the misidentification rate. Each of the various algorithms used has different working points: loose (defined to have high efficiency but 10% misidentification rate), medium (1% misidentification rate) and tight (0.1% misidentification rate), for jets with an average p_T around 80 GeV/c.

2.3.4 Missing transverse momentum

Despite having a hermetic detector to catch as many particles as possible and working on calibrations¹ to avoid mis-measurements, there are some stable particles which will always leave our detectors unseen. These can be the Standard Model neutrinos - which can go through Earth as if nothing was there, or some new particle, unknown so far, like a dark matter candidate or an LSP.

In electron-positron collisions, such as the collisions provided by the LEP collider, one can get a hint of what has been missed by relying on the momentum balance: the momenta of the final decay products must add up to be equal to the input momentum. If the overall output momentum does not match the input momentum, which we know precisely, we have some *missing momentum*, interpreted as neutrinos or some new particle(s).

In a hadronic collision, one never knows exactly what is the input energy of an interaction. Indeed, protons being composite particles, the actual interaction takes place between their parton constituents: quarks or gluons. When a proton carries 4 TeV, one does not know how

¹Calibration is the process of tuning the output energy signals collected from the detector to match the input energy of particles, whose energy is constructed to be known (such as in a test beam or a known physical process).

much of that energy is carried by those constituents. Thus, one cannot determine the initial center of mass energy of a parton-parton interaction that takes place within a proton-proton collision. Yet, knowing that protons are not accelerated in the transverse plane, we can apply momentum conservation in this plane: if one adds up the transverse momenta of all particles observed in a given event, it should be zero. Thus, in hadronic collisions, one only has access to *missing transverse momentum*, which is computed as the negative vector sum of all visible particles' transverse momenta¹.

2.4 Event simulation

In order to set up analyses, we rely on simulated events ((30), (31)), be it for known Standard Model processes or new physics signals . These simulated samples are often referred to *Monte Carlo* (MC) samples because they involve random numbers generation to numerically evaluate integrals or generate quantities from probability density functions.

2.4.1 Event generation

Each event may be thought of as a succession of steps :

1. hard process (with possible initial and final state radiations);
2. parton showering;
3. hadronization;

with, in parallel, an underlying event, and possibly secondary interactions. Many event generators have been developed, some of them general purposes, some targeting more specific steps.

¹Often referred to as missing transverse energy, but energy is a scalar, it cannot be transverse.

Event generation begins with evaluating the cross section of a given *hard process*. This is done at tree level(LO), or leading order, plus next-to-leading-order (NLO), including one loop corrections or extra radiations (sometimes even at next-to-next-to-leading order(NNLO)). Starting from the Feynman diagram for a given process, the corresponding cross-section is derived using the Feynman rules to write the matrix element.

Then, one needs to simulate the *parton shower*: when quarks or gluons are emitted at tree level in the hard process, they radiate gluons and quarks, yielding *parton showers* because of collinear and infrared emission detailed in 2.3.3.

Afterwards, these partons have to recombine to constitute the hadrons actually observed (or whose decay products will be observed): this is the *hadronization* step. While this step is not really understood, different models exist, such as the *Lund String Model* which is implemented in PYTHIA (32), while other generators such as Herwig (33) prefers the *Cluster Model*. To put it short, the Lund String Model (left plot on Figure 14) extends a string across quarks and gluons and combine them within color-less quark pairs or triplets. The Cluster Model, illustrated on the right plot of Figure 14, splits each gluon into quark-antiquark pairs, and builds hadrons out of the whole set of quarks and anti-quarks.

In parallel, event generators need to account for the underlying event, that is, what happens to the remnant of the protons not taking part in the main interaction.

2.5 Detector simulation

The simulated events are processed through a GEANT based detector simulation (36). GEANT allows to simulate a full detector geometry and to reproduce particles interaction with

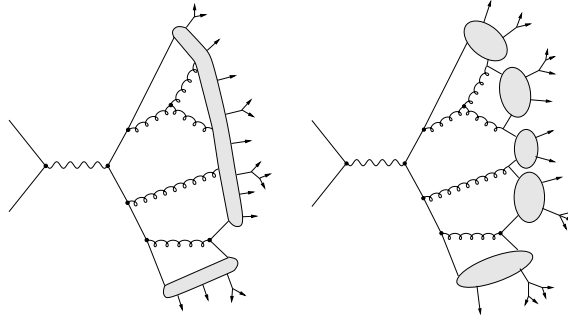


Figure 14. **Hadronization models** - The Lund String Model (left plot) relies on a string across all quarks and gluons and then splits it to build hadrons, while the cluster model (right plot), splits each gluon into two quarks and then cluster quarks within hadrons. (34), (35).

matter (including a wide variety of phenomena, such as nuclear interactions, bremsstrahlung radiations, scintillation, etc.). Thus, a GEANT based description of the CMS detector is able to provide us with a detector view of the events generated with the previous steps. While this *full simulation* provide a very good description of the detector, it can be rather slow. Thus, a *fast simulation* (37) has been developed for the CMS detector. The latter uses a simplified detector geometry, parameterized detector response and pattern recognition to achieve a reconstruction up to 100000 faster.

In the analysis detailed in chapter 4, we will use samples which have been generated with MADGRAPH v4.22 (38) for the hard scatter combine with PYTHIA6 for the subsequent showering and hadronization. Standard Model background processes have been processed through the Full detector simulation, while the signal samples (for which we need many mass points in the stop-LSP masses plane) relies on the Fast Simulation.

CHAPTER 3

TRIGGERING AT CMS

Trigger Rates for L=4e29

Contents:

- [Trigger Rates for L=4e29](#)
- [Rates for Full Table](#)
- [Rates for Primary Datasets](#)

Rates for Full Table

[hide](#)

- Evaluated using run 136100
- run over lumi sections 407~1148 (Ave. Lumi = 1e29cm²/2)
- artificially added HLT_L1MuOpen_DT with same parameters as HLT_L1MuOpen. This is a new path.
- For previous versions of this menu, see [JMTMDLumi4e29](#)

[hide](#)

Path Name	L1 condition	L1 Prescale	HLT Prescale	HLT Rate [Hz]	Total Rate [Hz]	Avg. Size [MB]	Total Throughput [MB/s]
HLT_L1Jet6U_NoBPTX	L1_SingleJet6	1	10000	0.00±0.00	0.00	0.15	0.00
HLT_L1Jet6U	L1_SingleJet6	1	200	6.95±1.27	6.95	0.15	1.04
HLT_L1Jet10U_NoBPTX	L1_SingleJet10	1	1000	0.00±0.00	6.95	0.15	1.04
HLT_L1Jet10U	L1_SingleJet10	1	20	10.43±1.55	17.38	0.15	2.61
HLT_Jet15U	L1_SingleJet5	1	10	5.56±1.13	22.94	0.15	3.44
HLT_Jet15U_HcalNoiseFiltered	L1_SingleJet5	1	10	5.56±1.13	22.94	0.15	3.44
HLT_Jet30U	L1_SingleJet20	1	1	6.26±1.20	27.34	0.15	4.10
HLT_Jet50U	L1_SingleJet30	1	1	1.39±0.57	27.34	0.15	4.10
HLT_DiJetAve15U_8E29	L1_SingleJet6	1	5	6.72±1.25	29.42	0.15	4.41
HLT_DiJetAve30U_8E29	L1_SingleJet20	1	1	3.71±0.93	29.42	0.15	4.41
HLT_FwdJet20U	L1_IsoEG10_Jet6_ForJet6	1	1	2.32±0.73	31.51	0.15	4.73
HLT_QuadJet15U	L1_QuadJet6	1	1	0.00±0.00	31.51	0.15	4.73
HLT_DoubleJet15U_ForwardBackward	L1_DoubleForJet10_EtaOpp	1	1	0.00±0.00	31.51	0.15	4.73
HLT_L1MET20	L1_ETM20	1	1	7.65±1.33	35.68	0.15	5.35
HLT_MET45	L1_ETM30	1	1	0.46±0.33	35.68	0.15	5.35
HLT_MET100	L1_ETM70	1	1	0.00±0.00	35.68	0.15	5.35
HLT_HT100U	L1_HTT50	1	1	0.70±0.40	35.68	0.15	5.35
HLT_BTagMu_Jet10U	L1_Mu3QE8_Jet6	1	1	0.46±0.33	36.14	0.15	5.42
HLT_BTagIP_Jet50U	L1_SingleJet30	1	1	0.23±0.23	36.14	0.15	5.42
HLT_DoubleLooseIsoTau15	L1_DoubleTauJet14 OR L1_DoubleJet30	1, 1	1	1.62±0.61	36.61	0.15	5.49
HLT_SingleLooseIsoTau20	L1_SingleTauJet20 OR L1_SingleJet30	1, 1	1	5.56±1.13	37.30	0.15	5.60
HLT_L1Mu14_L1SingleEG10	L1_SingleMu14 AND L1_SingleEG10	1, 1	1	0.00±0.00	37.30	0.15	5.60
HLT_L1Mu14_L1SingleJet6U	L1_SingleMu14 AND L1_SingleJet6	1, 1	1	0.46±0.33	37.76	0.15	5.66

Figure 15. **First step into Trigger World** - Recollection of the first (and puzzling) trigger meeting I attended: meaningless names and people arguing about the numbers.

3.1 Setting the scene: collisions at the LHC, why do we need a trigger system

For its first run, the LHC beams are made of 1368 bunches (during normal operations), each containing a hundred billion protons. Thus, approximately seven hundred pairs of bunches can collide within the CMS detector. Traveling extremely close to the speed of light (99.999997%), each bunch travels through the LHC ten thousand times per second.

As a result, the CMS detector sees about seven million bunch crossings per second, most of them being focused to actually collide bunches. When bunches collide, the beam diameter is no more than a few microns. No more? As small as a human hair, this is still way bigger (a billion) than the size of protons (which are actually so small that even the idea of size is losing sense)! Most of the protons miss each other and just go through as if nothing had happened¹. Even when a pair of protons interacts, most of the time, it is just a glancing interaction. What we are looking for are head-on collisions, liberating lots of energy - out of which we may create new particles.

Besides, seven million collisions per second, with an average event size of the order of one megabyte, amount to about seven terabytes of data per second, more than one petabytes per hour, even worse over a full year: we cannot store everything.

Therefore, each time two bunches collide in CMS, we analyze the collision very quickly to decide whether to record it or not. This selection is performed by the *trigger system*. CMS

¹Furthermore, interactions between protons are not just a matter of geometry: they also interact because of strong & electromagnetic forces. In fact, the total proton-proton cross-section is about 100mb. This yields about half a billion interactions per second.

relies on a two level trigger system, called Level 1 (L1) and High Level Trigger (HLT). The first one selects up to 100 000 collisions per second within a latency of $3.2\mu\text{s}$, while the second reduces further the amount of data down to 350-450Hz, within an average time of 160ms (in 2012).

Data selection is not something new: already at the time of bubble chambers experiments (Figure 16), event selection was needed, because of the relative rarity of the sought for events. With the data rate of LHC experiments, we need more sophisticated techniques though.

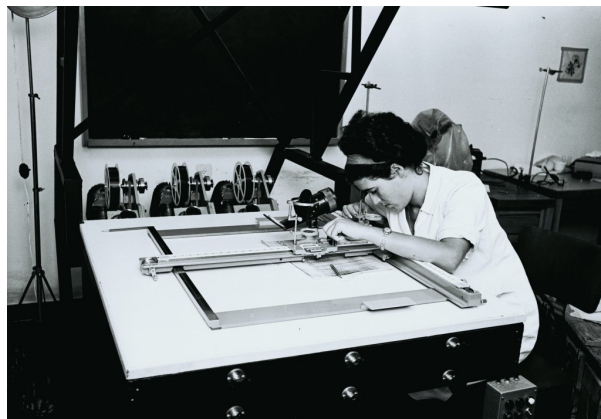


Figure 16. **Old fashioned data selection** - Back in the bubble chambers era, pictures used to be filtered manually. (Scanning, circa 1963. Image: Weizmann Institute Archives)

3.2 Data taking at CMS

Figure 17 illustrates how both the L1 and the HLT tightly combine with the Data AcQui-sition (DAQ) system (16): we cannot bypass the trigger system to take data, since events are

pushed through the acquisition workflow provided a L1 accept has been issued, and the HLT is part of this very workflow.

(39), (40) and (17) detail extensively the design of the CMS detector trigger system. Although some figures are outdated, the main concepts are still mostly valid.

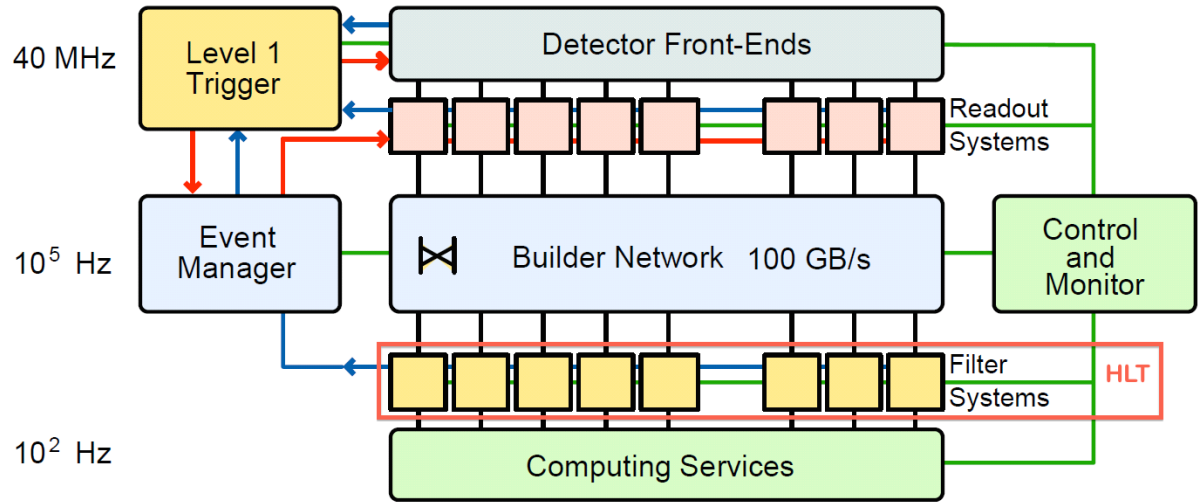


Figure 17. **CMS Data Acquisition (DAQ) system** - The HLT is a part of the DAQ, starting right after a L1 accept. Figures correspond nominal values. More realistic values are given in the text.

In practice, selecting or rejecting an event is a decision resulting from counting the number of objects above a given threshold in energy, possibly adding some kinematic requirements.

Possible objects are what we use in a physics analysis: electrons, muons, taus, photons, jets, missing energy - mostly.

The selection performed at the trigger level is often referred to as *online selection*, as opposite to the *offline selection*, done by analyses on recorded data. When the online selection rejects an event, it is definitive, there is no way to recover what might have been thrown away.

3.2.1 Level 1

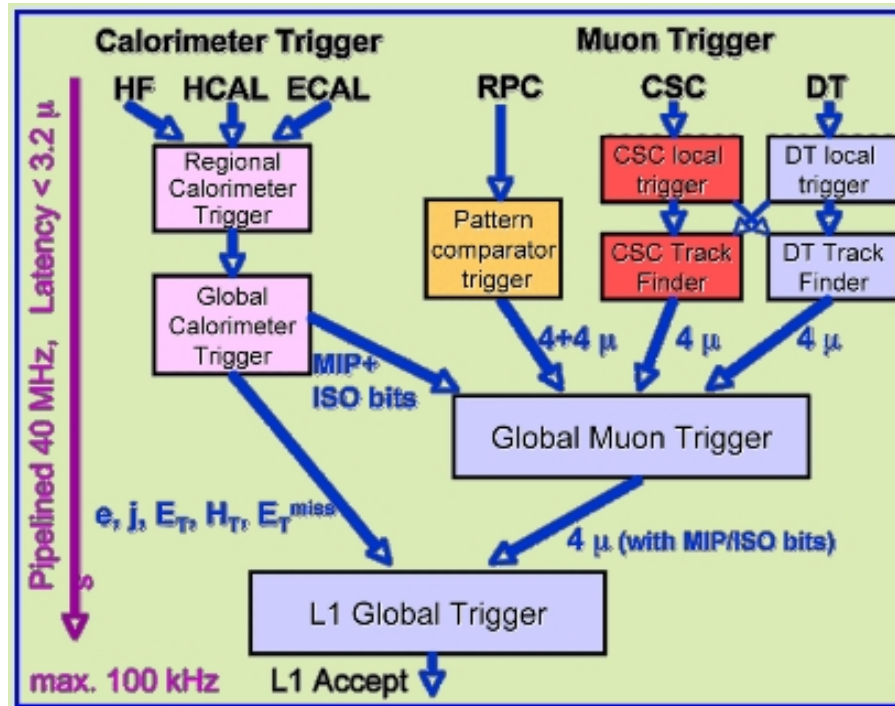


Figure 18. **Overview of the Level 1 trigger** - Two channels, relying on calorimetric and muon systems are used at L1. Calorimeters information is used to build electromagnetic and hadronic objects, along with energy (momentum) sums. Candidate muons are reconstructed by the muon system, possibly including isolation information obtained from the calorimeters. Muon & calorimetric systems objects are combined within the Global Trigger processor.

In the service cavern, two counting rooms host machines partly devoted to the Level 1. Hardware based, its duty consists in reducing the data rate from several megahertz down to about 100 kHz, within $3.2\mu\text{s}$. This latency allows to keep the event for some time into buffers, while building the objects needed to perform a first rough selection. If the decision to keep it comes back from the Level 1 machines, the event continues through the DAQ system, otherwise, it is discarded.

The architecture of Level 1 trigger is divided into two parallel channels (Figure 18): the muon system and the calorimetric system, eventually combined by the Global Trigger processor (GT). The tracking system is not used at Level 1. Within each channels, the reconstruction remains local: we do not use the full detector information at L1.

Only very simple arithmetic operations can be performed by the L1 system. Corrections may be incorporated through look-up tables.

3.2.1.1 Muon system

The muon system gathers signal from DTs, CSCs and RPCs to reconstruct up to four muon candidates, ranked by their p_t .

The muon trigger organization stems from the muon system structure: DTs are used for the barrel, CSCs in the endcap, supplemented by RPCs everywhere. Each of these subsystems reconstruct muon tracks, along with bunch crossing identification for the DTs and CSCs (slower detectors). The *Global Muon Trigger* (GMT) collects these tracks, correlates RPCs' tracks with DTs' or CSCs' tracks, determines isolation thanks to the calorimeters' information and select the four best candidates (best quality and hardest p_t) to transmit to the *Global Trigger* (GT).

3.2.1.2 Calorimetric system

The calorimetric system builds jets, taus & electron-photon candidates, using both calorimeters (ECAL & HCAL) information. It is implemented in different stages:

1. each individual calorimeter cell is used by the *Trigger Primitive Generators* (TPG) to compute local energy sums;
2. these are used by the *Regional Calorimeter Triggers* (RCT) to build candidate electrons, photons, taus, and jets. Isolation information can also be derived for electron/photon candidates by the RCT, as well as quiet regions to be transmitted to the Muon system to determine muon isolation information;
3. the *Global Calorimeter Trigger* (GCT) gathers RCTs' information (candidates and transverse energy sums), ranks the candidates by p_t and selects the four hardest of each kind to transmit them to the GT. The GCT also computes and transmits the total transverse energy (momentum) sum, along with the missing transverse momentum.

3.2.1.3 Global trigger

The Global Trigger (GT) processor collects muons and calorimetric candidates, synchronizes them, and feed them into up to 128 selection criteria, referred to as *Level 1 seeds*. These are logical combinations of GMT and GCT informations. Only events accepted by at least one of these L1 seeds may proceed to the next step.

The decision to keep or reject an event is transmitted to the DAQ system by the *Timing Trigger and Control* (TTC) system, after going through the *Trigger Throttling System* (TTS),

which allows to regulate the data flow from the L1 to the DAQ, including the filter farm (that is, the HLT). If there is a risk of overflowing the DAQ, prescales (3.3.1.2) can be applied or seeds might be turned off.

3.2.2 Data acquisition system

The DAQ system of the CMS detector, after a L1 accept, gathers the full detector information, reconstructs and filters it, before ensuring its storage for offline analysis.

Upon receiving a L1 accept from the TTC, the DAQ of the CMS detector (Figure 17) collects detector information from the Front End electronics with nearly 700 *Front-End Drivers* (FED). The data fragments, after being marked with headers and trailers for event identification, are pushed downstream by about 500 *Front-End Readout Links* (FRL), which transfer them to 64 *FED Builders*, switches that can dispatch data fragments to the following *Readout Units*. The Readout Units are grouped by eight and connected to eight *Readout Builders*, responsible for putting the full event together. After building full events, the Builder Units send them to *Filter Units*, that will preform the HLT selection.

A key feature of the DAQ lies in its eight slices running in parallel: one might be turn off if a problem arise, while preserving the data taking. Each slice can handle 12.5kHz.

The expected event size is 1MB, yielding a total throughput of 1GB/s. Yet, all along the DAQ line, *back-pressure* can be propagated to the GT to slow down the trigger frequency if the system is throttling.

3.2.3 High level trigger

The High Level Trigger, or HLT, is the second step of the online data selection. The *Event Filter* stage of the DAQ is performed by the *Filter Farm*, and serves several purposes listed below.

- Collecting events and perform consistency checks (full events are assembled by the Builder Units, filtering nodes should receive complete events to analyze);
- Running reconstruction algorithms as close as possible to the offline ones, but faster (because of higher/tighter thresholds, regional reconstruction or careful path design, see 3.3.1.4.2), to build electrons, photons, jets, taus, missing energy. More sophisticated quantities such as the supersymmetry searches variables α_T , M_T or the razor variables may also be computed. These are used to define selection criteria, called *trigger paths*. At the end of 2012, we had nearly five hundred paths running in parallel, serving hundreds of analyses. Since the whole detector information is available at the HLT, objects reconstructed are of offline quality: they could be as good as offline objects if the reconstruction time was longer, no information is missing compared to the local L1 reconstruction.
- Collect data quality information for the *Data Quality Monitoring* (DQM) system: during the data taking, we constantly monitor that the data collected looks relevant. For instance, if we look at the η distribution of electrons collected over thousands of events selected, we should observe a flat distribution, electron production is expected to be isotropic.
- Transfer the events selected to data storage areas.

The HLT reduces the data flow from 100kHz down to 350-450Hz, within an average time of 80ms in 2011, 160ms in 2012.

3.2.4 The Grid, or WLCG: Worldwide LHC Computing Grid.

Finally, the selected data is sent to CERN, to the *Tier0*, which is the first stage of data storage. Tier0 gathers all the *raw* data (unprocessed signals, but also trigger level objects) and reconstructs it, that is, builds objects for physics analyses. Thereafter, the data is transferred to secondary data storage areas called *Tier1*. Smaller data storage areas, called *Tier2* and *Tier3*, may have copies of the data as well, permanently or only short-term. To analyze the data, physicists may be writing their code anywhere in the world, after what they send it to run where the data is stored. This is the principle of the *grid*.

In practice, analysts tend to write a first step that makes basic selections on the data, to *skim* it and get back smaller samples with which they can work locally.

Such an organization was made mandatory by the amount of data produced. Indeed, even with the online data selection, we still collect too much data to store all of it in a single place: just a few months of data taking bring petabytes (10^{15}) of data. In comparison, a commercial hard drive is close to a few TB(10^{12}) at best. Besides, not only do we record lots of data, but also we simulate billions of events for various physics processes to understand our detectors, how signals and background (would) look like (2.4). Last, since our reconstruction software evolves continuously, we often keep several versions of similar samples. LHC's experiments rely on huge computing ressources, always pushing the limits further away.

A few points worth highlighting about the trigger system:

- two kinds of constraints have been pointed at: rate & timing;
- the rate constraints underline two bottlenecks:
 - 100kHz: this is the admitted limit on what the L1 should deliver to the HLT, based on the assumption that the average event size is about 1MB. In practice, the limit is the amount of data the filter farm can process. All collisions are not equal: when nothing happens but glancing collisions, there is not so much activity in the detector, events are small, we may be able to collect more than 100kHz of these. But we are not interested in these glancing collisions. We look for events depositing lots of energy in the detector: these are more complex, and require more processing time in the filter farm. These are the ones that may hit the limits of the computers running the HLT. During data taking, we carefully monitor the deadtime: if it goes too high (we tolerate up to 10%), this is a signal that the system is throttling;
 - 350-450Hz: the limit we set to the HLT output considers the amount of data that can be processed by the Tier0, based on reconstruction time (seconds, or, at worse, minutes);
 - the faster we are, the more approximate the objects we reconstruct. Tuning the trigger often boils down to trade-offs between accuracy and rapidity.

3.3 The trigger Studies Group: organizing the data taking for hundreds of analyses

None of them truly happy about having to make the definitive choices required by an online selection.

The Trigger Studies Group (whose organization is depicted on Figure 19) has to organize and operate the trigger system, supervise its development to serve hundreds of analyses, keeping the rates & timing under control, following LHC's evolution.

Between the start of the LHC in 2009 and the end of its first run in February 2013, many aspects evolved:

- the instantaneous luminosity started around $10^{30}\text{cm}^2\text{s}^{-1}$ in 2010 to reach $7.10^{33}\text{cm}^2\text{s}^{-1}$ at the end of 2012. The luminosity is linked to the number of events via: $N = L\sigma$. For a given process, the number of events (per second, or rate, if we use the instantaneous luminosity) should increase linearly with the luminosity. So, in principle, if we start with a trigger menu at 300Hz and do not touch it while the luminosity increases by a factor 10, we end up with 3kHz - putting aside other factors that may act on the rate. This is not sustainable, we have to tighten our selection criteria as the luminosity increases;
- the energy increased from 3.5TeV to 4TeV per beam;
- the number of *pile-up interactions* rose from just one or two up to thirty. When two bunches cross, if we are lucky, we may get a head-on collision between two protons, yielding some interesting physics process. But these happen amongst other soft interactions, or pile-up interactions, between protons, which add up some energy in the event. When

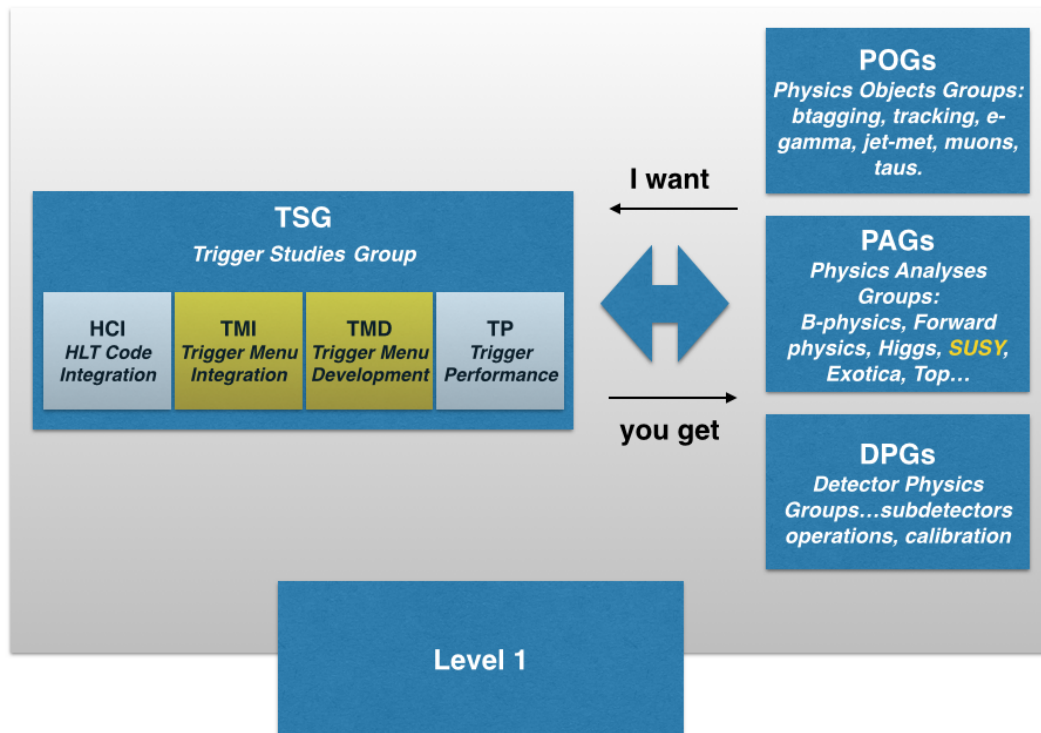


Figure 19. **A glimpse of the underlying organisation for trigger activities at CMS (Run I)** - While analysts are responsible for defining which L1 seeds & trigger paths they need, the TSG groups, HCI, TMI, TMD and TP provide technical support. Through trigger coordinators (one or two per POG, PAG), the TSG groups interact with POGs, PAGs and DPGs. They gather their requests and allocate the bandwidth, ensuring technical constraints are met while preserving the physics priorities, developing the system to accommodate LHC's evolutions. A few words about CMS organization, hinted at here: POGs, for Physics Objects Groups deal with the reconstruction of physics objects used in the analyses, such as jets, electrons, photons, muons, taus, b-tagged jets... PAGs refer to Physics Analyses Groups, each of them targeting a specific field of analyses, such as Higgs searches, B-physics, Forward physics, Heavy Ions physics, Supersymmetry analyses, Standard Model physics, Top physics, Exotica (everything else). *Yellow shades: I was involved in these activities; TMD: menu development, rate evaluation & monitoring; TMI: HLT on-call expert; SUSY: trigger coordinator for the SUSY group.*

deciding to record or not a collision, we need to be able to distinguish the output of the main interaction from the particles emitted in the pile-up ones to keep the rate under control.

To accommodate these changes (needed to increase our chances to produce new particles! although they do spice a bit the challenge), the trigger configuration must evolve. A trigger configuration is more commonly referred to as a *menu*.

3.3.1 Trigger menus

3.3.1.1 Foreword: data organization

When we record data, we need to organize the way we save and split it. The smallest unit in the recording is called a *lumisection*, which corresponds to 220 LHC orbits. During these, the instantaneous luminosity is considered as constant. A lumisection lasts approximately 93 seconds. Lumisections are gathered in a *run*. A run start defines a start of data taking. A run typically stops because an issue occurred or something in the detector configuration need to be changed. Ideally, when we have stable beams and collisions, we would like to have a run that starts just before we reach the stable beam mode and finishes after the beam dump - so as not to lose any beam data. Yet, we may have several runs during the same LHC *fill* (see 2.1). In between two runs, we continue taking data, mostly cosmics ray data, which is used for calibration.

3.3.1.2 L1 menu

A L1 menu consists in 128 L1 seeds meant for “physics” (that is, analyses), plus 63 technical seeds aimed at detector operations. Figure 20 shows the beginning of the list of L1 seeds for a

run taken during Run I. From this table, we may notice that we have several *prescale* columns. Indeed, while the list of L1 seeds running is uniquely defined for a given run, we may change the *prescale* column during a run. A *prescale* allows us not to accept every event satisfying the seed criteria, but one out of the prescale value. A very high value indicates that in the given column, the seed is in practice not active. The underlying motivation is that, during a fill (hence a run), the luminosity decreases (see Figure 21), and we can loosen our criteria accordingly. To do so without interrupting the data taking, and without reconfiguring the whole trigger menus, we just need to change the prescale column index on the fly. The change will become effective at the next lumisection.

3.3.1.3 HLT menu

An HLT menu (Figure 22) is built with hundreds of selection criteria or *paths*. In order to prepare the data sharing downstream, paths are gathered within *streams*, themselves subdivided into *datasets*. Paths meant for physics analyses end up in *stream A*, paths meant for alignment and detector calibration go to ALCA streams, etc... Within stream A, paths meant for single muonic analyses go to one dataset, *SingleMu*, paths for electronic analyses in *SingleElectron*: at the end of 2012, the HLT menu was collecting data split between thirty-five datasets such as:

- BTag: serving the B-tag POG (Physics Object Group), developing b-tagging methods;
- JetHT: single and double jets paths, HT (sum of jet pt), QCD analyses, but also SUSY & Exotica;
- DoubleMu: paths whose baseline selection requires two muons (plus some kinematic cuts);

Pre-scale Sets Details

n	L1 Algo Name	Index 0	Index 1	Index 2	Index 3	Index 4	Index 5	Index 6	Index 7	Index 8
0	L1_ZeroBias	9973	9973	9973	9973	9973	9973	9973	9973	262139
1	L1_AlwaysTrue	241	241	241	241	241	241	241	241	262139
2	L1_BeamGas_Hf_BptxPlusPostQuiet	10000	10000	10000	10000	10000	10000	10000	10000	262139
3	L1_DoubleJet20	1	1	1	1	1	1	1	1	1
4	L1_BeamGas_Hf_BptxMinusPostQuiet	10000	10000	10000	10000	10000	10000	10000	10000	262139
5		1	1	1	1	1	1	1	1	1
6	L1_IsoEG12er_ETM30	262139	262139	262139	262139	1	1	1	1	1
7	L1_DoubleEG6_HTT50	1	1	1	1	1	1	1	1	1
8	L1_BeamHalo	1	1	1	1	1	1	1	1	1
9	L1_TripleMu0	1	1	1	1	1	1	1	1	1
10	L1_Mu4_HTT125	1	1	1	1	1	1	1	1	1
11	L1_Mu3p5_EG12	1	1	1	1	1	1	1	1	1
12	L1_Mu12er_ETM20	1	1	1	1	1	1	1	1	1
13	L1_MuOpen_EG12	262139	262139	262139	262139	262139	262139	262139	1	1
14	L1_Mu12_EG7	1	1	1	1	1	1	1	1	1
15	L1_SingleJet16	200000	200000	200000	200000	150000	150000	150000	20	1
16	L1_SingleJet36	30000	6000	6000	6000	6000	6000	6000	10	1
17	L1_SingleJet52	7500	7500	7500	7500	7500	7500	7500	200	1
18	L1_SingleJet68	600	300	300	300	100	100	100	4	1
19	L1_SingleJet92	60	60	60	60	40	40	40	1	1
20	L1_SingleJet128	1	1	1	1	1	1	1	1	1
21	L1_DoubleEG6_HTT100	262139	262139	262139	262139	1	1	1	1	1
22	L1_DoubleEG6_HTT125	1	1	1	1	1	1	1	1	1
23	L1_Mu5_DoubleEG5	262139	262139	262139	262139	262139	262139	262139	1	1
24	L1_DoubleMu3p5_EG5	262139	262139	262139	262139	262139	262139	262139	1	1
25	L1_DoubleMu5_EG5	1	1	1	1	1	1	1	1	1
26	L1_DoubleMu0er_HighQ	262139	262139	262139	1	1	1	1	1	262139
27	L1_Mu5_DoubleEG6	1	1	1	1	1	1	1	1	1
28	L1_DoubleJetC44_ETM30	262139	1	1	1	1	1	1	1	1
29	L1_Mu3_JetC16_WdEtaPhi2	300	300	300	300	300	300	300	80	1
30	L1_Mu3_JetC52_WdEtaPhi2	10	10	10	10	10	10	10	1	1
31	L1_SingleEG7	800	800	800	800	800	50	20	400	1
32	L1_SingleIsoEG20er	1	1	1	1	1	1	1	1	1
33	L1_EG22_ForJet24	1	1	1	1	1	1	1	1	1
34	L1_EG22_ForJet32	1	1	1	1	1	1	1	1	1
35	L1_DoubleJetC44_Eta1p74_WdEta4	6	6	6	6	6	6	6	3	1
36	L1_DoubleJetC56_Eta1p74_WdEta4	1	1	1	1	1	1	1	1	1
37	L1_DoubleTauJet44er	1	1	1	1	1	1	1	1	1
38	L1_DoubleEG_13_7	1	1	1	1	1	1	1	1	1
39	L1_TripleEG_12_7_5	1	1	1	1	1	1	1	1	1
40	L1_HTT125	262139	262139	262139	262139	262139	1	1	262139	1
41	L1_DoubleJetC52	262139	262139	262139	262139	262139	262139	262139	262139	1
42	L1_SingleMu14er	1	1	1	1	1	1	1	1	262139
43	L1_SingleIsoEG18er	262139	262139	262139	1	1	1	1	1	1
44	L1_DoubleMu_10_Open	262139	262139	262139	1	1	1	1	1	1
45	L1_DoubleMu_10_3p5	1	1	1	1	1	1	1	1	1
46	L1_ETT80	262139	262139	262139	262139	262139	262139	262139	3	1
47	L1_SingleEG5	4500	4500	4500	4500	4000	500	200	3000	1

Figure 20. L1 menu as visualized in WBM (*Web Based Monitoring tool, which gathers information about online operations*) - There are up to 128 physics L1 seed in the L1 menu. In the beginning of the table, shown here, we can see single objects seeds with different energy thresholds (L1.SingleJetXX, L1.SingleEGXX), or multiple objects seeds, mixing leptonic and hadronic requirements for instance. The various columns show the different sets of prescales that can be used. Following column 0 which corresponds to high luminosity, the columns are ordered by decreasing expected luminosity. Seeds not in bold font are implemented but masked (rows 3, 5, 7, 8): they are not used in practice.

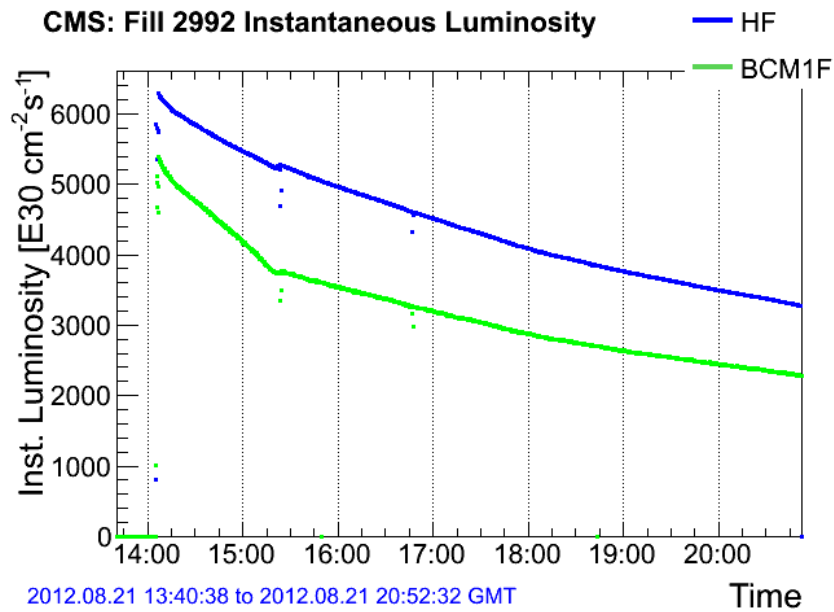


Figure 21. **Luminosity decrease during an LHC fill** - We can observe nearly a factor two decrease over the fill in this case. This is due to emittance growth and beam intensity decrease over the time.

- MuOnia: serving B-physics;
- + 31 more datasets.

Datasets are defined taking into account the rate they gather: in order to be handled by storage areas (and analysts' code), they should not be too big.

Each time a menu is updated we must check that the overall rate & timing fit within the budget, possibly adjusting the dataset content.

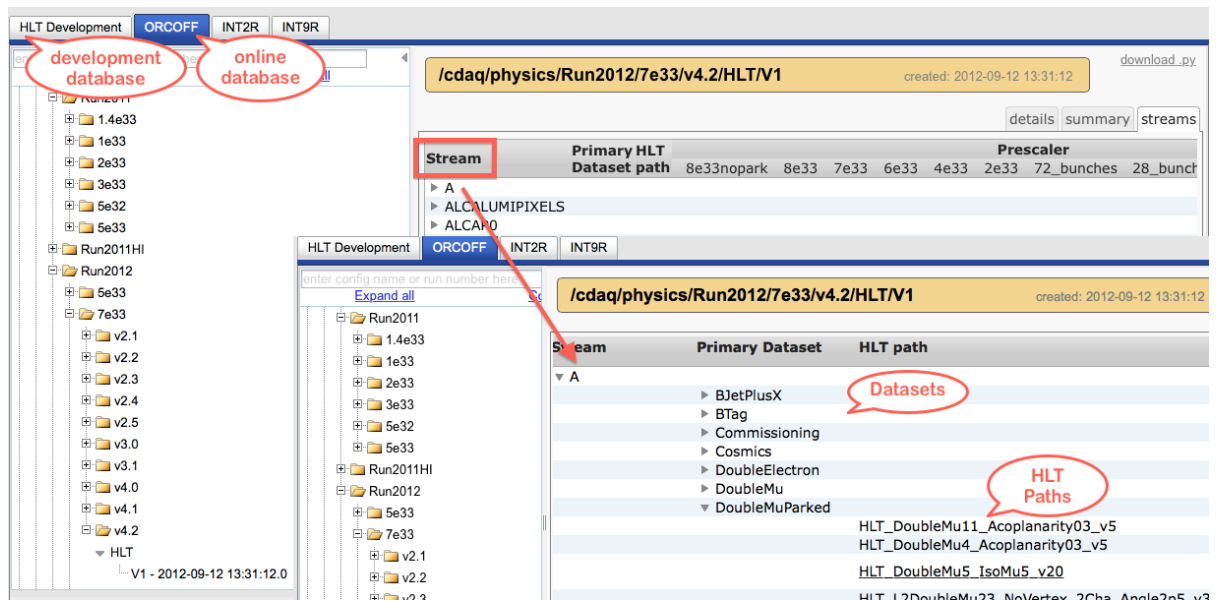


Figure 22. **Visualizing an HLT menu using the database browser** - The database browser is one of the key tool for the HLT development and operation. As we can see, we have different databases; the main ones are the development one (offline menu preparation), and the online one. Within the databases, menus are organized in subdirectories following the year of use, the mode of operations (proton-proton physics, heavy ions, special menus), the maximal luminosity expected for the menu, and different versions. Within a menu, we have different streams, divided into datasets containing the trigger paths. As for the Level 1, we have several prescale columns to adjust to varying luminosity.

3.3.1.4 Menu development: rate & timing

Each analysis is responsible for developing the trigger paths it wishes to implement. A trigger study proposing a new path should present:

- physics motivation;
- efficiency;
- rate & timing.

In principle, a trigger request encapsulates three paths: the main path desired, a backup one (tighter, in case the rate of the main one goes too high), and a monitoring path, to evaluate the performance of the main path.

3.3.1.4.1 Rate evaluation

In 2011 and 2012, as the conditions required tighter and tighter selection criteria, rate evaluation was mostly done using data.

Choosing a good run for rate evaluation

In the data already recorded, we chose a run that is smooth enough (no issue with subdetectors, no interruption, no misconfigurations), long enough (to have enough events to test our menu), rather recent (conditions not so different from what we are preparing for).

The OpenHLT analyzer during Run I

We have two options to study the rate of a new path or, even, of the whole menu: either we run the HLT code¹, or we use *OpenHLT* trees. The principle of this tool consists(ed) in running the

¹This works only if we have access to the data in RAW format - the format read by the filter farm.

HLT code on a reference run raw data, storing objects quantities typically used in cuts in a flat ROOT(41) tree. For example, when running the code producing HLT muons, we store their p_t , η , ϕ , energy...The result of existing paths (at the time of the reference run) is also saved.

Thus, we can distinguish two cases:

- the path already existed in the previous menu, we just need to apply a scale factor corresponding to the ratio of the new luminosity divided by the luminosity of the reference run to obtain the new rate for the menu in preparation;
- the path is new, but tighter than what was previously in the menu, e.g., we want HLT_Mu40 (selecting events containing HLT muons with $p_t > 40\text{GeV}$) while we used to have HLT_Mu30: we write an emulation of the new path, using the flat tree. Of course, if our data contains only muon events with muons above $p_t > 30\text{GeV}$, we cannot use it to evaluate the rate of, for instance, HLT_Mu20.

The pure rate of each path is not sufficient: when designing a menu, we need to take into account the *overlap*. Indeed, paths are not exclusive, and two paths with different criteria may trigger the same event: it is not impossible to have one muon above 40GeV and two muons above 12GeV plus one jet above 30GeV! As a result, to determine the overall rate of a dataset, of a menu (where the limits are), we need to be able to take into account the overlap between paths. The OpenHLT tool allows us to do so.

3.3.1.4.2 Controlling the timing

A path consists in a series of filters, as can be seen on Figure 23. Each filter reduces the amount of data transferred to the next one. This is how we control the average time needed to

run the whole menu: not only do we apply faster algorithms (tighter thresholds for instance), but also we try to sequence the steps such that we keep the time-consuming ones for the end, to avoid unnecessary computing time consumption. It can happen that an event takes more than a 160ms to process, but most events will be treated much faster.

An HLT path may be viewed as a succession of steps, or *modules*. To begin with, we start by requiring that a specific (set of) L1 seed(s) selected the event (hadronic seeds for hadronic paths, EG seeds for electrons and photons, muon seeds for muon paths). Then, for the given example, we first build calorimeter jets, filter on $calo-H_T$ (the sum of calorimeter jets' p_t) before building *particle flow* (2.3.1) jets, and finally filter on the corresponding $PF-H_T$ (the sum of particle flow jets' p_t). This sequence allows to reduce the input rate at each step, hence control the time needed to run the full path. First, only events fired by some specific L1 seed(s) will result in the computation of calorimeter jets: instead of feeding this module with 100kHz, we send only a few kHz. Then, the input to the particle flow jets builder is even more reduced by the pre-filtering on calorimeter jets: particle flow jets are more suitable objects to use to control the rate and target a better purity(3.3.2.2), but they take longer to build.

Like for the rate, there is some overlap between paths in terms of timing: several paths may trigger jet reconstruction for instance, but the jet reconstruction will be done only once per event - this is a good reason to have a given event processed by a single filtering node. Thus, the tools allowing us to evaluate the timing should take it into account. Besides, while we are interested in the average time needed to run the path, which had to be less than 10ms in 2012, we also look at the extra time needed when we add it to a full menu.



Figure 23. **Structure of an HLT path** - An HLT path begins with a L1 seed requirement: if the desired seed did not *fired*, the following steps are not performed. If it did, for the given example, we first build the rather fast calorimeter jets, filter on the corresponding H_T (sum of jets' p_t), before moving to the Particle Flow objects, more accurate but time consuming. To preserve the efficiency we would obtain without the calorimeter pre-filtering, the cut on the calorimeter based H_T is set lower than the cut on the PF- H_T .

3.3.2 Trigger paths for supersymmetry searches

Supersymmetry is one of the key topics studied at the LHC. About one fourth to one third of the trigger bandwidth is used by searches for supersymmetry. A similar bandwidth is typically used by Higgs searches. Exotica (all other searches beyond the Standard Model) is also a major bandwidth consumer. Overlaps exist between these groups' requests, as well as with smaller consumers such as top quark analyses, B-physics, Standard Model physics: trigger coordination for supersymmetry searches also requires coordination with other groups' choices & requests.

Supersymmetry signatures present a huge variety: decay products can be hadronic objects like jets, missing transverse energy, leptons, photons, as we can see on Figure 24.

Originally, searches for supersymmetry were organized following the expected final states: *all hadronic* (jets and missing energy) searches ((42), (43), (44), (45), (46)), *single lepton* (electron, muon, tau - (47), (48), (24)), *di-lepton* ((49), (50), (51)), *multilepton* (at least three, (52)), searches with photons (53). Each analysis may present different requests. Initially, they could use rather simple paths requiring a lot of activity in the event, energetic leptons. Yet, as the luminosity increased, paths had to become more selective. One of the issues of the decay such as the one shown on Figure 24 comes from the fact that, while we have many objects, these may be rather soft. Thus, raising *thresholds* (energy cuts), is not such a good option. Rather, as supersymmetry kept evading us, since the luminosity & number of pile-up interactions kept increasing, we resorted more and more to paths playing with high multiplicity of objects and low energy thresholds. We also resorted to variables used in analyses for background rejection, such as α_T (54), M_{T2} (44), razor variables (55). The diversity of supersymmetry signatures is

highlighted on Figure 25 which shows how the bandwidth requested by supersymmetry searches was split at the end of 2011.

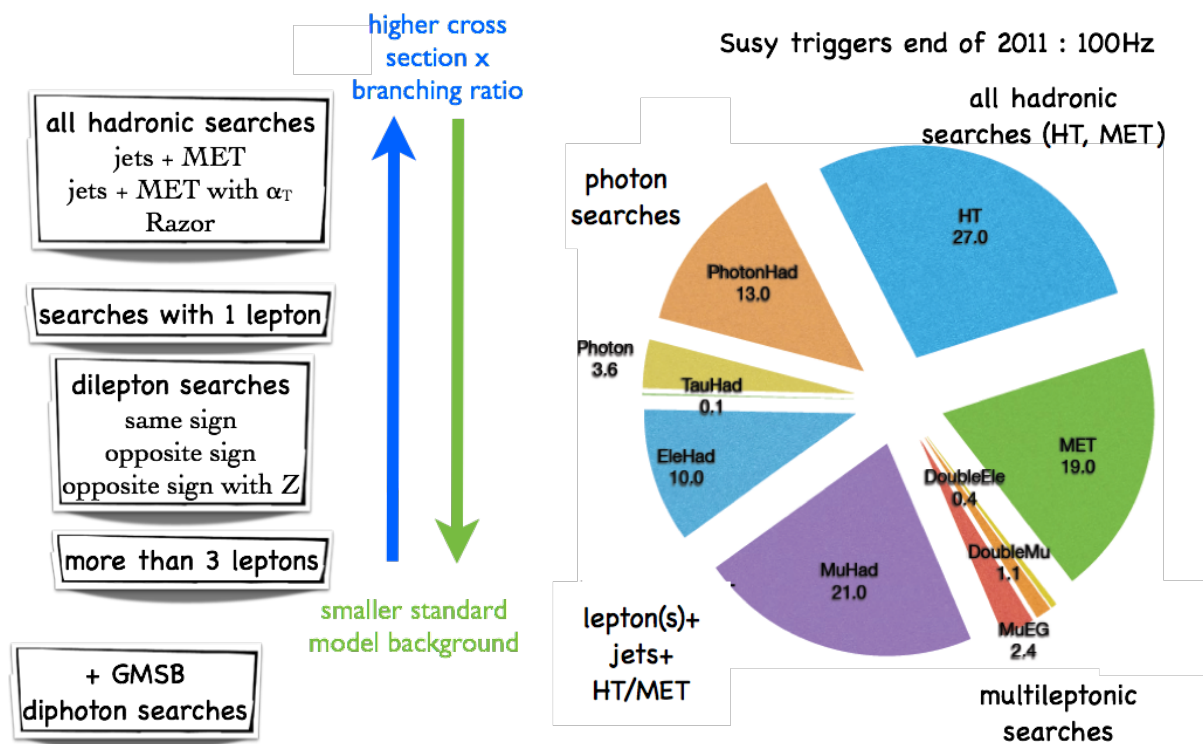


Figure 25. **Sharing of the bandwidth requested by supersymmetric searches in 2011** - The diversity of the analyses is reflected by the variety of the paths: hadronic, leptonic, combination of both. Opposite strategies for searching for supersymmetry are underlined here: one the one hand, hadronic decays have a large cross section, but have to face huge background yields, while, on the other hand, leptonic signatures can be cleaner but suffer from low cross sections.

3.3.2.1 Addressing the increasing number of pile-up interactions at HLT

The main issue encountered by paths for supersymmetry searches across Run I arose from the increase in the number of pile-up interactions. Indeed, supersymmetry searches focusing on hadronic activity had to adapt their selections so as to ignore the energy coming from those secondary interactions, to consider only the main one.

Pile-up interactions may contribute in several ways, as can be seen on Figure 26 and Figure 27:

- adding some new particles, hence some extra energy to the jets from the main interaction;
- adding some new particles which are going to be clustered in extra jets;
- modify the set of jets from the main interaction.

During summer 2011, the number of pile-up interactions increased from an average of 5-6 ($L = 2.3 \times 10^{32} \text{cm.s}^{-2}$) per bunch crossing to 14-15 ($L = 3.2 \times 10^{33} \text{cm.s}^{-2}$), with tails up to 30. Using the relation $R = L\sigma$, where R is the rate, L the instantaneous luminosity and σ the cross section, we can define a cross-section for a trigger path as:

$$\sigma = \frac{R}{L}. \quad (3.1)$$

When plotted as a function of luminosity, this cross section should be flat: the rate increases linearly with the instantaneous luminosity for a well behaved trigger path. What we could observe for some hadronic paths such as the ones combining momentum sums H_T and $M_{\vec{H}T}$ (vectorial sum of jet $\vec{p}_t$) was that, on the contrary, the cross section was increasing more and

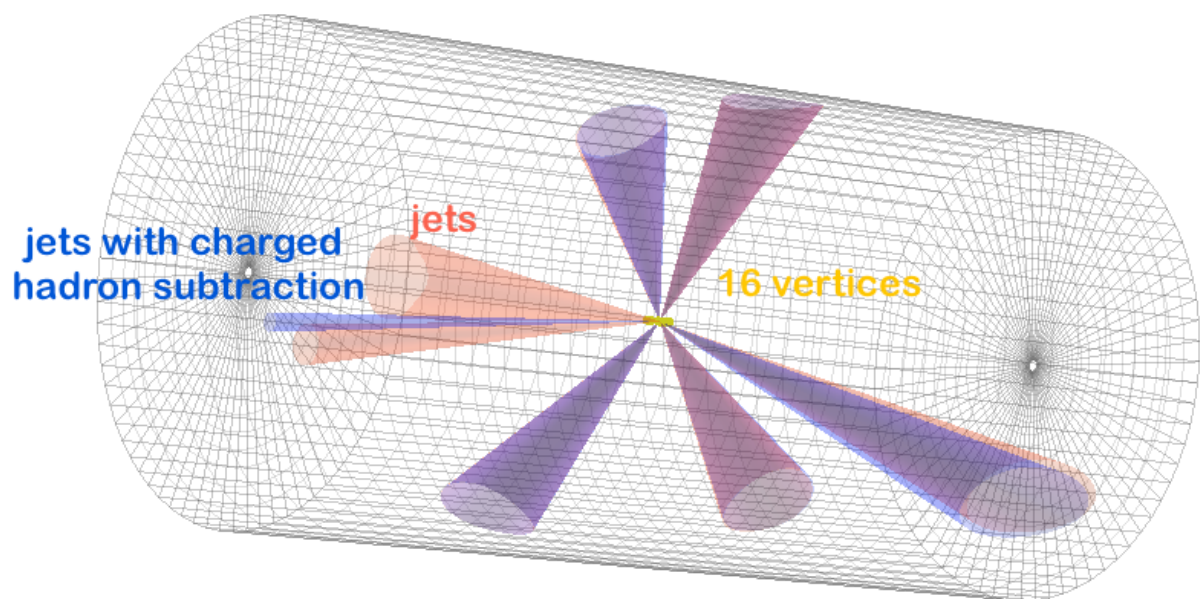


Figure 26. **Effect of pile-up interactions on jets -1** - In this event, we had 16 interactions. In reddish shades, jets which have been clustered with no pile-up subtraction, in blue, jets clustered after removing particles identified as coming from secondary vertices. Jets displayed here have a $p_t > 20\text{GeV}$. Most of the jets overlap, but we can see also two reddish jets near only one blue. Particles from pile-up interactions may affect the jet clustering.

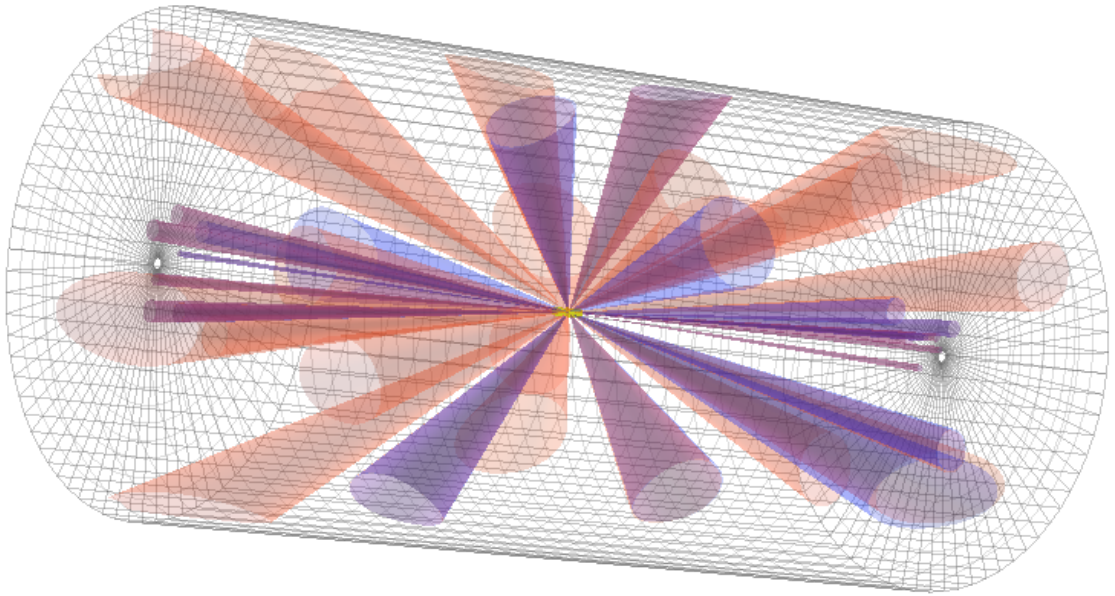


Figure 27. **Effect of pile-up interactions on jets -2** - Same as the previous figure, except that the cut on the jets' p_t has been lowered down to 10GeV: pile-up interactions tend to add soft jets.

more (Figure 28). Hence, the rate was blowing up, threatening the data taking in the extent that too much data was sent to the Tier0.

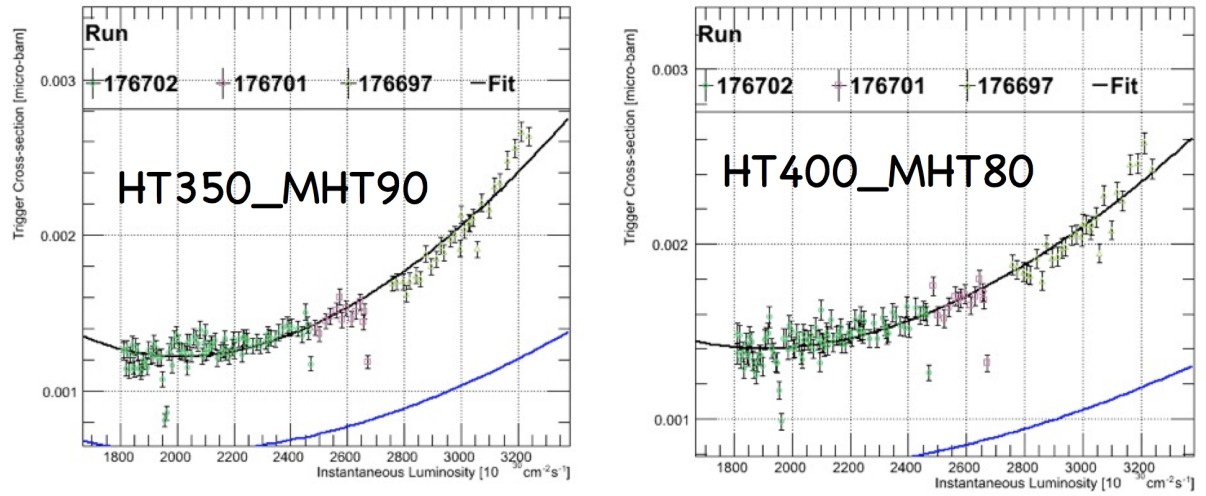


Figure 28. **A pathologic trigger path cross section versus luminosity** - The cross section should be flat. The rate for these paths is blowing up because of the increase in the number of pile-up interactions.

A quick and easy solution for the well-being of the trigger system would have been to increase the thresholds. However, increasing these thresholds at the trigger level means that the analyses cuts need to be raised as well, implying a loss of sensitivity for the searches. So, from a physics point of view, we needed a better solution, taking into account what was happening: extra interactions, not interesting for us, artificially increasing the energy sums.

The first solution implemented was the use of pile-up corrections, already used in analyses, providing a jet by jet energy correction. The principle of these correction, developed in (56), consisted in evaluating the pile-up energy density through:

$$\rho = \text{median}_{\text{all } k_t \text{ jets}} \frac{p_t}{\text{area}} \quad (3.2)$$

and correct each jet p_t :

$$p_{t,\text{corr}} = p_t - \rho \cdot \text{area}(\text{jet}). \quad (3.3)$$

The area of *anti- k_t* jets is constant, while the area of k_t jets may vary: the latter are more greedy and provide a better estimation of the pile-up energy density (for thorough studies about jet areas, please refer to (57)).

Two checks had to be performed before adopting these corrections:

- they actually help reducing the rate & stabilize the cross section;
- they do not harm the path efficiency.

Regarding the first point, the use of these pile-up corrections turned out to be very efficient as the plots from Figure 29 shows. The cross-section flattens and the rate decreased by up to a factor two.

In the meanwhile, the efficiency was not affected.

Measuring a path efficiency

To assess the efficiency of a trigger path, we resort to *efficiency* or *turn-on curves*. Consider a path selecting events if their H_T is greater than 400GeV -we would name it HLT_HT400. We

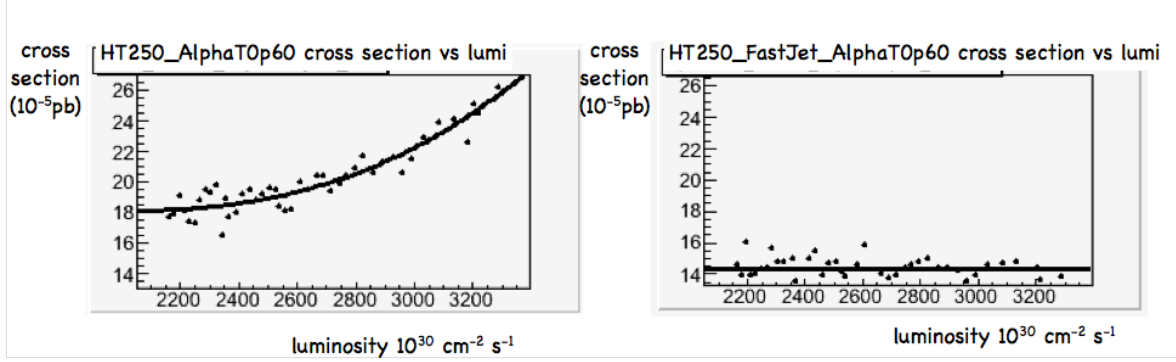


Figure 29. **Effect on pile-up corrections on the cross section of paths relying on H_T and α_T** - α_T is a variable used by searches for supersymmetry in order to discriminate between supersymmetry and QCD background: QCD events mostly have $\alpha_T < 0.5$, supersymmetric events extend towards $\alpha_T > 0.5$ (54). With the use of pile-up corrections (right plot), the cross section flattens, which means that the path comes back under control.

know that, because of time constraints, the online H_T does not match perfectly the offline one, and we want to know as of which offline H_T value our path is fully efficient. We chose an unbiased sample - typically, an orthogonal sample, selected with cuts which have nothing to do with H_T . For instance, we pick up a `SingleMu` dataset sample, which we further filter by requiring that events passed the lowest unrescaled single muon path, `HLT.Mu40`. This is our *denominator* cut. We feed events satisfying this cut into an histogram storing the number of events as a function of H_T . Among these events, we select those also passing our trigger path: this is our *numerator* cut. They feed another histogram storing the number of events as a function of H_T . Dividing the second histogram by the first one, we obtain a *turn-on* curve as

displayed on Figure 30. Such a plot tells us that a path cutting on H_T at 400GeV will be fully efficient as of 450GeV with respect to the offline H_T .

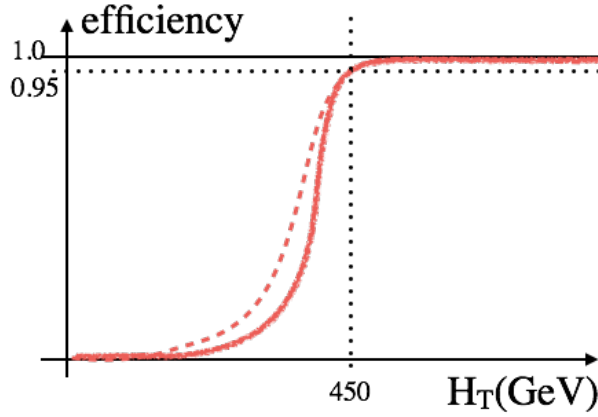


Figure 30. **A *turn-on curve* or efficiency plot for a path cutting at 400GeV on online H_T , HLT_HT400** - This kind of plots tells us that our trigger path (full line) is 95% efficient for an offline H_T reaching 450GeV. The dashed line show a path with a similar efficiency but a slower turn-on curve, such as what we could observe for a path not using the pile-up corrections, which would be collecting more events at low H_T .

If the efficiency was not affected, it is mainly due to the fact that the offline H_T already used these pile-up corrections. Applying them online resulted in a better match between online and offline objects, the turn-on curve sharpened but the efficiency threshold was not affected.

3.3.2.2 Towards better jets

Turn-on may also be observed for other paths, such as jets' paths: a path selecting events with an online jet p_t above 300GeV will only catch all events above 350GeV. An analysis relying

on this path will have to put a cut on the leading jet p_t at 350GeV (or would have to take into account a non-fully efficient trigger, which is not commonly done at CMS). This means that the events collected by this path having a leading jet with an offline p_t smaller than 350GeV will be thrown away. In the trigger realm, this is waste of bandwidth. Thus, narrowing the offline-online gap is a key point to improve our bandwidth usage and the purity of the events we collect.

Purity

The efficiency of a path is a key information for physics studies. For the trigger group, the purity is quite important as well, although not always so easy to define. In principle, it sounds simple: given the number of events collected by your path, which percentage was actually used by the analysis? Yet, if you really consider it in these terms, once you apply a full analysis selection, probably very few events are kept, the purity is low, which is intuitively not good. Indeed, the cuts applied online are much simpler, so as not to discard too many events because of the not so good online reconstruction. A final selection may look like (see 4.3.2):

- single mu trigger path;
- exactly one tight muon, no electron, isolated track veto;
- four jets above 30GeV;
- $R^2 > 0.15$, $MR > 350\text{GeV}$;
- ...

Of course, we cannot expect a path asking for one muon to collect only the events that will end up in our analysis. A more realistic way to define the purity consists in looking at the percentage of event collected above the efficiency threshold out of all of the events collected. Then, the slower the turn-on curve, the lower the purity.

Bringing online objects closer to the offline ones is not only a matching game: offline objects should be the closest we can get to the actual physics particles and their properties. In that extent, a unique feature of the CMS reconstruction software during Run I was provided by the *particle flow*, or *global event description* workflow (2.3.1).

Quite new at the beginning of Run I, *particle flow* jets become the most commonly used in offline analyses.

Introducing pile-up energy corrections online already helped reducing, the online-offline gap. Figure 31 shows three turn-on curves: for calorimeter jets with the pile-up energy corrections discussed in the former paragraph, for *particle flow jets* and for particle flow jets with the particle flow derived pile-up energy corrections, as a function of offline jets' p_t . The latter particle flow jets have an additional piece of pile-up energy correction: the particle flow particles, if they are charged, can be associated to a vertex thanks to their track. Thus, if they point to a vertex that is not the main one¹, we can discard them. The key information from Figure 31 lies in the sharpness of the blue curve, with respect to the other ones: with online, pile-up corrected, particle flow jets, the efficiency plateau comes earlier, and we collect fewer events towards low

¹The main vertex is taken as the one for which the sum of the track p_t , for tracks pointing to the vertex, is maximum.

jet p_t . We control the rate, with up to 30% rate reduction, and we improve the purity of our paths.

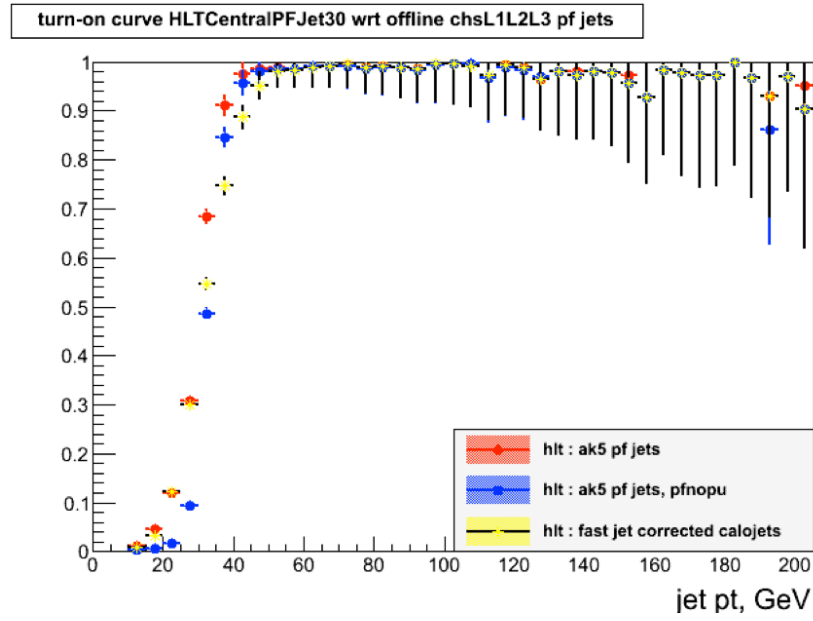


Figure 31. **Turn-on curves for different set of online jets requiring an online $p_t > 30\text{GeV}$, as a function of offline jet p_t , obtained on $Z \rightarrow \mu\mu$ events.** - The sharper the turn-on, the better the purity, the less waste of bandwidth: pile-up corrected particle flow jets win.

3.3.2.3 Addressing pile-up interactions at L1

Even at L1 we had to cope with pile-up interactions: the rate of hadronic seeds relying on jets or H_T was getting too high and sending too much data to the filter farm.

At L1, the available information, and the possibilities to play with it are more limited. A jet at L1 resembles what Figure 32 shows: a 3x3 set of calorimeter regions, the central cell carrying a higher p_t than the surrounding ones.

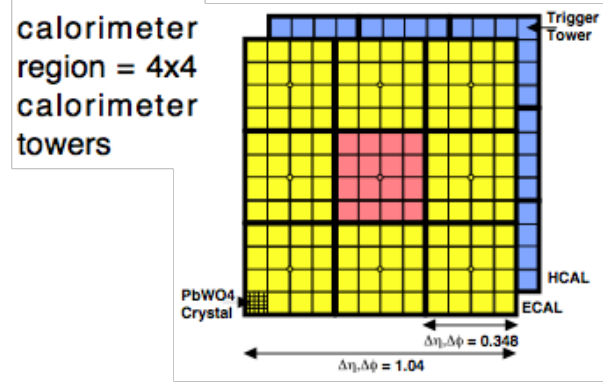


Figure 32. **A L1 jet in the $\eta - \phi$ plane** - A L1 jet consists in 3x3 calorimeter regions: each region contains 4x4 calorimeter towers, corresponding, for the barrel, to one HCAL tower combined with a 5x5 array of ECAL crystals. This is considered as a L1 jet if the central cell has a higher energy than the surrounding ones.

Just like for the HLT paths, the cross section of L1 seeds relying on hadronic activity, such as `L1_HTT`, cutting on a L1 H_T , sum of L1 jets p_t , became more and more non-linear as the number of pile-up interactions increased, as shown on Figure 33.

At L1, we cannot apply the same strategies we use at HLT. The workaround proposed consisted in adding an extra requirement on the jets: not only should the central tower region p_t be higher than the p_t of the surrounding ones, but, also, it should be greater than a given

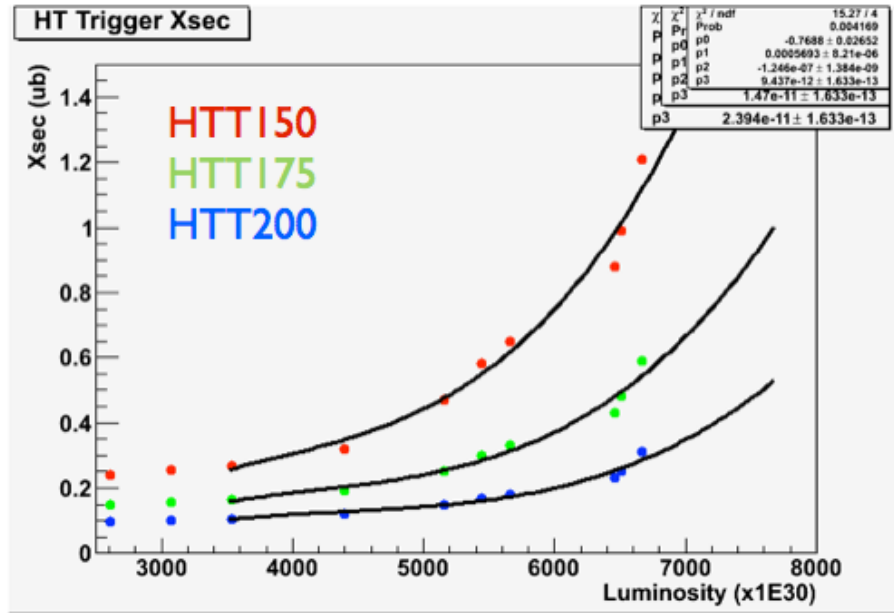


Figure 33. Cross section of L1 seeds cutting on H_T , as a function of luminosity - The cross sections are clearly not flat, the corresponding rate is blowing up because of the increase of the number of pile-up interactions at high luminosity.

threshold. In practice, this threshold was set to 5GeV: allowing a rate reduction possibly as high as 40%, this barely affected the efficiency plateau, as can be seen on Figure 34¹.

3.3.3 Selected tales of trigger activities

This chapter highlighted various aspects of trigger activities during Run I, from the tools used to the organization, including some of the essential concepts and issues encountered. I started in the trigger group with rate evaluation to assess the viability of the whole menu, which got me not only to learn the tools, but also *build a feeling* about online operations. Moving

¹The GCT is where the threshold may be applied - hence the name *GCT setting* in Figure 34.

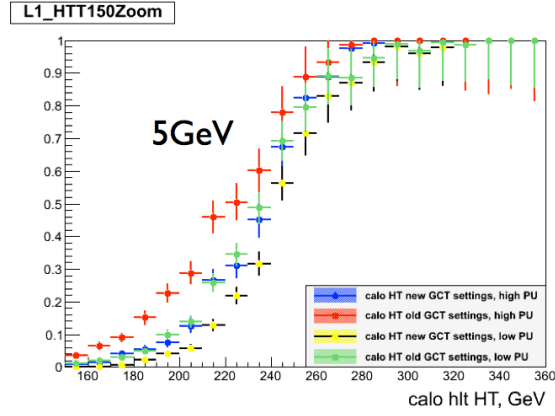


Figure 34. **Efficiency (or turn-on) curves for L1_HTT150, for different GCT settings, in high and low pile-up conditions** - The goal of the threshold requirement for the central region of the L1 jets is to be less sensitive to the extra energy from pile-up interactions.

Indeed, these tend to add low energy jets, while the main interaction should give hard jets. Yet, we want to ensure that we do not lose events if the number of pile-up interactions, which varies across a fill, decreases. Thus, we plot the efficiency curves for two different conditions: high number of pile-up interactions (red and blue curves), low number of pile-up interactions (green and yellow curves), with and without the extra threshold requirement. We also want to check that the efficiency plateau is not affected. This plots confirms that the plateau efficiency remains similar to the no threshold situation, while the turn-on curve sharpens: fewer events are collected at low H_T , this is where we save some bandwidth.

to trigger coordination for supersymmetric searches and on-call expert tasks, built up on this technical knowhow and further required to learn as much as possible about the system and its ramifications, to anticipate, “guide”...coordinate. It is not about inventing wheels. It is about making them run.

To conclude, a few daily life examples, underlining that:

- when designing a trigger, the efficiency to collect a signal need to be thought of taking into account the amount of background in the corresponding phase space;
- the rate overlap game is interesting, but cannot buy everything;
- some paths may provide information about the current detector status.



Figure 35. **Selected tales - trigger world** - Know the people. Know your tools. Listen(eavesdrop), read, sort out, organize, anticipate...Keep common sense.



Figure 36. **Selected tales - trigger world** - Know the people. Know your tools. Listen(eavesdrop), read, sort out, organize, anticipate...Keep common sense.

CHAPTER 4

SEARCH FOR THIRD GENERATION SPARTICLES WITH THE RAZOR VARIABLES

4.1 Introduction

This chapter develops a search for supersymmetry focusing on direct stop production in the proton-proton collisions provided by the LHC, at a center of mass energy of 8TeV. The stops, scalar partners of the Standard Model top quarks, are assumed to be pair-produced and decay to top quarks plus LSPs. The guiding model used to set up this analysis is the simplified model known as T2tt, which has two unknown parameters: the stop and LSP masses (see 1.2.5).

The LSP, which can correspond to the MSSM neutrino $\tilde{\chi}_1^0$, is a stable and weakly interacting particle that escapes the detector unseen, thus yielding events with missing transverse momentum. The top quark decay may lead to different signatures depending on the decay channel of the W boson:

- *all hadronic*, both W bosons decay hadronically - 44.5%;
- *semi-leptonic*, one of the W decay yields a lepton (electron, muon or tau, the latter not being considered in the analysis) - 44,5%;
- *di-leptonic*, both W decays yield leptons - 11%, not considered in the analysis.

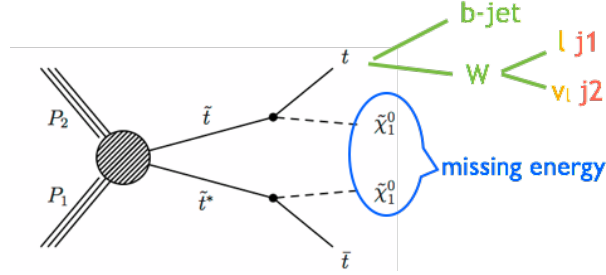


Figure 37. **The T2tt simplified model** - The T2tt simplified model is used to design the analysis: two stops, scalar partners of the Standard Model top quarks, decaying to tops plus LSP, $\tilde{\chi}_1^0$. The tops undergo decays to b quarks and W bosons, themselves decaying into hadrons or leptons. The LSPs are stable, they do not decay and escape the detector unseen.

The analysis targets different channels by splitting events into distinct categories or *boxes*: the all-hadronic channel with a *BJet* box, and two semi-leptonic channels, either one electron or one muon, with a *Mu* box and an *Ele* box.

4.2 The razor variables

Choosing your glasses: playing with frames

4.2.1 Resonances, supersymmetry event

4.2.1.1 Bump search

Over the past century, many new particles have been discovered through *bump* searches. Collecting the decay products, one can reconstruct the invariant mass of a (hypothetical) parent particle by adding their four momenta and squaring the result. This procedure typically leads to a picture like Figure 39. Each new particle appears as a *bump*, or *resonance*, at the corresponding invariant mass. Recalling the corresponding Feynman diagram (Figure 38), the cross section for resonance production follows a Breit-Wigner distribution:

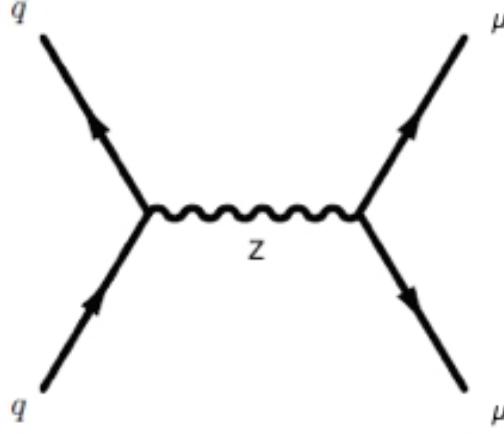


Figure 38. **Resonance production and decay diagram** - Example of the Z boson. The Z boson can be produced as a resonance, thereafter decaying to muons.

$$\sigma(E) \propto \frac{\Gamma^2/4}{E^2 - E_0^2 + \Gamma^2/4}, \quad (4.1)$$

where E is the energy in the center of mass frame of reference, E_0 is the mass (or rest energy) of the resonance and Γ its width (28). Thus, the cross section is peaking at the mass of the resonance. For a two-body decay, denoting $\vec{p}_1$ and $\vec{p}_2$ the momenta of the decay products, assumed massless, and ϕ the angle between them, the mass of the resonance is given by:

$$M = \sqrt{2|\vec{p}_1||\vec{p}_2|(1 - \cos \phi)} \quad (4.2)$$

When particles, like the W boson for instance, decay to a visible particle and an invisible one (lepton plus neutrino in the case of the W boson), we miss the neutrino and cannot reconstruct properly the invariant mass of the mother particle, since we do not know the longitudinal component of the neutrino

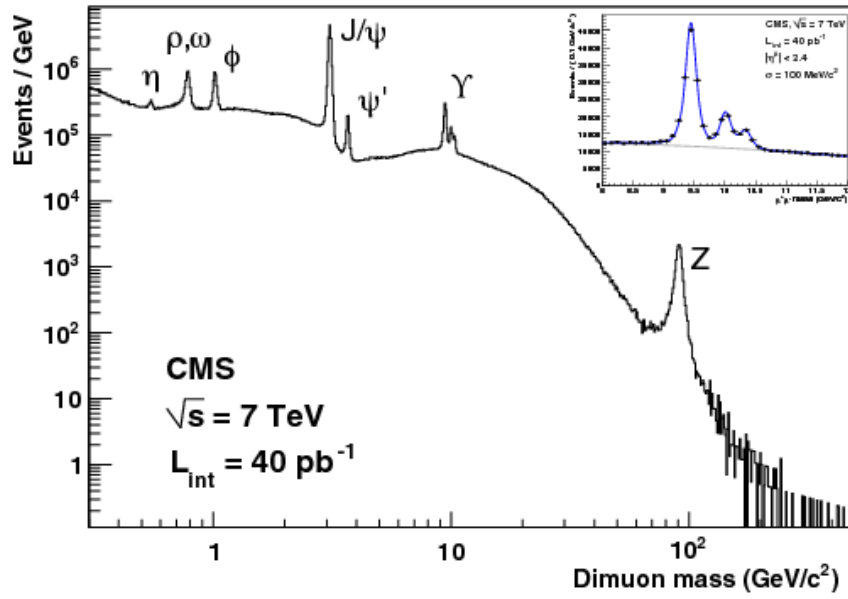


Figure 39. **Di-muon invariant mass spectrum at the LHC** - For events with two muons, one computes their invariant mass and obtain this spectrum. Each peak corresponds to an intermediate particle state. In proton-proton collisions, the amount of energy involved varies from one interaction to another. The first peak appears at the η mass of ≈ 547 GeV; if the energy is smaller than that, no particle decaying to two muons is observed. The next peak that can be seen corresponds to the ρ and ω mesons. The more energy provided in the collision, the heavier the particles that can be produced and observed, such as the J/ψ or Z .
CMS Collaboration.

momentum. However, since we know that the total momentum should balance in the transverse plane, one can define the *transverse mass*:

$$M_T^W = \sqrt{2p_l^T p_\nu^T (1 - \cos \phi)} \quad (4.3)$$

where ϕ is the angle between the lepton and the neutrino, and p_l^T and p_ν^T are the transverse momenta of the lepton and neutrino respectively (the latter being reconstructed as missing transverse momentum, see 2.3.4). Instead of having a distribution that peaks at the particle's mass, the transverse mass distribution will show an end point at the mass of the mother particle, since, in the case where the transverse momenta equal the three-dimensional momenta, Equation 4.3 and Equation 4.2 are similar.

4.2.1.2 Supersymmetry events: an underconstrained system and no bump

In a typical supersymmetric event, such as the one depicted in Figure 40, one ends up with at least two particles, the LSPs, that escape the detector without leaving any signal. These two particles lead to a single missing transverse momentum vector, and one does not know how to split it into two components. As a result, it is not possible to fully reconstruct the event to observe the masses of the squarks $\tilde{q}$ as *resonances* or *bumps* using Equation 1.1¹.

¹This topology may look very similar to dileptonic $t\bar{t}$ decays. Yet, some extra complications prevent us from using techniques that have been developed for top mass determination in $t\bar{t}$ dileptonic decays. For instance, not only do we ignore how to split the missing transverse momentum between the two LSPs, but also we ignore both stops and LSP masses. Thus, a method (e.g. (58)) scanning top masses to find the best compatibility with the observed missing momentum arising from the two weakly interacting particles is not applicable. We have more unconstrained variables: the squarks and LSP masses, each LSP four-momenta. Variables sensitive to the mass of the new particles through an edge point, such as M_{T2} may be used, as in (44). We pursue a different strategy here.

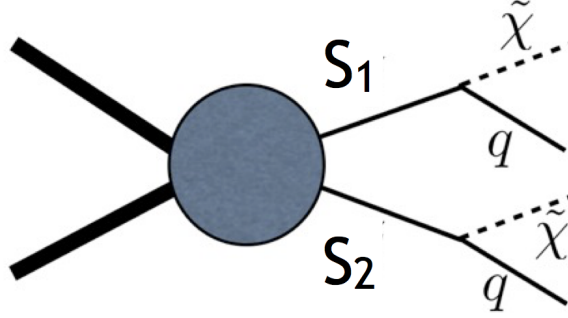


Figure 40. **A *canonical* supersymmetry event** - In the interaction, two squarks S_i are produced, each decaying to a visible (quark, hence jet) part and an invisible (LSP) part.

4.2.2 Introducing the razor variables

We introduce two new kinematic variables, the *razor* variables, which will allow us to formulate a search for new heavy resonances as an effective bump search. The following paragraphs rely on the event depicted in Figure 40.

4.2.2.1 In the squark's $S_{i=1,2}$ rest frames

In Figure 40, each squark S_i decays to a visible part, typically a quark q_i yielding a jet, and an invisible part, an LSP $\tilde{\chi}_i$, yielding missing transverse momentum. In each squark's rest frame, we can apply four momentum conservation:

$$\mathbf{P}_{S_i} = \mathbf{P}_{q_i} + \mathbf{P}_{\chi_i} \quad (4.4)$$

where the four momentum vector $\mathbf{P}$ is defined to have energy E and three momentum vector $\vec{p}$ such that $\mathbf{P} = (E, \vec{p})$. From which we obtain, assuming massless quarks:

$$E_{q_i} = |\vec{p}_{q_i}| = \frac{M_q^2 - M_{\chi}^2}{2M_q} = \frac{M_{\Delta}}{2}. \quad (4.5)$$

Thus, the energy of the visible parts of the decay can be expressed as half of M_Δ , which, rather than providing us with the masses of the new sparticles, gives us a characteristic scale for new physics. In the case of large mass splitting between the squarks and the LSP, M_Δ is probing M_S directly. On the contrary, if the mass splitting is small, M_Δ is probing the mass splitting itself.

Unfortunately, we do not have a direct access to the squarks' rest frames: we can only measure quantities in the laboratory frame.

4.2.2.2 From the squarks' rest frames to the razor frame

A key feature of each of the squarks' rest frames worth pointing out is the following: in these frames, both quarks (or jets) carry the same momenta, equal to half of the characteristic mass scale we are looking for. Hence, a natural question follows: can we find or approximate such a frame, starting from laboratory frame quantities? A series of approximations helps to do so: first, we assume that sparticles are produced at rest, so that either of the squarks' rest frames and their center of mass frame are identical. Second, we assume that the boost from the laboratory frame to the squarks center of mass frame is only longitudinal along the beam, or z -axis: $\vec{\beta}_L = \beta_z \hat{z}$. The different frames and their relationships are illustrated in Figure 41. Then, requiring that in our *rough approximation* or *razor* frame the magnitude of the momenta of the two massless jets $\vec{q}_{i=1,2}$ are equal yields¹:

$$|\vec{q}_1^R| = |\vec{q}_2^R| \quad (4.6)$$

$$\Rightarrow \gamma(E_1^l - \beta_z q_{1,z}^l) = \gamma(E_2^l - \beta_z q_{2,z}^l) \quad (4.7)$$

$$\Rightarrow \beta_z = \frac{E_1^l - E_2^l}{q_{1,z}^l - q_{2,z}^l}. \quad (4.8)$$

¹Superscripts R denote quantities in the razor frame, superscripts l indicate quantities in the laboratory frame. Arrows indicate three vectors, and E is the energy.

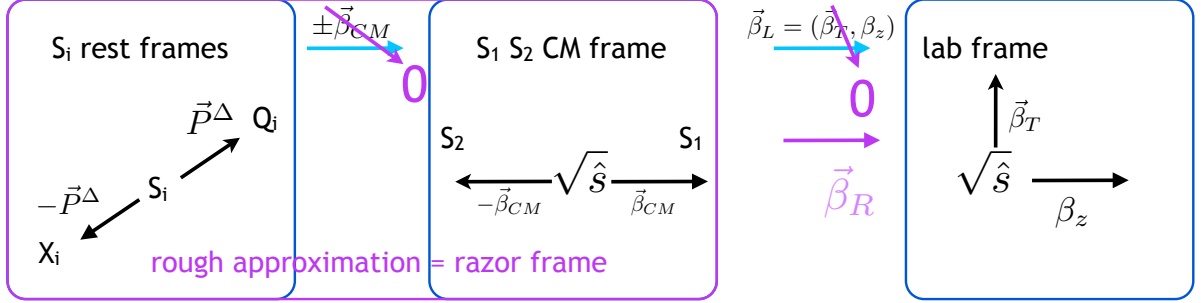


Figure 41. **Frames used in the M_R derivation** - The razor variable M_R is an approximation of the characteristic scale M_Δ . From left to right: the two squarks rest frames, the squarks center-of-mass frames, the laboratory frame. Assuming squarks are produced at rest, the squarks rest frames and their center-of-mass frame are identical, defining the *rough approximation* or *razor frame*. We also assume that the boost between the laboratory frame and the razor frame is only longitudinal.

Having derived the boost that allows us to move from the laboratory frame to the razor frame, we can now express $|\vec{q}_i^R|$ as a function of laboratory quantities, and derive an approximation M_R for the characteristic scale M_Δ :

$$M_R = 2|\vec{q}_i^R| = 2\sqrt{\frac{(E_1^l q_{2,z}^l - E_2^l q_{1,z}^l)^2}{(q_{1,z}^l - q_{2,z}^l)^2 - (E_1^l - E_2^l)^2}} \quad (4.9)$$

This mass variable M_R , which is our approximation for the new physics scale M_Δ , will display a peaking behavior with a maximum close to M_Δ , as shown in Figure 42.

Yet, M_R alone is not enough: the background processes may result in huge yields, making it impossible to distinguish a peaking feature due to a faint new signal over the background processes. A key example is provided by QCD di-jet production, resulting in two back-to-back jets. For this final state, $M_R = \sqrt{\hat{s}}$, which is of the order of 100 GeV - 1 TeV in the LHC collisions: this is precisely the

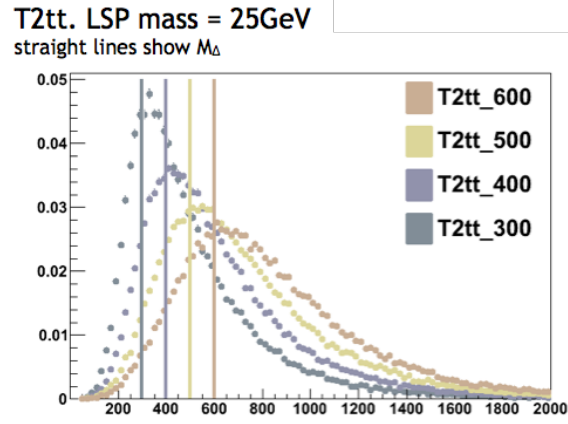


Figure 42. **Distribution of M_R , for different stop masses** - The T2tt simplified model follows the canonical topology displayed in Figure 40, with squarks being stops. For this plot, the LSP mass is fixed to 25 GeV, while the stops masses range from 300 to 600 GeV. All of those distributions exhibit a peak, close to the M_Δ values and hence M_R allows us to probe a possible scale of new physics through a *bump* search. On the contrary, for Standard Model backgrounds, M_R shows a falling distribution, since either the Standard Model particles produced in the proton-proton collision are much lighter, or the decay chain does not match the canonical topology from Figure 40.

mass range where we are searching for new physics. To be able to distinguish a peaking signal out of this background, we need to further reject these events. To do so, we build another variable, R , or R^2 , sensitive to the momentum imbalance in the transverse plane. For signal events, we expect to have some imbalance due to the LSPs, while for background events, only mis-measurements can provide a non-zero missing transverse momentum. To construct R , we formulate the following mass quantity using the missing transverse momentum and the transverse momenta of the quarks (or jets) in the laboratory frame:

$$M_R^T = \sqrt{\frac{E_T^{miss} (q_T^{j1,l} + q_T^{j2,l}) - \vec{E}_T^{miss} \cdot (\vec{q}_T^{j1,l} + \vec{q}_T^{j2,l})}{2}} \quad (4.10)$$

Just like the transverse mass defined in Equation 1.1, which has an endpoint value at the parent particle mass, M_R^T will show an endpoint at our characteristic scale M_Δ for signal processes that satisfy the canonical event in Figure 40:

$$\begin{aligned} (M_R^T)^2 &\leq \frac{E_T^{miss} (|\vec{q}_T^{j1,l}| + |\vec{q}_T^{j2,l}|) + \vec{E}_T^{miss} \cdot (\vec{q}_T^{j1,l} + \vec{q}_T^{j2,l})}{2} \\ &\leq E_T^{miss} (|\vec{q}_T^{j1,l}| + |\vec{q}_T^{j2,l}|) \end{aligned}$$

with quarks emitted in the same direction, opposite to the LSPs' direction;

$$\leq (|\vec{q}_T^{\chi_{1,l}}| + |\vec{q}_T^{\chi_{2,l}}|) (|\vec{q}_T^{j1,l}| + |\vec{q}_T^{j2,l}|)$$

for signal events where the invisible parts are the $\tilde{\chi}_{i=1,2}$

$$\leq (|\vec{q}^{\chi_{1,l}}| + |\vec{q}^{\chi_{2,l}}|) (|\vec{q}^{j1,l}| + |\vec{q}^{j2,l}|)$$

if everything happens in the transverse plane

$$\leq M_\Delta^2 \text{ if } \beta_z = 0.$$

Hence, this additional mass quantity is also a variable that is sensitive to the scale of new physics, M_Δ , relying only on quantities in the transverse plane. We now define the ratio:

$$R = \frac{M_R^T}{M_R}. \quad (4.11)$$

For events with no real missing transverse momentum, R , or R^2 , will concentrate at low values, whereas for signal events, since it is the ratio of two quantities probing the same scale, it will extend up to 1. Thus, R , or R^2 , may be used to discriminate further between signal and background events (59).

Further refinements...

The boost from the laboratory frame to the razor frame $\vec{\beta}_L$, derived previously, sometimes encounters unphysical behavior: it can be greater than one. We may solve this issue by removing one of the simplifying assumptions made earlier: the squarks might not be produced at rest, which implies that the squarks' rest frames and their center-of-mass frame may be different. This affects the expressions of the boosts and M_R , but the behavior remains the same: we look for a peaking signal near a characteristic mass scale over a falling background. In this case, M_R becomes (Details in (60)):

$$M_R = \sqrt{\left(|\vec{q}_1^l| + |\vec{q}_2^l|\right)^2 - (q_{1z}^l + q_{2z}^l)^2}. \quad (4.12)$$

The remainder of this work uses M_R as given above in Equation 4.12.

4.3 Analysis

The present analysis follows from a series of CMS analyses developed with the razor variables: the 7 TeV inclusive (55) and multijet (61) analyses, as well as the 8 TeV inclusive razor analysis (62). This work is similar to the 7 TeV multijet analysis, but uses data at a higher energy. This work differs from the 8 TeV inclusive analysis mainly by its lower cuts on the razor variables, especially R^2 (Figure 43).

To achieve a lower cut on R^2 , which is motivated by a desire to reject the QCD background, we rely on

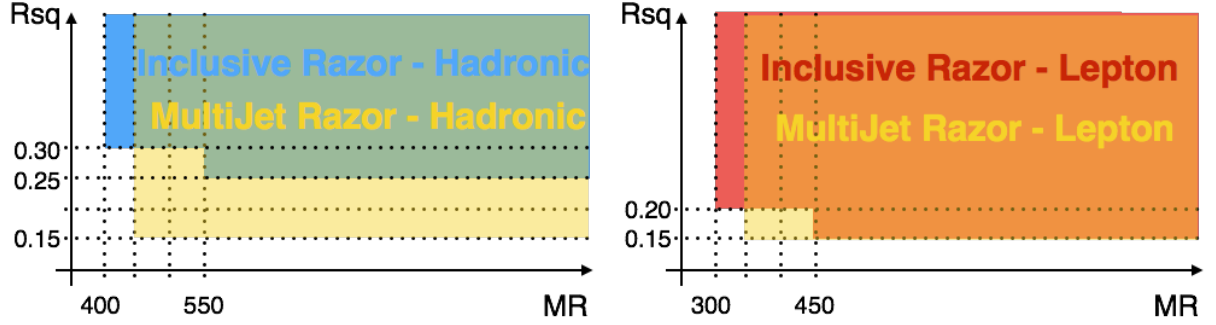


Figure 43. **Kinematic cuts applied on the razor variables for the inclusive and multijet razor analyses.** - The multijet analysis, developed in this chapter, extends towards lower R^2 and M_R values compared to the inclusive analysis. The leptonic boxes cuts are shown on the left plot and the hadronic box is displayed on the right. We can reach lower cuts by relying on different trigger paths, but in doing so, we accept more background events. To counter this, we increase the jet multiplicity requirements.

a higher jet multiplicity requirements, as detailed in 4.3.2.

4.3.1 A shape analysis

Our search for pair-produced top quark spartners is designed as a shape analysis. The principle of the analysis consists in modeling the backgrounds with a falling functional form. This functional form is tested and adjusted using known background samples, made with simulated events, following the inclusive analysis (62). This functional form is motivated empirically, by observing that the SM backgrounds distributions for the razor variables fall exponentially in Monte Carlo simulated samples.

The functional form used to model the Standard Model backgrounds is the following:

$$P_{SM}(M_R, R) = \left[b (M_R - M_R^0)^{1/n} (R^2 - R^{2,0})^{1/n} - 1 \right] \exp \left(-bn (M_R - M_R^0)^{1/n} (R^2 - R^{2,0})^{1/n} \right). \quad (4.13)$$

where M_R^0 , $R^{2,0}$, b , n are free parameters which are determined by an unbinned maximum likelihood fit (using ROOFIT (63)) to the data across the full R^2 , M_R plane. Equation 4.13 has the following properties:

- upon integrating over one of the razor variables, either R^2 or M_R , this functional form corresponds to a one dimensional exponential in the other razor variable, whose tail is characterized by M_R^0 or $R^{2,0}$;
- b and n describe the slope of the 1D projections along R^2 and M_R ;
- n specifically characterizes the concavity of the tail: for $n = 1$, it is flat, for $n > 1$, it is concave, for $n < 1$, it is convex.

We first test the fit model on known Monte Carlo samples, before attempting a data fit. To evaluate the fit performance, we build *French flag* plots in the following way:

- an initial fit is performed to a reference simulated sample; we then generate alternate background models, or fit functions, by varying the fit parameters according to the covariance matrix from the reference fit;
- we produce about 3000 sets of pseudo-data following these alternate forms;
- we divide the R^2 , M_R plane into cells and we build, for each cell, a histogram with the corresponding number of events that the pseudo data sets yielded (Figure 44); we smooth each histogram to obtain a probability density function $P(n; R^2, M_R)$ for the number of events in each cell;

- for each histogram, hence, each R^2, M_R cell, we compute the p -value to observe the number of events or more in the the reference sample as follows:

$$p(R^2, M_R) = \int_{obs}^{\infty} P(n; R^2, M_R) dn \quad (4.14)$$

- finally, we report the p -values of the Standard Model background-only fit to the reference sample for each cell in the R^2, M_R plane, both in terms of number of standard deviations and color-coded such that shades of red represent the degree of an excess and shades of blue represent the degree of a deficit (Figure 45).

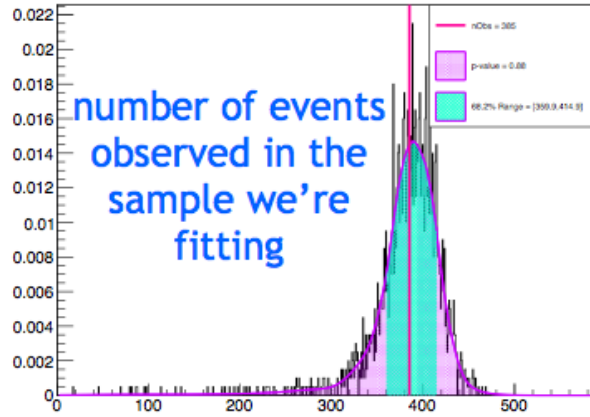


Figure 44. **Distribution of the number of events per toy in a representative R^2, M_R cell** - We generate samples of pseudo-data using alternate with forms, obtained by varying the fit parameters according to the covariance matrix. Then, dividing the $R^2 - M_R$ plane into cells, we can fill, for each cell, a histogram with the number of events observed for this cell.

We can then compute the p -value corresponding to our initial sample yield.

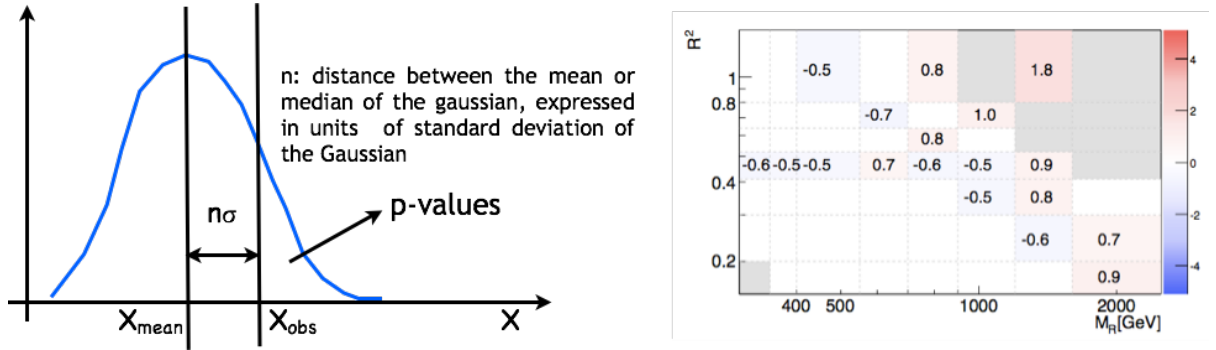


Figure 45. **Building *French Flag* plots** - The p-values are translated into a number of standard deviation, and reported in a $R^2 - M_R$ plot. Thus, this plot represents, both in color shades and numbers, deficits (negative values, blue shades) and excess (positive values, red shades) between the observed yields and the fit predictions.

If the deviations are small, and we do not see a cluster of deep red or blue cells, the background model provided by the fit works well.

4.3.2 Splitting into boxes and analysis selection

As indicated earlier, the analysis is set up so as to focus on direct stop production, described by the T2tt simplified model.

To begin with, we need to cast events into the canonical topology shown in Figure 40. Indeed, we do not directly see the top quarks from the stops decays, we collect their decay products. To meet the canonical topology features, we cluster particles into two megajets, or hemispheres, by trying all possible combinations of jets and selecting the one such that the sum of the squares of the two megajets masses is minimized, as recommended in (64). Afterwards, we compute the razor variables replacing the quarks' momenta with the corresponding hemisphere quantities in the expressions derived previously (Equation 4.12).

4.3.2.1 Standard Model backgrounds

For leptonic (Ele and Mu) boxes, the final state of a T2tt event contains four jets (including two jets arising from b quarks), one lepton, missing transverse momentum. The Standard Model background processes are listed below.

- *QCD processes*

Multijet production arises from the production of quarks or gluons followed by parton showering and hadronization of the products. Mismeasurements of the jets energy can result in fake missing transverse momentum. In principle, the lepton requirement will remove most of the QCD background. Yet, if a jet is misidentified as a lepton, some QCD events could be misinterpreted as signal. This process is difficult to fully simulate, since the cross sections can be huge, as can be seen from table Table I. The QCD samples (labelled as QCD-HT100To250, QCD-HT250To500, QCD-HT500To1000, QCD-HT1000ToInf in Table I) are binned in HT , the scalar sum of jets' p_T , in GeV.

- *Drell-Yann processes (DY sample)*

The (possibly virtual) production of a Z boson or photon that decay to oppositely charged leptons may constitute a background, if we misidentify one lepton and the events come along with jets and missing transverse momentum, which can arise from experimental mismeasurements.

- *WJets*

Leptonic decays of W bosons produce missing momentum and a lepton. If such an event is produced with other jets, including a b-tagged jet, this background can resemble our signal.

- *Single top*

Top quarks decay to b quarks and a W boson, which itself can decay either to a lepton and missing momentum or to two jets. Single top events have at least one leg in common with our signal. If the

top decays leptonically and other jets are produced in the event, or if the top decay hadronically but there is a lepton (real or fake) plus missing momentum, possibly from mismeasurements, single top events may be similar to our signal.

- *$t\bar{t}$ ($TTJets$ sample)*

The production of top-quark pairs (often referred to as $t\bar{t}$) constitutes our main background because the decay topology is similar to our signal events, yielding two top decays. However, this background is reducible because signal events will typically contain more missing momentum due to the presence of the two LSPs.

For the hadronic (BJet) box, the final state of a $T2tt$ event consists of six jets, including two jets arising from b quarks, and missing transverse momentum.

- *QCD processes*

Multijet production arises from the production of quarks or gluons followed by parton showering and hadronization of the products. Mismeasurements of the jets energy can result in fake missing transverse momentum. This process is difficult to fully simulate, since the cross sections can be huge, as can be seen from table Table I. For hadronic boxes, we do not benefit from the presence of a lepton to reject this background.

- *Drell-Yann processes (DY sample)*

The (possibly virtual) production of a Z boson or photon that decay to neutrinos may constitute a background, if the events contains additional jets, with one of them being either mis-tagged as b-jet or possibly a real b-jet.

- *$WJets$*

Leptonic decays of W bosons produce missing momentum and a lepton. If the lepton is misidentified and such an event is produced with other jets, including a b-tagged jet, this background

can resemble our signal. Further hadronic decays of W bosons produce jets, and possibly missing transverse momentum from experimental mismeasurements. If additional jets are present, this background can again resemble our signal.

- *Single top (T_tW and $T_{\bar{t}}tW$ samples)*

Top quarks decay to b quarks and a W boson, which itself can decay either to a lepton and missing momentum or to two jets. Single top events have at least one leg in common with our signal. If the top decays hadronically and other jets are produced in the event, or if the top decays leptonically but the lepton is misidentified with other jets produced in the event, then single top events may be similar to our signal.

- *$t\bar{t}$ ($TTJets$ sample)*

The production of top-quark pairs constitutes our main background because the decay topology is similar to our signal events, yielding two top decays. However, this background is reducible because signal events will typically contain more missing momentum due to the presence of the two LSPs.

Simulated samples for each of the backgrounds described above were centrally produced for the whole CMS collaboration. Table I summarizes the cross sections, number of events generated and equivalent luminosity for each of these samples.

4.3.2.2 Leptonic boxes analysis selection

The analysis selection for the leptonic boxes consists of:

- a single lepton trigger path (see 3.2.3).

HLT_IsoMu24 is used for the Mu box, seeded with L1_IsoMu17, it requires one muon with a $p_t > 24$ GeV for the Mu box, and is fully efficient for offline muons with a $p_t > 25$ GeV;

Process	Cross section(pb)	Number of events	Equivalent luminosity (pb ⁻¹)	Weight
QCD-HT100To250	1.04e+07	50129518	4.82	4004.03
QCD-HT250To500	276000.	27062078	98.05	196.83
QCD-HT500To1000	8426.	30599292	3631.53	5.31
QCD-HT1000ToInf	204.	13843863	67862.07	0.28
DY	2950.	30459503	10325.26	1.87
WJets	30400.	57709905	1898.35	10.17
T_tW	10.7	497658	46510.09	0.41
Tbar_tW	10.7	493460	46117.76	0.42
TTJets	245.8	6923750	28168.23	0.69

TABLE I

SIMULATED (MC) SAMPLES, CROSS SECTIONS, NUMBER OF EVENTS GENERATED, RESULTING WEIGHT TO MATCH 19.3FB⁻¹ OF DATA.

HLT_Ele27_WP80 is used for the Ele box, seeded with L1_EG22, it requires one electron with a $p_t > 27$ GeV for the Ele box, and is fully efficient for electrons with an offline $p_t > 30$ GeV;

- at least four jets with $p_t > 30$ GeV, $|\eta| < 2.4$;
- at least one CSVM b-tagged jet (CSVM: CSV algorithm, with the medium working point);
- exactly one tight lepton of the box flavor, exactly one loose lepton of the box flavor, no lepton of the other flavor, and an isolated track veto to reject taus. To ensure full trigger efficiency, muons are required to have a $p_t > 25$ GeV, while electrons must have a $p_t > 30$ GeV;
- $M_R > 350$ GeV, $R^2 > 0.15$.

The leptonic requirements target essentially box categorization. Asking for four jets including one b-tagged jet rejects W +jets background. The jet multiplicity criterium and the analysis cuts on the razor variables are set so as to further reject the QCD background (which is already small thanks to the leptonic selection). The value of the cut on M_R should be high enough so as to avoid the peak near the top mass ($M_\Delta = M_{\text{top}} = 176$ GeV for top events, see Figure 46). In practice, these kinematic

requirements are also set such that the resulting sample can be fitted with the functional form given previously. Table II and Table III illustrate the effects of each cut, for background processes and an example T2tt mass point. In Table II and Table III, the huge QCD yields are seen to be greatly reduced by the initial kinematic cuts on the razor variables of $M_R > 300$ GeV and $R^2 > 0.05$; these initial cuts are there for practical reasons and keep the size of the data samples to a manageable level. The leptonic selection also shrinks these yields significantly. The jet multiplicity and b-tagged jet requirement bring down the W +jets contribution. The final kinematic cuts on M_R and R^2 remove events with low real missing momentum (W +jets, Single Top), and di-top events with low M_R values. As expected, the yields diminish similarly in the case of the Ele box.

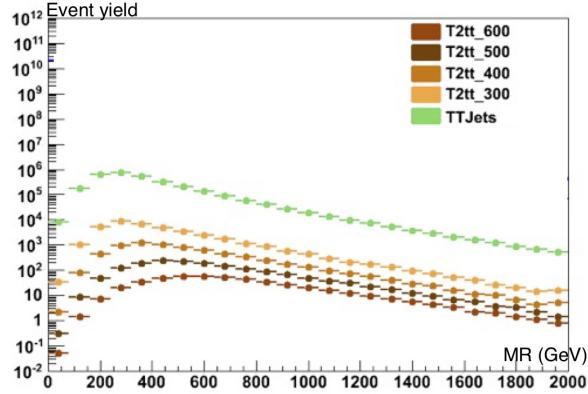


Figure 46. M_R distribution for background and signal events, before selection - The topology of standard model di-top events is quite similar to our sought signal events, with a peak at low M_R value, around 200GeV. Our analysis selection includes a cut on M_R so as to avoid this peak.

Process	Weight	No cut	$M_R > 300$ $R^2 > 0.05$	Lepton selection	Jet multiplicity	≥ 1 b-jet	$M_R > 350$ $R^2 > 0.15$
QCD-HT100To250	4004.03	2.10^{11}	2.10^6	4004	0	0	0
QCD-HT250To500	196.83	5.10^9	10^7	5118	2953	1772	0
QCD-HT500To1000	5.31	10^8	3.10^6	957	542	149	0
QCD-HT1000ToInf	0.28	4.10^6	9.10^4	20	13	3	0
DY	1.87	6.10^7	10^4	1996	899	269	11
WJets	10.17	6.10^8	10^5	31404	14538	3457	244
T_tW	0.41	2.10^5	10^4	1639	1110	1023	120
Tbar.tW	0.42	2.10^5	10^4	1627	1102	1015	118
TTJets	0.69	5.10^8	6.10^5	73787	57642	55360	5400
T2tt(500, 25)	0.013	123492	619	64	52	50	34

TABLE II

MC YIELDS FOR BACKGROUND PROCESSES, ELE BOX, AFTER EACH CUT. THE NUMBER OF EVENTS CORRESPONDS TO 19.3 FB^{-1} .

4.3.2.3 Hadronic box analysis selection

For the all-hadronic BJet box, the analysis selection is somewhat similar and consists of:

- a multijet trigger path

HLT_QuadJet50 requires four jets with an online p_t above 50 GeV, seeded with L1_QuadJet36;

- at least six jets with $p_t > 30 \text{ GeV}$, $|\eta| < 2.4$;
- at least one b-tagged jet (medium working point);
- no loose leptons, no isolated track;
- $M_R > 450 \text{ GeV}$, $R^2 > 0.15$.

The jets multiplicity and kinematic requirements are tighter to compensate for the fact that, without the lepton requirements, the QCD background is much larger. With these requirements, the trigger path HLT_QuadJet50 is fully efficient, as shown in Figure 47.

Process	Weight	No cut	$M_R > 300$ $R^2 > 0.05$	Lepton selection	Jet multiplicity	≥ 1 b-jet	$M_R > 350$ $R^2 > 0.15$
QCD-HT100To250	4004.03	$2 \cdot 10^{11}$	$2 \cdot 10^6$	4004	0	0	0
QCD-HT250To500	196.83	$5 \cdot 10^9$	10^7	1575	984	787	0
QCD-HT500To1000	5.31	10^8	$3 \cdot 10^6$	207	143	138	0
QCD-HT1000ToInf	0.28	$4 \cdot 10^6$	$9 \cdot 10^4$	3	1	1	0
DY	1.87	$6 \cdot 10^7$	10^4	2258	1062	327	19
WJets	10.17	$6 \cdot 10^8$	10^5	45018	20516	4951	407
T_tW	0.41	$2 \cdot 10^5$	10^4	2047	1388	1289	148
Tbar.tW	0.42	$2 \cdot 10^5$	10^4	2034	1377	1278	144
TTJets	0.69	$5 \cdot 10^8$	$6 \cdot 10^5$	93881	73529	70420	7238
T2tt(500, 25)	0.013	123492	619	84	69	66	45

TABLE III

MC YIELDS FOR BACKGROUND PROCESSES, MU BOX, AFTER EACH CUT. THE NUMBER OF EVENTS CORRESPONDS TO 19.3 FB^{-1} .

4.3.2.3.1 Rejection of the QCD background against ttbar events using a BDT

Another selection step is applied for the hadronic box, making use of a *Boosted Decision Tree* (BDT) through the TMVA tool (65). A BDT is a multivariate analysis method intended for event classification. Given a set of input variables with known distributions for background and signal samples¹, a multivariate analysis method attempts to use as much information as possible to differentiate signal-like from background-like events. The method of simple cuts, used to define the analysis selections above and usually done in a physics analyses, can also be seen as a multivariate analysis; but such a simple method does not exploit as much of the available information as a BDT multivariate analysis.

To build a Decision Tree (Figure 48), we need a known sample of signal and background events. The tree begins with a root node, where events are split into two subsets depending on whether they satisfy

¹Signal and background here are defined locally with respect to a particular multivariate analysis being performed. They may not necessarily be directly related to the previously described signal and background physics processes that are relevant to the whole analysis searching for stops.

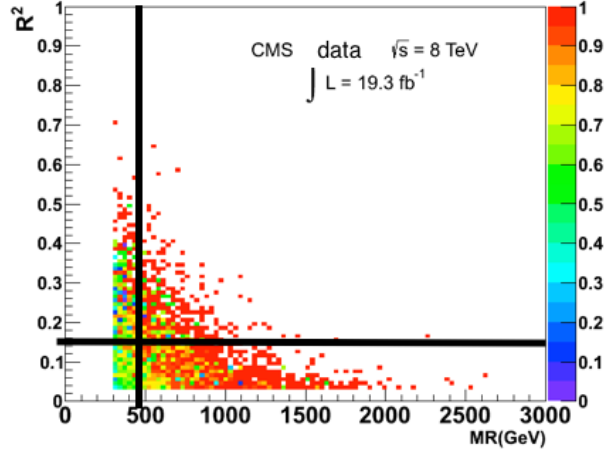


Figure 47. **Efficiency of the trigger path HLT_QuadJet50, used for the hadronic BJet box** - The path requires four jets with an online p_t above 50 GeV. In the $R^2 - M_R$ plane, the trigger plateau efficiency is reached for the kinematic requirements of the hadronic box analysis selection. The denominator cut is HLT_IsoMu24, requiring one isolated muon with $p_t > 24$ GeV.

or not a given cut on one of the input variables. The cut is chosen such that it gives the best signal / background separation possible, considering the set of input variables and scanning different cut values for each of these variables. Thus, two branches are created. The same procedure is repeated for each branch, with a different cut value. Input variables may be reused or ignored. The process is repeated, so as to reach a maximum purity for the signal sample. We stop building the tree when the sub-samples contain less than a given number of events, or after a given number of iterations (defining the depth of the tree). Boosting a Decision Tree consists of repeating the training using the same samples, but weighting the events taking into account the mistakes from the previous results. It makes the procedure more robust against overtraining.

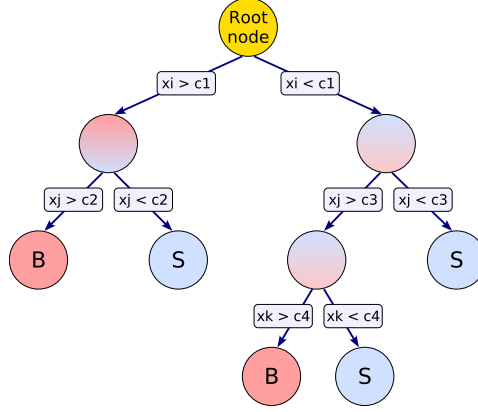


Figure 48. **Sketch of a (Boosted) Decision Tree** - A Decision Tree recursively splits events into signal-like and background-like sub-samples, depending on the results of cuts on the input variables (65).

We use a BDT to further improve the purity of the hadronic box, that is, further reject QCD events and select events containing top-quark pairs (di-tops), while maintaining good efficiency. Hence, our background training sample consists of events obtained from data, where we apply the hadronic box analysis selection except for the b-tagging requirement. Instead of requiring at least one b-tagged jet, events with a b-tagged jet are vetoed. This provides a QCD-enriched sample for the background. A pure signal sample containing only di-top events is obtained from Monte Carlo simulation 2.4. The input distributions to the BDT are top-tagging (66) and quark-gluon discrimination (67) variables (described in the next paragraphs). The BDT returns, for each event, a value between -1 and 1 indicating whether events are more signal-like or background-like. Using the BDT output distribution for known background and signal samples, we can then decide on a cut to differentiate signal from background events within a sample of unknown composition, on a event by event basis, as shown in Figure 49. Using the BDT output

(Figure 49), we decide to cut at an output value of -0.2 to split the hadronic box into two sub-samples: a low purity sample and a high purity one, the latter defining the *BJeHP* box.

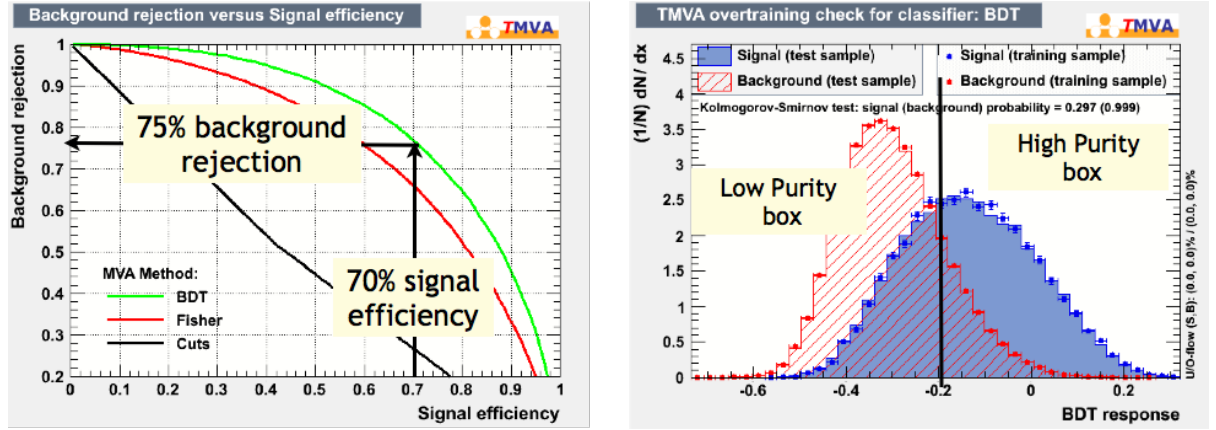


Figure 49. **BDT performance** - The BDT helps to select di-tops events and reject further the QCD background: a 75% background rejection can be achieved while keeping a 70% efficiency. Using the right plot, we decide to cut a BDT output value of -0.2 to split our initial hadronic box into two sub-samples: a low purity part and a high purity part. Only the latter is used in the analysis.

4.3.2.3.2 Top tagging variables

The T2tt simplified signal model contains two top quarks in the final state, which decay hadronically for the BJet search box: we want to assess the compatibility of each event with two hadronic decays of top quarks. To accomplish that, we start by forming a detector view of a di-top event; the situation depicted in Figure 50 should help visualizing the problem at hand. After clustering reconstructed particles into hemispheres, with the requirement that each hemisphere contains at least three jets to

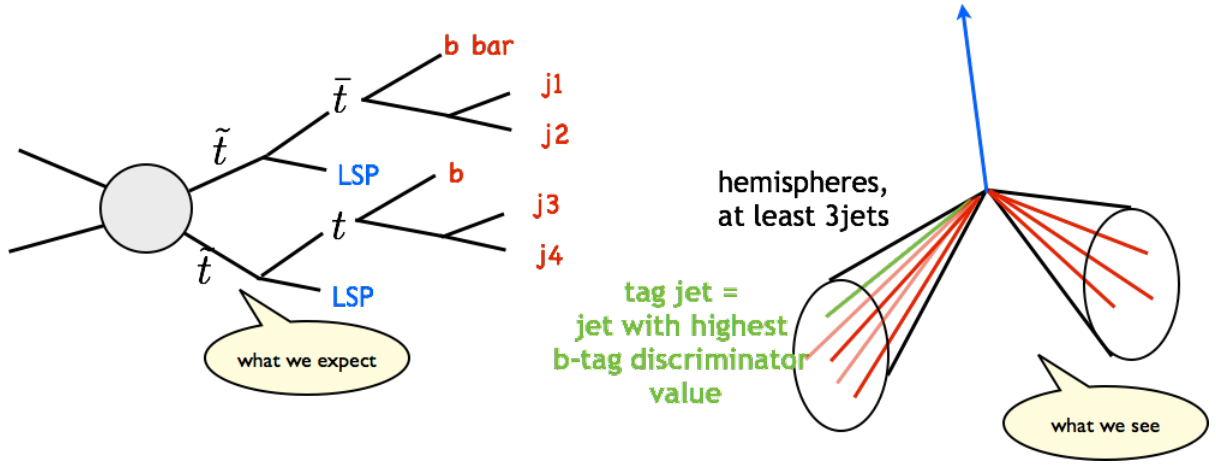


Figure 50. **Deriving top tagging variables: casting what we see into what we expect**

- In our events, instead of two tops, we see two hemispheres which should match each top decay. To meet this requirement, we require that each hemisphere contains at least three subjects, as expected for a top quark fully hadronic decay.

be consistent with a top decay, we compute m_W^H , m_t^H and the helicity angle Θ^H for each of the two hemispheres. In other words, each hemisphere is assumed to correspond to a separate top quark.

- m_W^H represents the invariant mass of a hadronically decaying W boson. A top quark decays to a b quark and a W boson, itself decaying to jets (for the hadronic box, leptons are vetoed). For each hemisphere, we compute the invariant mass for pairs of jets within the hemisphere, ignoring the jet with the highest CSV discriminant value, hence the jet with the highest probability for being a b-jet. The dijet combination whose mass is closest to the known W boson mass, 80GeV, provides m_W^H .
- m_t^H is computed, for each hemisphere, as the invariant mass of the previously selected b-jet and the two jets used to construct m_W^H .

- Θ^H is the helicity angle of the W boson (Figure 51). It corresponds to the angle between the top flight direction and the softest of the W decay products, where softest is defined using momenta in the laboratory frame. For QCD events, there is no real W, and the helicity angle is divergent for small values, while the distribution is almost flat for top events¹.

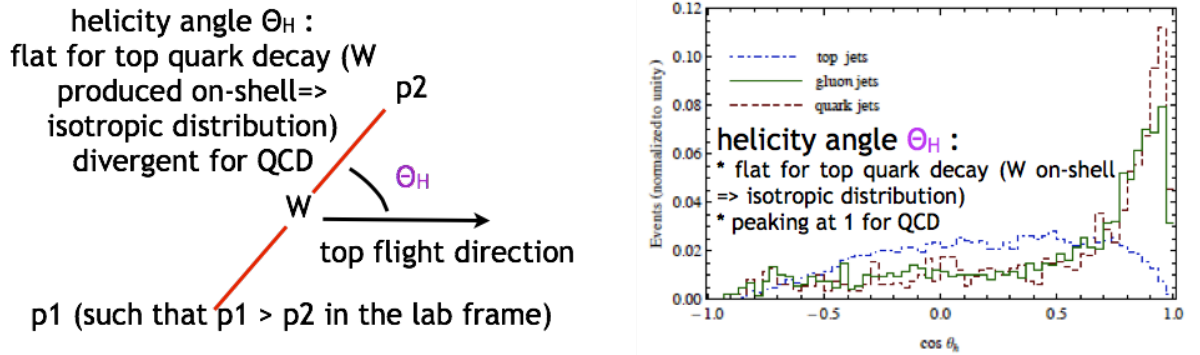


Figure 51. **Helicity angle** - While the helicity angle is rather flat for top decays including an actual W, it is divergent and peaking at 1 for QCD (no real W).

Figure 52 shows the distributions for these top tagging variables, corresponding to signal (simulated $t\bar{t}$ events) and background (QCD-enriched data) samples, for both hemispheres.

¹The helicity angle in top events has initially been used in semi-leptonic decays ((68), (69)), where there is no ambiguity about which W decay product to use since we only see the lepton. For these events, we are interested in the precise shape of the helicity angle distribution, from which we can extract the different helicity fractions: longitudinal, F_0 , left, F_- , or right, F_+ . Because of the V-A structure of the weak interaction, the latter is strongly suppressed. Precise measurement of F_+ , if they reach the point where a deviation from 0 is observed, may sign new physics processes.

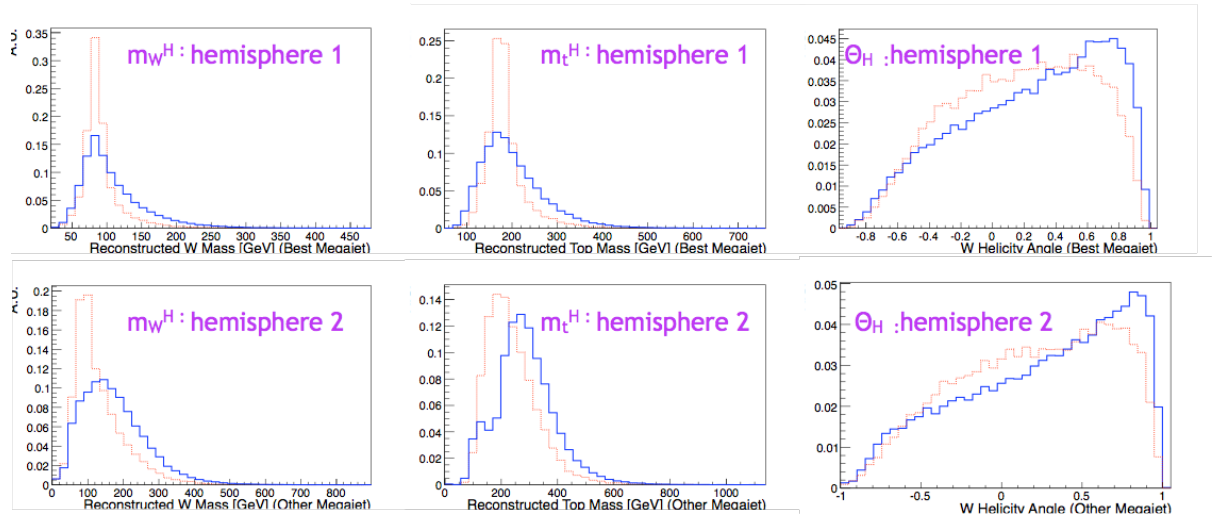


Figure 52. **Top-tagging variables** - Distributions for both hemispheres, for our QCD data sample and simulated di-top events. These variables are used as inputs for the BDT, along with quark-gluon discrimination variables. While it may be complicated to decide on strict cuts on these variables to distinguish QCD from $t\bar{t}$, a more advanced multivariate analysis technique, such as a BDT, provides a better signal/background disentanglement on an event by event basis.

4.3.2.3.3 Quark-gluon discrimination variables

T2tt and ttbar events produce top quarks, hence quark jets, while QCD multijet processes produce mostly gluon jets. The strong couplings for gluons and quarks are rather similar, with amplitudes given by:

$$\mathcal{M}(q \rightarrow qg) \propto \frac{2\alpha_s C_F}{\pi} \frac{dE}{E} \frac{d\theta}{\theta}; \quad (4.15)$$

$$\mathcal{M}(g \rightarrow gg) \propto \frac{2\alpha_s C_A}{\pi} \frac{dE}{E} \frac{d\theta}{\theta}; \quad (4.16)$$

In the above equations, E is the energy of the emitted gluon, θ the angle between the final radiating parton direction and the emitted particle direction. The expressions are valid for small E and small θ . Yet, since $C_A = 4$ while $C_F = 4/3$, gluons have a higher probability to emit radiation. As a result, gluon jets tend to be wider and show a higher particle multiplicity. Consequently, the following two variables are used to help to select quark jets and to reject gluon jets (67) :

- charged track multiplicity (number of charged particles);
- linear radial momentum (a measure of the jet broadening):

$$\mathbf{lrm} = \frac{\sum_{i \in \text{jet}} p_{t,i} r_i}{p_{T,\text{jet}}}, \quad (4.17)$$

where $p_{t,i}$ is the transverse momentum of the i th constituent, r_i is the angular distance (in radians) between the constituent and the jet axis.

Figure 53 shows the distributions of these variables for quark and gluon jets in simulation, confirming that gluon jets tend to have higher charged track multiplicity and are wider.

Table IV shows the effect of the cuts on background processes, including the BDT cut, and an example signal mass point.

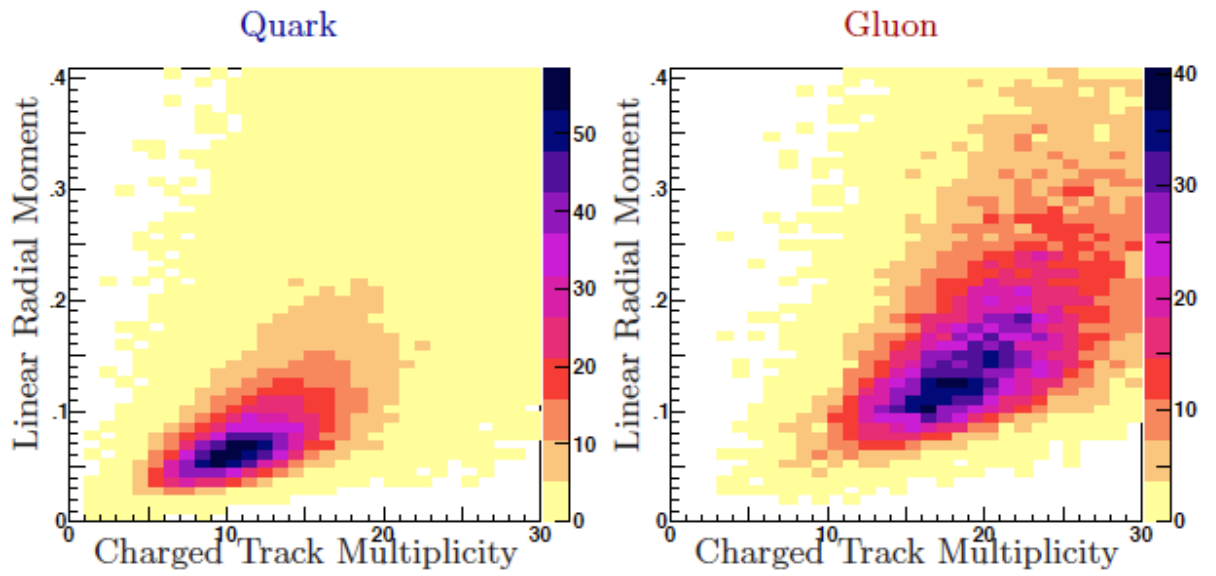


Figure 53. **Quark - gluon discrimination variables** - Charged multiplicity and linear radial momentum, which are used to help disentangle quark jets from gluon jets, and hence to distinguish di-top events from QCD background (67).

Process	Weight	No cut	$M_R > 300$ $R^2 > 0.05$	Lepton veto	Jet multiplicity	≥ 1 b-jet	$M_R > 450\text{GeV}$ $R^2 > 0.15$	BDT
QCD-HT100To250	4004.03	2.10^{11}	2.10^6	8.10^5	3.10^5	76077	0	0
QCD-HT250To500	196.83	5.10^9	10^7	6.10^6	2.10^6	964892	420	0
QCD-HT500To1000	5.31	10^8	3.10^6	2.10^6	10^6	475340	85	69
QCD-HT1000ToInf	0.28	4.10^6	9.10^4	8.10^4	45717	20423	0	9
DY	1.87	6.10^7	10^4	755	351	211	2	2
WJets	10.17	6.10^8	10^5	6700	3091	1179	61	10
T_tW	0.41	2.10^5	10^4	3971	2477	2352	38	15
Tbar_tW	0.42	2.10^5	10^4	3977	2488	2362	37	14
TTJets	0.69	5.10^8	6.10^5	2.10^5	161097	156452	1939	891
T2tt(500, 25)	0.013	123492	619	300	220	213	122	45

TABLE IV

MC YIELDS FOR BACKGROUND PROCESSES, BJETHP BOX, AFTER EACH CUT. AGAIN, THE FINAL KINEMATIC CUTS REMOVE EVENTS WITH SMALL MISSING MOMENTUM (WJETS, SINGLE TOP), AND DI-TOP EVENTS WITH LOW M_R VALUES. THE FINAL BDT CUT REJECTS FURTHER THE QCD BACKGROUND. THE NUMBER OF EVENTS CORRESPONDS TO 19.3 FB^{-1} .

4.3.2.4 Data samples

The data used in the present analysis corresponds to the 8 TeV data collected by the CMS experiment in 2012. We rely on three datasets:

- *SingleMu*: for the Mu box, relying on the single mu trigger path `HLT_IsoMu24`, seeded with `L1_SingleMu16`;
- *SingleElectron*: for the Ele box, relying on the single electron trigger path `HLT_Ele27_WP80`, seeded with `L1_SingleEG22`;
- *MultiJet1Parked*: for the Hadronic box, relying on the multijet trigger path `HLT_QuadJet50`, seeded with an OR of L1 seeds requiring four L1 jets above 36 GeV or a L1 HT above 150 GeV or two L1 jets above 56 GeV.

All the trigger paths used were unprescaled during the whole 2012 data taking period. Thus we use all of the 19.3 fb^{-1} of data collected by the experiment, except for the hadronic box, which relies on *Parked Data*, corresponding to 18.6 fb^{-1} . *Parked Data* indicates that the data was recorded but, due to the limited availability of offline computing resources, the reconstruction of these samples was postponed rather than performed promptly. Indeed, since the 2012 data taking was to be followed by a long shutdown, during which the computing resources normally devoted to reconstruction would not be used, CMS recorded more data (up to 1 kHz) than usually allowed. This data parking option allowed CMS to trigger events with looser criteria, hence extending the sensitivity reach.

4.3.2.5 Comparison of data versus simulation

Figure 54 and Figure 55 show projections of the R^2 and M_R distributions for data and (stacked) backgrounds (taken from simulation) after all analysis selection steps have been applied. The main background for leptonic boxes come from $t\bar{t}$ events, while we still have some QCD events for hadronic boxes. These plots demonstrate that the data and the simulation of Standard Model processes agree

well, especially for the core of the distributions. More discrepancies are to be expected in the tails, due to the high weights of some samples, such as the QCD and W+jets samples. Yet, even in the tails of the distributions, data and simulated simulation remain rather consistent. The dominant background for the leptonic boxes after the analysis selection is the $t\bar{t}$ bar+jets background. W+jets events follow, to a lesser extent. The main background processes for the BJetHP box are the top and QCD backgrounds. It is important to note that *these plots are illustrative only and are not used in the quantitative background estimation, which is data driven.*

Figure 56 shows the expected efficiency for the $T2tt$ signal after the analysis selection for the Ele, Mu, and BJetHP boxes. The efficiency is low along the diagonal corresponding to small mass splitting between the stop and LSP, as expected: when the mass splitting is small, R^2 is low, and signal events fail the $R^2 > 0.15$ cut. Correspondingly, the selection efficiencies are highest for large stop masses or large mass splittings.

4.3.3 Establishing the background modeling

After defining the boxes, we need to test our fit form for the background modelling. To do so, we use simulated events, corresponding to the main background processes for each channel. After performing a first fit of the reference sample, we evaluate the fit performance as described in 4.3.1.

4.3.3.1 Leptonic boxes

For the leptonic boxes, we use two instances of the form, corresponding to events with one b-tagged jet, and separately for events containing two or more b-tagged jets. A simultaneous fit involving both forms is performed on all events (any number of b-tagged jets). We use simulated $t\bar{t}$ bar events, which correspond to the main background for the leptonic boxes, to test the fit. The free parameters after fitting take the values given in Table V for the Ele box, and Table VI for the Mu box.

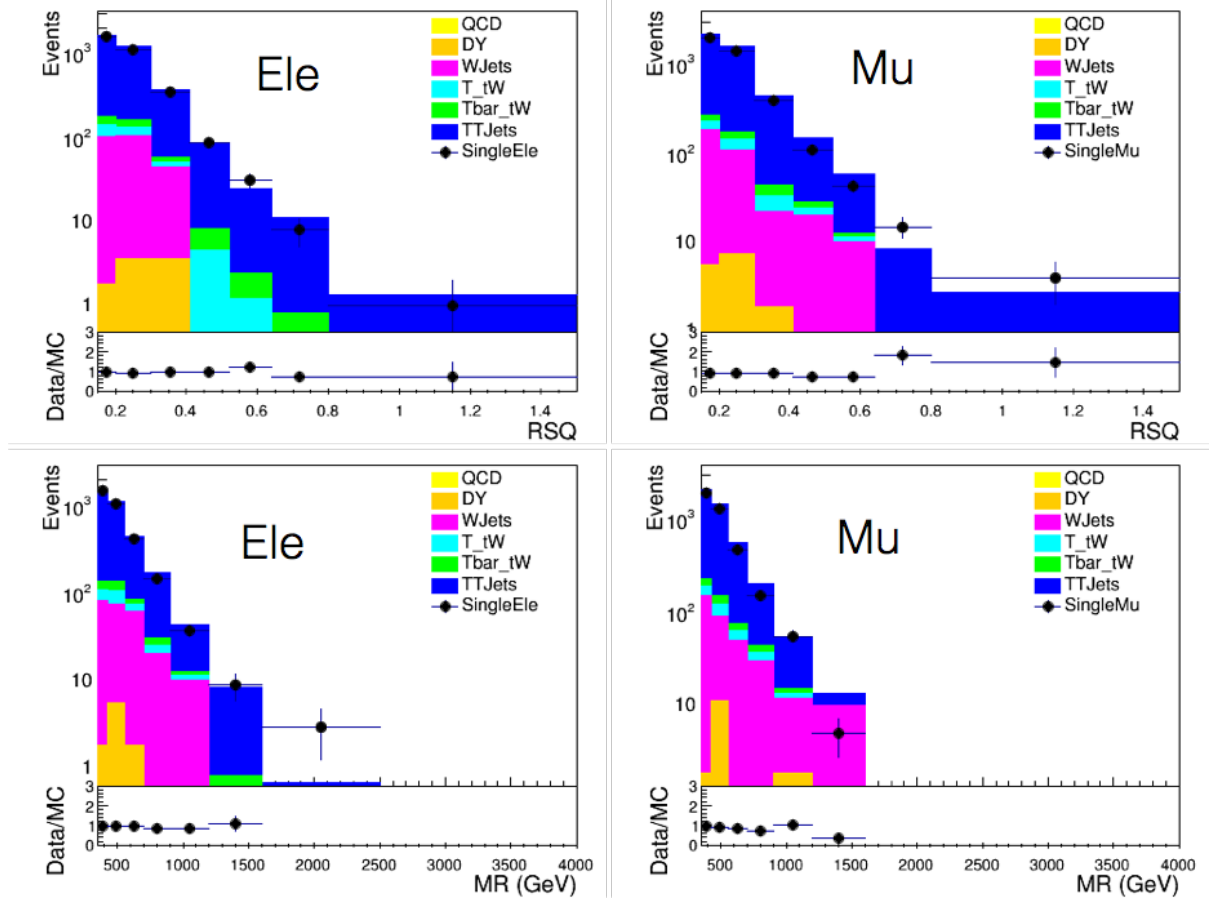


Figure 54. **Comparison of data with simulated events for R^2 and M_R distributions, leptonic boxes** - The left column corresponds to the Ele box, while the right column corresponds to the Mu box. The data points (black) and the simulated events (colored, stacked) histograms agree well, especially for the core of the distributions. More discrepancies are to be expected for the tails, due to the high weights of some samples, such as the QCD and WJets samples. Yet, even in the tails of the distributions, data points and simulated events histograms remain rather consistent. The dominant background for the leptonic boxes after the analysis selection is the $t\bar{t}$ +jets background. W+Jets events follow, to a lesser extent.

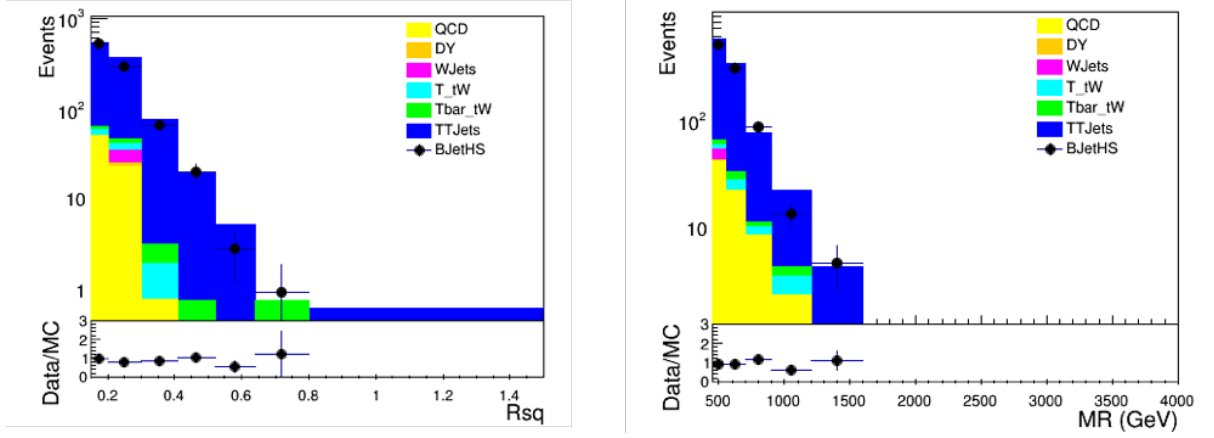


Figure 55. **Comparison of data with simulated events for R^2 and M_R distributions, hadronic BJethHP box** - The data points (black) and simulated events (colored, stacked) histograms agree rather well. The main background processes for the BJethHP box are the top and QCD backgrounds.

Figure 57 and Figure 58 show the results of the fit: for both boxes, the fit projections and the simulated data points agree well, as indicated by the French flag plots. We thus conclude that the fit is able to describe the $t\bar{t}$ backgrounds properly, and that our background modeling is adequate.

4.3.3.2 Hadronic boxes

For the hadronic BJethHP box, we again use two instances of the fit form, one form corresponding to events with one b-tagged jet, and a separate form for events containing two or more b-tagged jets. And, a simultaneous fit of all events is performed. In order to test the fit, we build a cocktail of simulated events containing $t\bar{t}$ +jets, Single Top and QCD events (our main backgrounds as we can read from Table IV), whose contributions match the weights given in table Table IV. Table VII shows the free parameters values after fitting, and Figure 59 shows the results of the fit: both fit projections and the

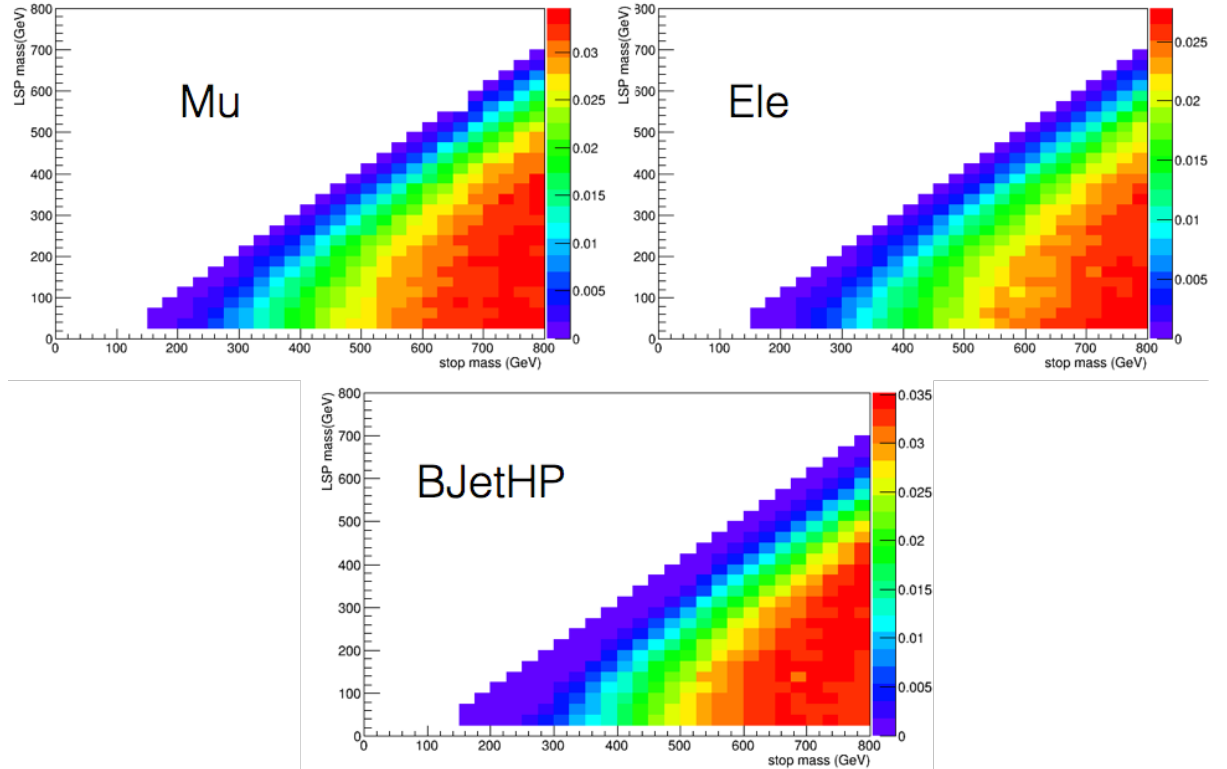


Figure 56. **Efficiency map after the analysis selection, for the T2tt simplified model**
 - Leptonic, Mu and Ele boxes for the top plots, BJetHP box for the bottom plot. As expected, the efficiency decreases for small mass splittings (softer objects, failing the kinematic cuts). The efficiency reaches about 2.5-3% for the leptonic boxes, with a difference explained by the 5GeV difference on the leptonic p_t cuts, itself due to different trigger cuts. The efficiency for the hadronic box is a bit higher, up to 3.5%.

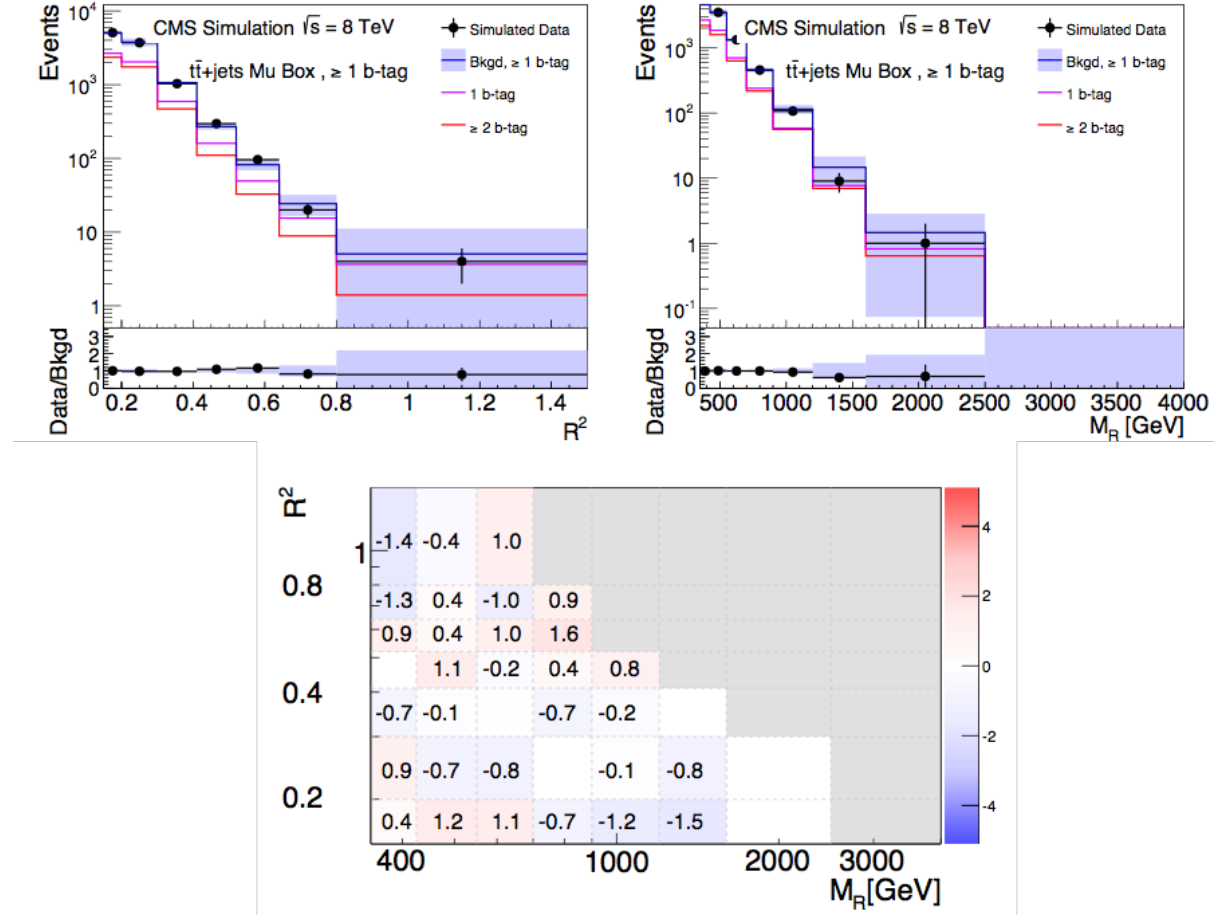


Figure 57. **Background modeling test using $t\bar{t}$ +jets for the Mu box** - In order to establish our background modeling principle, we test it on simulated events. For leptonic boxes, we use events corresponding to the main background process, that is, di-top events. The fit projections, along with the error bands, accommodate the simulated data points well. Besides, there is no significant deviation between the observed yields and fit predictions according to the French flag plot. Thus, the fit describes properly our $t\bar{t}$ simulated sample, the method is satisfactory for the Mu box. Fit systematics are included in the error bands thanks to the toy generation procedure described in 4.3.1.

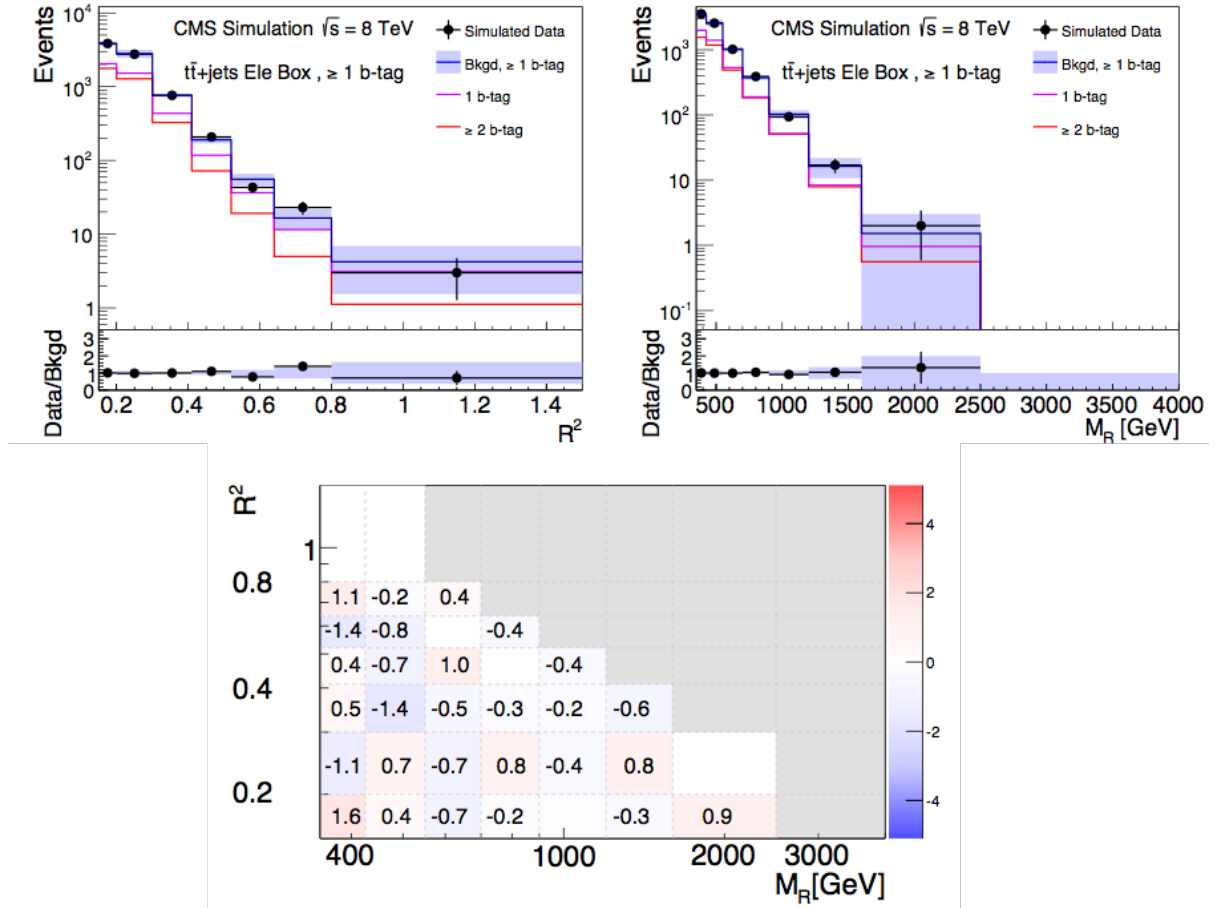


Figure 58. **Background modeling test using $t\bar{t}$ bar+jets for the Ele box** - The simulated data points fall within the error bands of the fit projections, and the French flag plot reveals no significant deviation between the observed yields and the fit prediction: the fit describes properly our $t\bar{t}$ bar simulated sample. Thus, the background modeling works for the Ele box as well. Fit systematics are included in the error bands thanks to the toy generation procedure described in 4.3.1.

Floating Parameter	InitialValue	FinalValue $\pm$ Error
MR0_TTj1b	-3.7950e+02	-3.7847e+02 $\pm$ 1.31e+02
MR0_TTj2b	-7.1123e+02	-7.1147e+02 $\pm$ 2.57e+02
Ntot_TTj1b	3.0720e+03	4.1760e+03 $\pm$ 6.46e+01
Ntot_TTj2b	2.1670e+03	3.4690e+03 $\pm$ 5.89e+01
R0_TTj1b	-3.1541e-01	-3.1469e-01 $\pm$ 8.36e-02
R0_TTj2b	-4.3758e-01	-4.3772e-01 $\pm$ 1.43e-01
b_TTj1b	3.4895e-01	3.9475e-01 $\pm$ 8.11e-02
b_TTj2b	1.3832e-01	1.3830e-01 $\pm$ 1.20e-02
f3_TTj2b	1.1126e-01	1.1127e-01 $\pm$ 5.34e-03
n_TTj1b	1.9553e+00	2.0356e+00 $\pm$ 1.39e-01

TABLE V

FREE PARAMETERS VALUES AFTER FITTING THE SIMULATED TTBAR EVENTS SAMPLE, ELE BOX. TTJ1B AND TTJ2B REFER TO THE COMPONENT: ONE FOR ONE B-TAGGED JET EVENTS, ONE FOR TWO OR MORE B-TAGGED JETS EVENTS.

French flag plot are satisfactory, and we conclude that the fit is able to describe our main backgrounds properly.

4.3.4 Probing our sensitivity to signal: expected limits and signal injection tests

Before attempting to look for a T2tt signal in the data, we must ensure that we understand the sensitivity of our workflow. One way to achieve this is to determine, in terms of the parameters that describe the model (in this case, the stop mass and LSP mass), which upper limits on signal strength can be set with our guaranteed background only fit results. Then, we can test our ability to find a signal just above detectability in a simulated signal plus background sample. The two following sections describe the procedures for doing this: 4.3.4.1 addresses the limit setting procedure, while 4.3.4.2 proceeds with signal injection tests.

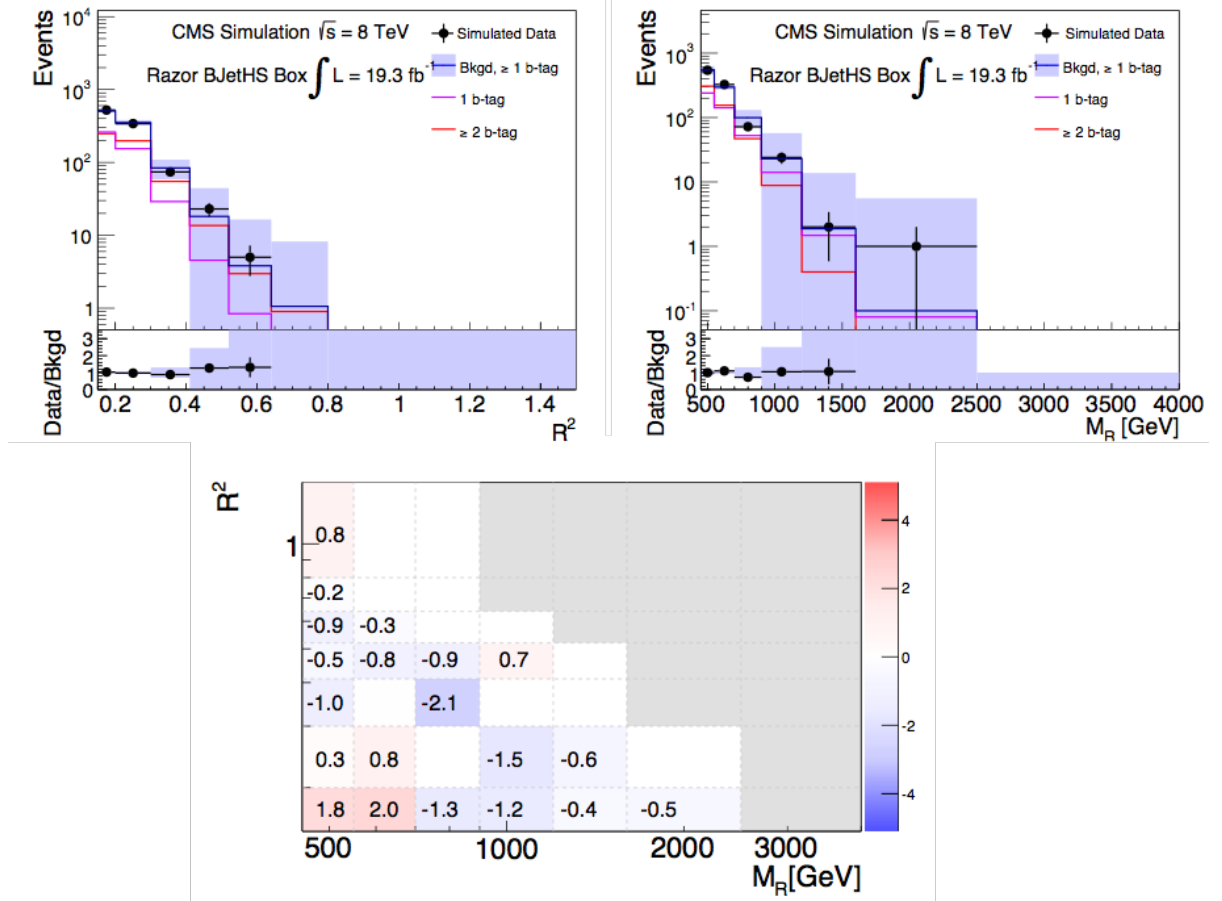


Figure 59. **Background modeling test using simulated Standard Model processes for the BJetHP box** - The simulated data points are well accommodated by the fit projections including the error bands. Despite some fluctuations, there is no significant excess in the French flag plot. The fit describes properly our sample of simulated events (top and QCD): the model provides a good description for the backgrounds for the BJetHP box. Fit systematics are included in the error bands thanks to the toy generation procedure described in 4.3.1.

Floating Parameter	InitialValue	FinalValue $\pm$ Error
MR0_TTj1b	-3.9203e+02	-3.9203e+02 $\pm$ 1.13e+02
MR0_TTj2b	-6.3764e+02	-6.3762e+02 $\pm$ 1.92e+02
Ntot_TTj1b	3.8050e+03	5.5230e+03 $\pm$ 7.43e+01
Ntot_TTj2b	2.7150e+03	4.7101e+03 $\pm$ 6.86e+01
R0_TTj1b	-3.4621e-01	-3.4621e-01 $\pm$ 7.56e-02
R0_TTj2b	-4.4493e-01	-4.4492e-01 $\pm$ 1.16e-01
b_TTj1b	2.3235e-01	2.3235e-01 $\pm$ 7.96e-02
b_TTj2b	1.8250e-01	1.8250e-01 $\pm$ 1.03e-02
f3_TTj2b	1.1019e-01	1.1019e-01 $\pm$ 4.56e-03
n_TTj1b	1.7334e+00	1.7334e+00 $\pm$ 1.64e-01

TABLE VI

FREE PARAMETERS VALUES AFTER FITTING THE SIMULATED TTBAR EVENTS SAMPLE, MU BOX. TTJ1B AND TTJ2B REFER TO THE COMPONENT: ONE FOR ONE B-TAGGED JET EVENTS, ONE FOR TWO OR MORE B-TAGGED JETS EVENTS.

4.3.4.1 Limit setting - LHC CLs

If no evidence for a model is observed in the data, we can extract from our results an upper limit on its possible strength, or cross section. If this observed upper limit is lower than the theoretically predicted value, we can conclude that the corresponding model does not describe properly the data, and is excluded. The limit setting procedure follows what has been established for the Higgs boson searches at the LHC (70).

Statistical methods aimed at setting limits look for the degree of incompatibility of the data with a background only hypothesis (28). This is typically quantified with a *confidence level*, CL , and the usual requirement is that $CL < 0.05$: there is less than 5% chance that the data sample corresponds to background only.

4.3.4.1.1 Building a test statistics, CL_s definition

In order to assess if our data is better described by a hypothetical signal having a given strength μ together with background (hypothesis $H1$) or by the alternate situation (background only, hypothesis

Floating Parameter	InitialValue	FinalValue $\pm$ Error
MR0_TTj1b	-3.0547e+02	-3.9042e+02 $\pm$ 4.37e+02
MR0_TTj2b	-1.4943e+03	-4.9133e+03 $\pm$ 6.04e+03
Ntot_TTj1b	4.7100e+02	4.5005e+02 $\pm$ 2.12e+01
Ntot_TTj2b	5.2700e+02	5.1801e+02 $\pm$ 2.28e+01
R0_TTj1b	-2.8143e-01	-2.2238e-01 $\pm$ 1.94e-01
R0_TTj2b	-9.7194e-01	-3.4921e+00 $\pm$ 4.10e+00
b_TTj1b	1.3512e-01	5.6705e-02 $\pm$ 9.56e-02
b_TTj2b	2.0087e-03	2.4333e-03 $\pm$ 2.74e-03
f3_TTj2b	1.6298e-01	1.7182e-01 $\pm$ 1.66e-02
n_TTj1b	1.4535e+00	1.2053e+00 $\pm$ 3.59e-01

TABLE VII

FREE PARAMETERS VALUES AFTER FITTING THE SIMULATED MC COCKTAIL EVENTS SAMPLE, BJETHP BOX. TTJ1B AND TTJ2B REFER TO THE COMPONENT: ONE FOR ONE B-TAGGED JET EVENTS, ONE FOR TWO OR MORE B-TAGGED JETS EVENTS.

$H0$), we begin by building a test statistic. A test statistic is a quantity that should exhibit distinguishable distributions for signal plus background and background only samples. A missing transverse momentum distribution could be used for instance. The test statistic used by the *CLs* method(s) derives from the likelihood ratio¹:

$$q = -2 \ln \frac{\mathcal{L}(X|H1)}{\mathcal{L}(X|H0)}. \quad (4.18)$$

Here, $\mathcal{L}(X|H1)$ corresponds to the likelihood of observing our data X , if $H1$ is true; while $\mathcal{L}(X|H0)$ corresponds to the likelihood of observing our data X , if $H0$ is true. If $H1$ is the signal plus background hypothesis and $H0$ the background only hypothesis, the test tends to be negative if our data X is more compatible with $H1$, while it is more likely to be positive if we only have background in our sample.

¹according to the Neyman-Pearson lemma, the likelihood ratio or a monotonic function of it provides the most powerful test. The power of a test is defined as $1 - \beta$, where β is the probability for not rejecting $H0$ against $H1$ when $H1$ is true.

With the test statistic defined as above, one can throw toy-experiments, that is generate samples of pseudo-data, following the $H1$ hypothesis to build the probability density function for some variable q under the $H1$ hypothesis. The latter can then be used to compute the p -value corresponding to the observed value of q_{obs} , for our real data sample, such that $p = \int_{q_{\text{obs}}}^{\infty} pdf(q) dq$, where pdf is the probability density function for q . This p -value is referred to as the signal+background confidence level, CL_{s+b} . It is the probability for the data to exhibit a greater incompatibility with $H1$ than the one actually observed. If $CL_{s+b} < 0.05$, $H1$ is rejected.

Yet by construction, any analysis has a 5% to observe $CL_{s+b} < 0.05$ with a zero signal strength: such a procedure is not fully safe against the possibility of rejecting the background only hypothesis, when it is actually true, i.e. when the background experiences a downward fluctuation.

A possible workaround, used by the LEP experiments, consists in redefining the confidence level as:

$$CL_s = \frac{CL_{s+b}}{CL_b} \quad (4.19)$$

where CL_b is obtained from the background-only probability density function for the test statistic q . Figure 60 illustrates the different quantities.

4.3.4.1.2 Nuisance parameters treatment

To define statistical methods used in order to set limits, we need to specify not only the test statistic, but also the treatment of nuisance parameters, denoted as θ . In this work, nuisance parameters corresponds to the various uncertainties, statistical or systematic, that may affect the predictions for both signal and background yields or shapes. Some affect both signal and background processes; some are specific to signal, to the extent that we need to account for any discrepancies in data versus simulation.

Signal uncertainties

The following uncertainties affecting the signal yields are considered .

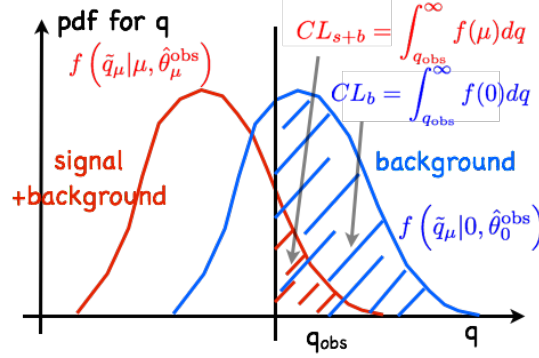


Figure 60. **Computing CL_s** - The quantity CL_{s+b} is computed as the p -value corresponding to q_{obs} for the s+b pdf: it is the probability for the data to have equal or greater incompatibility with the H_1 or s+b hypothesis. CL_b is computed as the p -value corresponding to q_{obs} for the B-only pdf: it is the probability for the data to have equal or greater incompatibility with the H_0 or background only hypothesis. CL_s then is the ratio of the two quantities.

- Trigger efficiency: estimated to be 95% with an uncertainty of 5%;
- Luminosity: officially given as 4.4% uncertainty;
- Shape uncertainties:

Some uncertainties result in larger and smaller values for the razor variables (or other cut variables used in the analysis) for each event, possibly preventing an event from passing our analysis selection or, on the contrary, promoting it to our search region. Thus, they can affect the shape of the razor variables' distributions. To build the signal probability density function $f_s(R^2, M_R, \theta)$, for each signal mass point $(m_{\text{stop}}, m_{\text{LSP}})$, we fill 2D histograms in the R^2, M_R plane, recording the upward and downward fluctuations corresponding to one standard deviation for each uncertainty. Out of these, a completely parameterized probability density function $f_s(R^2, M_R, \theta)$ can be built, where the uncertainties enter the probability density functions through the nuisance parameters

θ . The following systematic shape uncertainties are then considered in the analysis to account for any data versus simulation discrepancies:

- *Parton distribution functions (PDFs) (35)*

The physics processes produced in proton-proton interactions depend on the type of partons involved and the parton momentum fractions they carry. Uncertainties in the PDFs can therefore affect center-of-mass collision energy (in the parton-parton frame of reference) and hence can affect the kinematics, thus the shape of the signal probability density function f_s in the M_R, R^2 plane.

- *Jet energy scale*

Using simulated events, we can compare jet energies obtained from clustering particles at *generator* level to jet energies after those particles have gone through a detector simulation and jet reconstruction. This allows one to calibrate jet energies, which also possess an uncertainty. Uncertainties on jet energies imply uncertainties on the razor variables, which then affect the shape of their distributions.

- *b-tagging and lepton identification efficiencies*: both may be different for data and simulated events, which we need to take into account. They may promote or demote events from the signal region, therefore modifying the shape of the razor variables.

It is important to recall that the free parameters from Equation 4.13 along with normalization are fit to the data, producing a covariance matrix for the background model, as described in Section 4.3.1. Those fit errors are considered as systematic uncertainties on the background model, and hence are taken into account in the construction of each background pdfs $f_b(R^2, M_R, \theta)$.

With the LHC CLs method, each nuisance parameter is introduced in the test statistic through the likelihood, multiplying the Poisson part by its probability density function, $p(\tilde{\theta}|\theta)$, that encapsulates both its *best estimate*, $\tilde{\theta}$, and its distribution:

$$\mathcal{L}(\text{data}|\mu, \theta) = \text{Poisson}(\mu s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta). \quad (4.20)$$

where the signal plus background (Poisson) likelihood may be written, for our analysis, as:

$$\mathcal{L}(X|\mu, \theta) = \frac{\exp\left(\mu\epsilon L + \sum_{j \in \text{SM}} N_j\right)}{N!} \prod_{i=1, N} \left(\mu\epsilon L f_s(R_i^2, M_{R,i}, \theta) + \sum_{j \in \text{SM}} N_j f_j(R_i^2, M_{R,i}, \theta) \right) \cdot p(\tilde{\theta}|\theta). \quad (4.21)$$

The strength modifier μ corresponds in our case to the cross section, θ refers to the set of nuisance parameters describing systematic errors, ϵ is our efficiency for the signal considered and L is the integrated luminosity. The sum and product over SM run over our two standard model background components. N is the number of events in our dataset. The probability density function for the signal, $f_s(R^2, M_R, \theta)$, is evaluated thanks to Monte Carlo simulated signal events, and $f_b(R^2, M_R, \theta)$ is the razor probability density function for Standard Model background processes obtained in the previous analysis step using Equation 4.13. The likelihood may be evaluated for an actual data sample or for samples made of simulated events.

The distribution $p(\tilde{\theta}|\theta)$ reflects our knowledge of the nuisance parameter behavior. If nothing is known about it, a flat distribution is recommended. For parameters that can take positive and negative values, the normal distribution can be used (this is what we do for the shape uncertainties described previously), whereas the log-normal distribution is more suited to positively defined variables (used for the luminosity and trigger uncertainties detailed before) (70).

With the LHC CLs method, nuisance parameters, possibly including the signal strength μ , are *profiled*, as indicated by the caret symbol, to ensure good coverage properties. This means that we determine them so as to maximise the likelihood. The corresponding test statistics then becomes the LHC CLs test statistic:

$$q_\mu = -2 \ln \frac{\mathcal{L}(X|\mu, \hat{\theta})}{\mathcal{L}(X|\hat{\mu}, \hat{\theta})}, \quad 0 \leq \hat{\mu} \leq \mu. \quad (4.22)$$

The LHC CLs test statistic compares a trial cross section, μ , against the best fit, which maximizes the likelihood. We determine two specific sets of nuisance parameter values:

$$\hat{\theta}_0^{\text{obs}} : \mathcal{L}(\text{data}|\mu=0, \hat{\theta}_0^{\text{obs}}) \rightarrow H0 \text{ hypothesis} \quad (4.23)$$

$$\hat{\theta}_\mu^{\text{obs}} : \mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu^{\text{obs}}) \rightarrow H1 \text{ hypothesis}. \quad (4.24)$$

The set of nuisance parameters maximizing the likelihood for our data sample under the background only hypothesis is $\hat{\theta}_0^{\text{obs}}$, while $\hat{\theta}_\mu^{\text{obs}}$ is the set of nuisance parameters maximizing the likelihood for our data sample under the hypothesis of signal of strength μ plus background.

The final two probability density functions, which are required to construct CL_s are:

- $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$: the probability density function for the test statistic under the $\mu s + b$ ($H1$) hypothesis. To build this, we generate pseudo-data using the signal (strength μ) plus background pdf, with nuisance parameters and μ fixed. All of the nuisance parameters are then allowed to float when evaluating the test statistics.
- $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})$: the probability density function for the test statistic, under the background only hypothesis, $H0$. To build this, we generate pseudo-data using the a signal strength of zero. Again, while fixed for the pseudo-data generation, the nuisance parameters are allowed to float when evaluating the test statistic.

Using the above two probability density functions, the CL_s quantity can now be evaluated for a set of cross sections or signal strength μ . If $CL_s < 0.05$, the $\mu s + b$ hypothesis is rejected with a 95% confidence level.

4.3.4.1.3 Observed and expected limits

The procedure previously described provide the limit which is *observed*, either in the real data itself or as a closure test of the procedure using simulated data. Generating many sets of pseudo-data (using simulated data) under the background only hypothesis, we can compute CL_s for each of them, following the same procedure. For each of these pseudo-datasets, we compute CL_s , build a cumulative probability density function, and extract the median *expected* limit (50% quantile), along with the $\pm 1 \sigma$ and $\pm 2 \sigma$ values.

Asymptotic formulae have been derived as approximations for the full toy-based method, for both the test statistic and its probability density functions in (71), so that one can obtain the probability density function for the test statistics without throwing toy experiments, hence saving time and computing resources.

4.3.4.1.4 Limits from our background only fits to simulated events

Using the LHC-CLs procedure, we can test the sensitivity of the analysis using our previous fit results on simulated data samples for the background pdfs in Equation 4.21. The obtained limits on the signal strength μ will correspond to the signal mass points that we would expect to be able to exclude with 95% confidence level.

To validate the asymptotic approximation for this work, Figure 61 shows both the observed (red) and the expected (black dotted with uncertainty bands) limits for the BJetHP box on simulated data, using the asymptotic method, along with a comparison to the output from the full toy-based method for a few selected key mass points near where we cross the theoretically expected values that yield the exclusion range. The predicted supersymmetric cross sections are also plotted (72). Both methods are

consistent, and stop masses are seen to be excluded from 200 GeV up to 690 GeV. We thus conclude that the asymptotic approximation agrees very well with the full toy-based method, and so we will use only the asymptotic method for the remainder of this work.

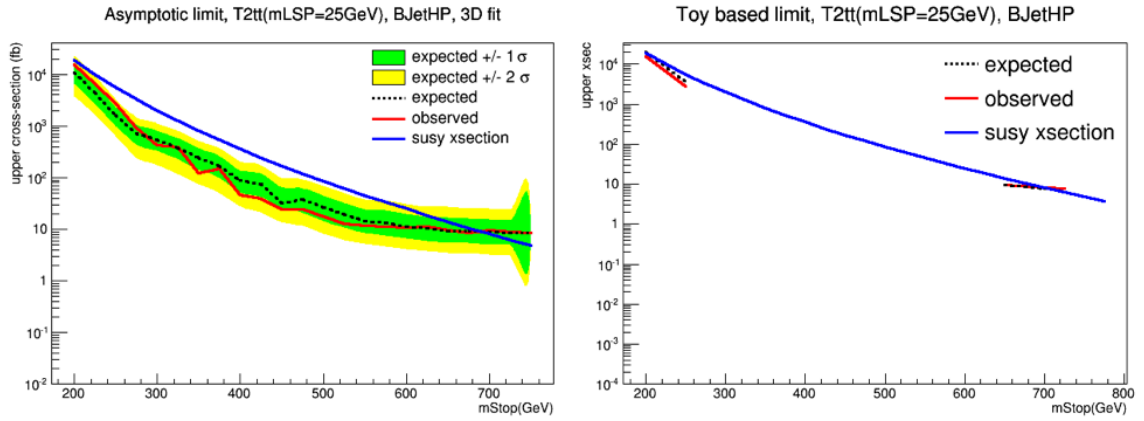


Figure 61. **Expected (black dotted with uncertainty bands) and observed (red) limits using simulated data** - The left plot shows the asymptotic limit result, while the right plot shows the output of the toy-based method for two specific points, to check that both methods agree in terms of exclusion range. The blue line gives the predicted supersymmetric cross sections (the fit results used are from 4.3.3.2)

Using simulated data, we establish the expected sensitivity of the leptonic search boxes to stop masses in the T2tt model, for light LSP masses. The left plot of Figure 62 shows the exclusion limits for the Ele box, while the right plot corresponds to the Mu box. The bottom plot shows the exclusion limits when the results from the Ele box and the Mu box are combined. In all plots, the red line corresponds to the observed limit on the simulated data and the blue line gives the predicted supersymmetric cross

sections. Low stop masses are excluded when both boxes are combined (the fit results used are from 4.3.3.1), indicating that the leptonic boxes are sensitive to the low mass part of the stop spectrum.

4.3.4.2 Signal injection tests

To validate that our workflow allows us to distinguish signal from background, in the presence of an actual signal, we perform signal injection tests. The question we ask is the following: using the procedure described previously, are we able to detect a signal with a given strength in the presence of the known backgrounds?

4.3.4.2.1 BJetHP box

For our closure tests, the signal used is T2tt, with a stop mass of 500GeV and an LSP mass of 25GeV. The signal is injected with a test cross section set to 1000fb (1pb), which is well above the expected limit from the previous paragraph. Figure 63 shows the 1D projects and the full 2D R^2 and M_R distributions for that test signal. A clear peak is seen near 500 GeV in the 1D distribution for M_R and a corresponding cluster is seen in the 2D R^2 , M_R plane.

The first step in this closure test consists in trying to fit our background only model to a sample containing simulated backgrounds (cross-section weighted according to the luminosity of the data), plus some signal events with the chosen test cross-section. Eventhough signal has been injected into the simulated sample, Figure 64, demonstrates that the fit of the background model is rather robust. For example, the blue line (and error bands) cover the simulated data points (black) well in the 1D projections of M_R and R^2 . Further, the French flag plot, while showing deviations at the 2σ level in some cells, suggests that the background functional form is still able to model the background sufficiently well.

The second step in the closure test consists in fitting our signal plus background sample using the signal plus background probability density function, and floating the cross section for the signal as a fit parameter. As we can see on Figure 65, such a fit is able to recover the injection cross-section with a

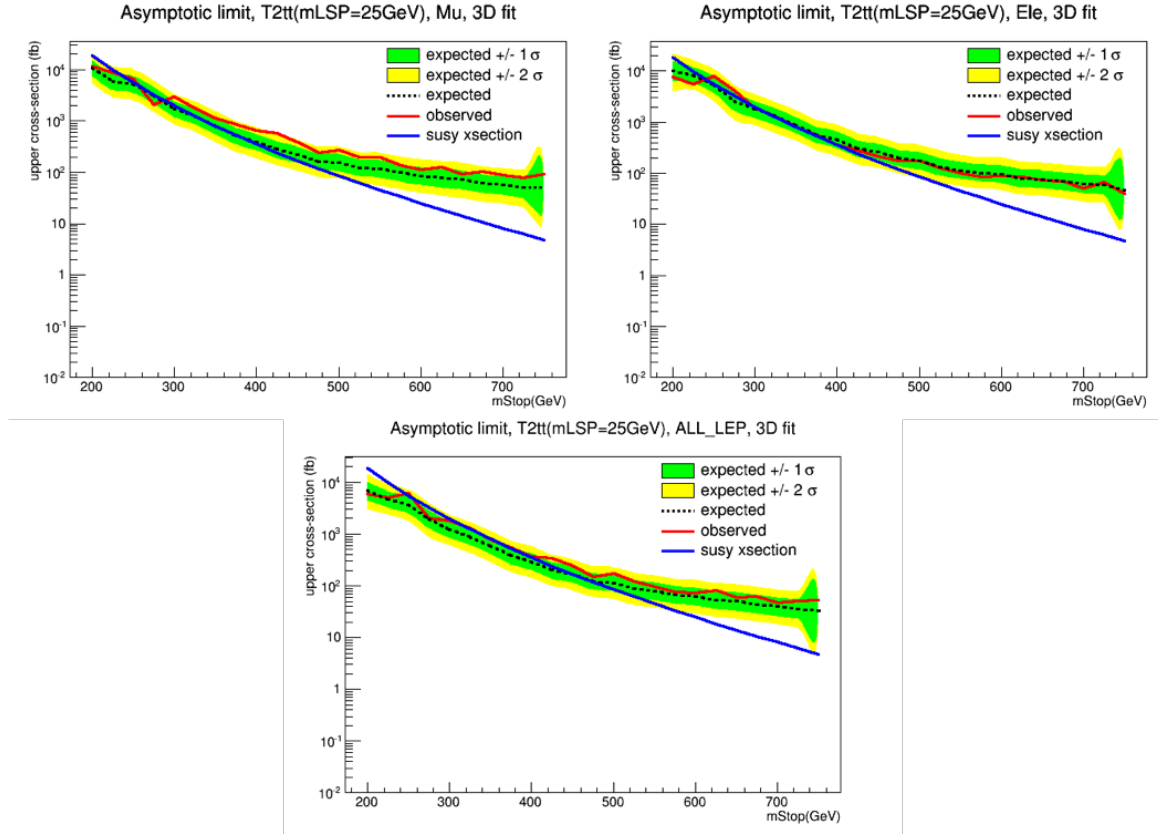


Figure 62. **Expected (black dotted with uncertainty bands) and observed (red) limits using simulated data** - The upper left plot shows the limit for the Ele box, while the upper right plot corresponds to the Mu box. The bottom center plot shows the exclusion limit when the results from the Ele box and the Mu box are combined. In all plots, the blue line gives the predicted supersymmetric cross sections. Low stop masses are excluded when both boxes are combined (the fit results used are from 4.3.3.1)

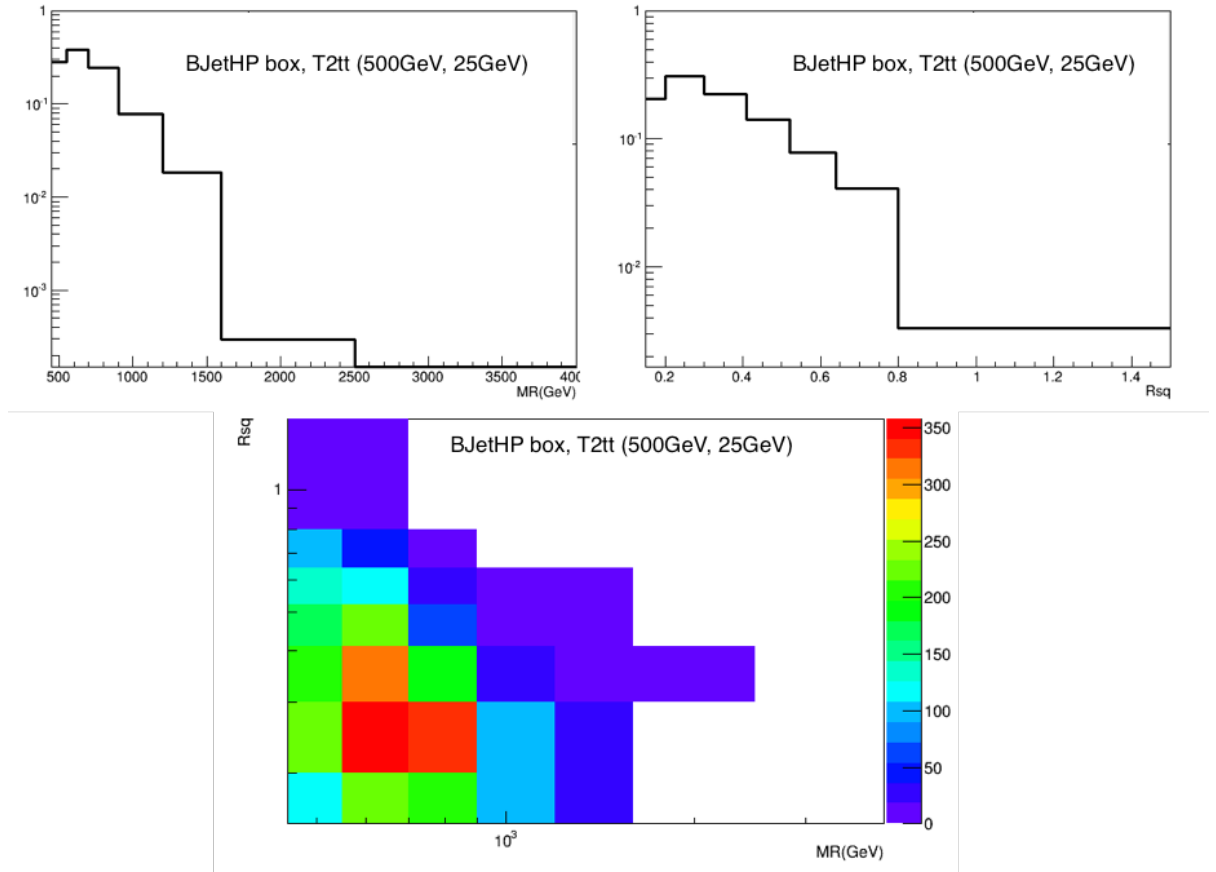


Figure 63. **T2tt** signal used for signal injection tests of the **BJetHP** box - The upper left plot shows the 1D M_R distribution. The upper right plot shows the 1D R^2 distribution. The bottom center plot shows the 2D M_R, R^2 plane; the temperature axis corresponds to the event yield in arbitrary units. In all plots the following T2tt model parameters are used: the stop mass is 500 GeV and the LSP mass is 25 GeV.

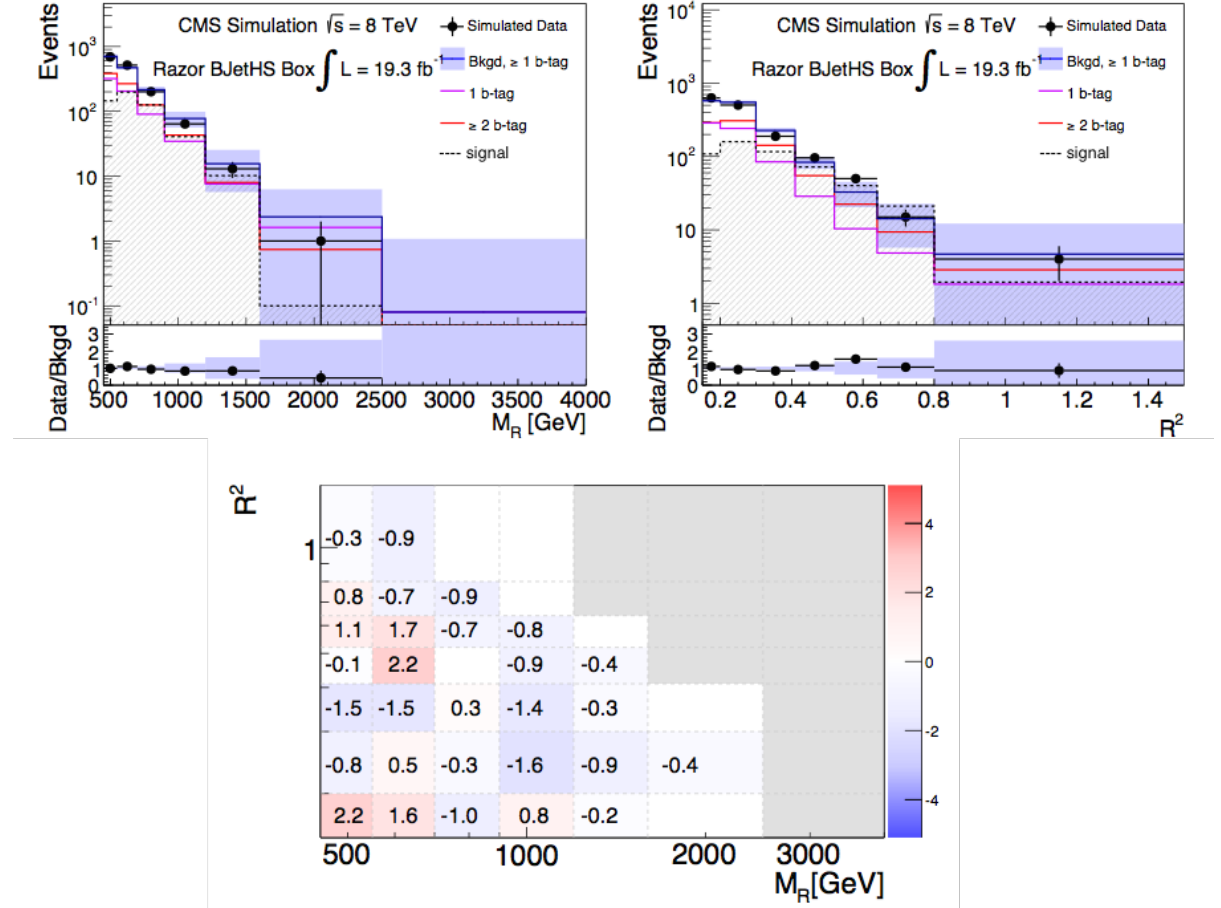


Figure 64. **Signal injection test for BJetHP box - Background only fit to a background+signal sample** - The upper left plot shows the 1D M_R distribution. The upper right plot shows the 1D R^2 distribution. The bottom center plot shows the 2D M_R, R^2 plane. In the upper two plots, the simulated data (background+signal) is shown as black points, the fitted background model is displayed as a blue line (with blue error bands from the fit), the one b-tag component of the background model is shown in magenta, the two b-tag component of the background model is shown in red, and finally the injected signal is shown as a dotted black line.

10-15% error at most. Hence, our signal injection procedure closes and confirms that the background modeling is robust against signal injection, that is it behaves as expected even in the presence of a large signal.

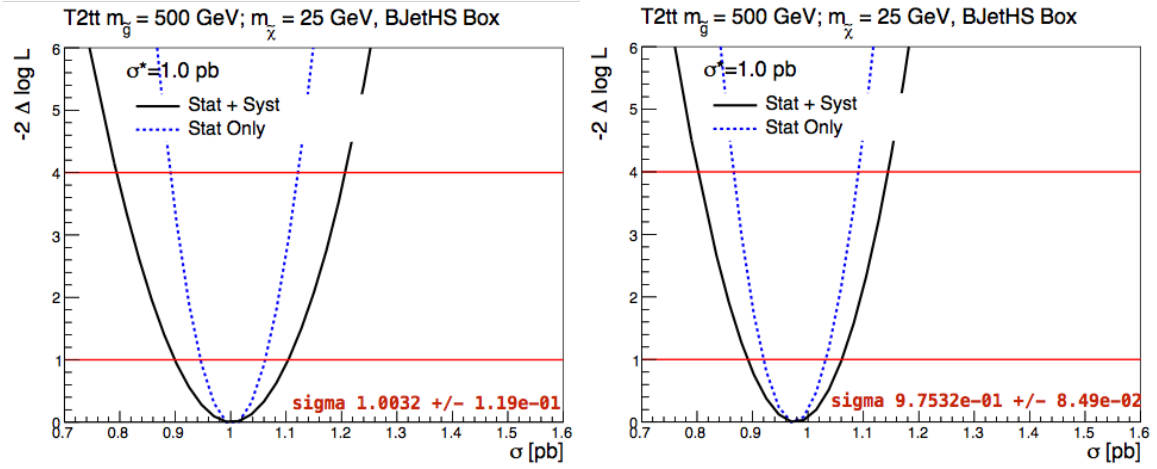


Figure 65. **Signal injection test - Finding back the injection cross section** - After building a background plus signal sample (our data points here correspond to simulated data), we fit the resulting dataset with the signal plus background probability density function, with the cross section as free parameter allowed to float. For the left plot, the cross-section is initialized to 1 pb before performing the fit, while it is initialized to 100 pb for the right plot. In both cases, the fit finds back the cross section of 1 pb quite well.

4.3.4.2.2 Leptonic boxes

For the leptonic boxes, we use a different mass point of the T2tt signal, with $m_{\text{Stop}}=200\text{GeV}$ and $m_{\text{LSP}}=25\text{GeV}$, and we test the following three different cross sections:

- the supersymmetry predicted value of the cross-section which is above our upper limit, and hence excluded,
- a much higher cross-section than the supersymmetry predicted one above (and hence even more excluded!)
- a much lower cross-sections, below the analysis sensitivity and hence not excluded.

According to Figure 62, this T2tt test model point with $m_{\tilde{t}} = 200$ GeV and $m_{\tilde{\chi}} = 25$ GeV is excluded. Thus, we check that we are able to differentiate a signal from the background, by injecting signal with the theoretically predicted supersymmetry cross-section of about 20 fb. Using the same closure procedure describe above for the BJetHP box, we can see in Table VIII that we recover the injection cross-sections quite well for high enough cross sections. Figure 66 (Ele box) and Figure 67 (Mu box) show the corresponding fit projections, French flags plots and how well we can find back the injection cross-section. Overall, the trend follows expectations: the data points and fit projections move apart, and the French flag plots are increasingly chaotic. Yet, the French flag plots, while they give a indication when the signal is there, remain limited, and it may be not so easy to determine a signal presence from these when the injection cross-section is too close to the observed limit. Two main tests indicate that the method works and closes well, without bias:

- one is that the method is able to determine the cross-sections of all injected signals very well, as demonstrated in Figure 66 and Figure 67;
- second, the expected and observed limits match for background only samples, as shown on Figure 61 and Figure 62.

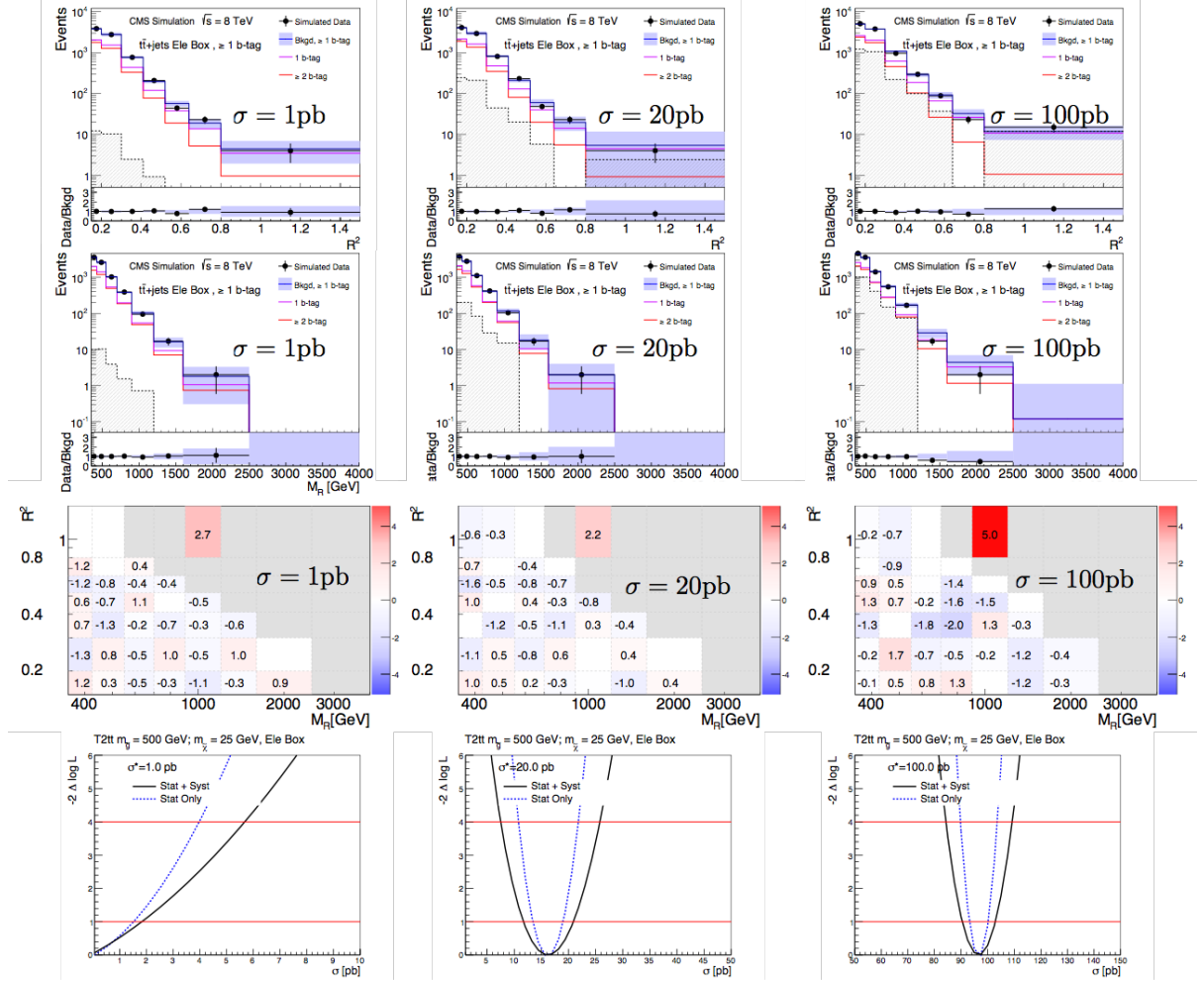


Figure 66. **Signal injection test - Ele box** - Fit projections, French flag plots, finding back the cross-section for signal injection tests. Data points here correspond to simulated data, combining $t\bar{t}$ and T2 $t\bar{t}$ events. 1pb is below the limit we obtain, 20pb is roughly at the limit we see, 100pb is above this limit. We can see the French flag plots getting more and more chaotic, with larger deviations, increasing number of blue and red cells.

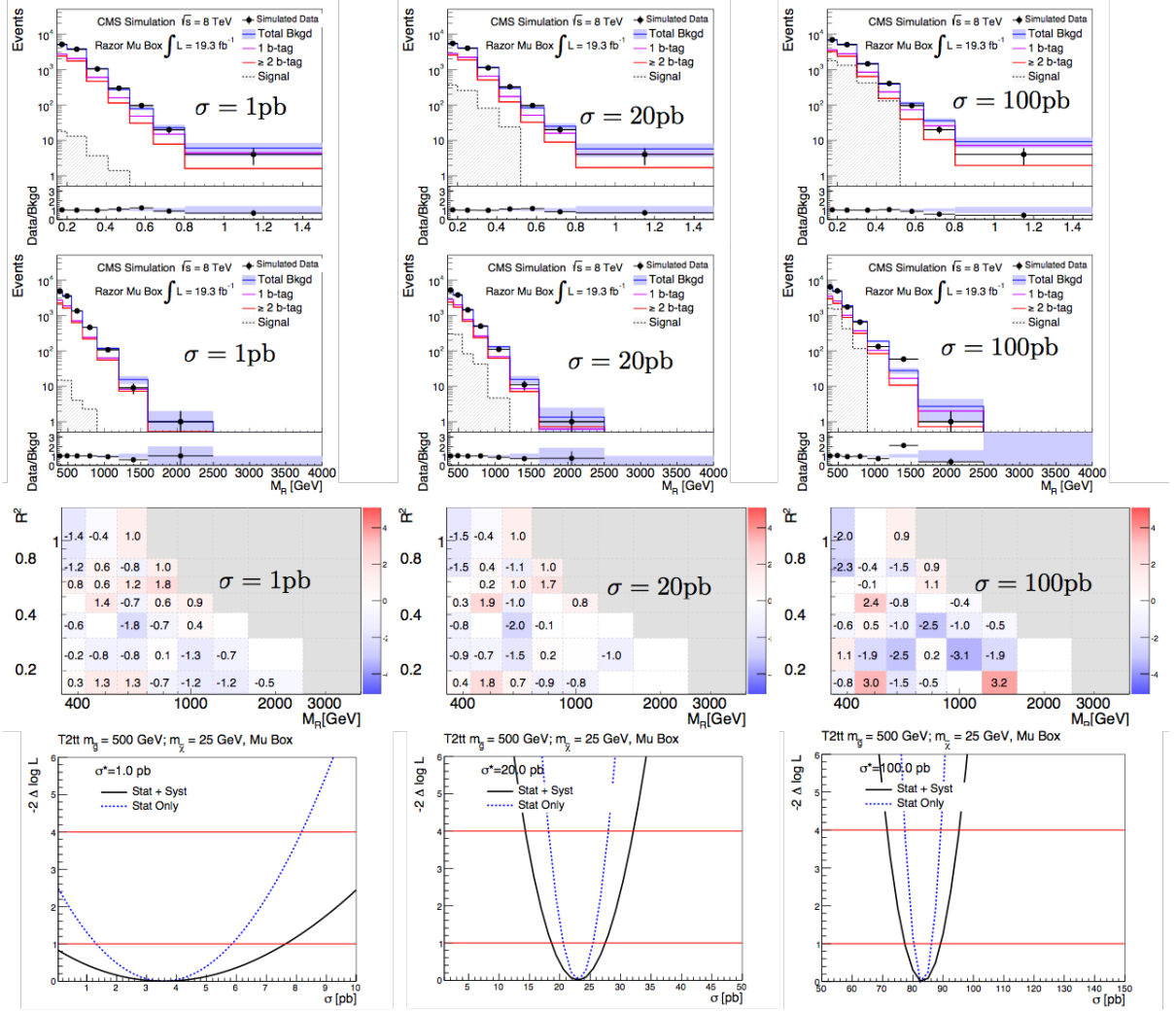


Figure 67. **Signal injection test - Mu box** - Fit projections, French flag plots, finding back the cross-section for signal injection tests. Data points here correspond to simulated data, combining $t\bar{t}$ and $T2t\bar{t}$ events. 1pb is below the limit we obtain, 20pb is roughly at the limit we see, 100pb is above this limit. We can see the French flag plots getting worse and worse.

	Injection cross section(pb)	Initial value before fit	Final value
Ele box	1.0	1.0000e+00	-9.9999e-02 +/- 3.58e+00
	20.0	1.0000e+00	2.2843e+01 +/- 4.58e+00
	20.0	2.0000e+01	1.6191e+01 +/- 4.63e+00
	100.0	1.0000e+02	8.8799e+01 +/- 6.16e+00
Mu box	1.0	1.0000e+00	3.6575e+00 +/- 4.14e+00
	20.0	2.0000e+01	2.3001e+01 +/- 4.49e+00
	100.0	1.0000e+02	9.2603e+01 +/- 5.75e+00

TABLE VIII

SIGNAL INJECTION TESTS, LEPTONIC BOXES. AFTER BUILDING A BACKGROUND PLUS SIGNAL, WITH A TRIAL CROSS-SECTION, SAMPLE, WE FIT IT WITH THE SIGNAL PLUS BACKGROUND PROBABILITY DENSITY FUNCTION, ALLOWING THE CROSS SECTION TO FLOAT. THE FIT RESULT RETURNS BACK THE INJECTION CROSS-SECTION QUITE WELL FOR VALUES FOR WHICH WE HAVE SENSITIVITY, THAT IS, FOR WHICH WE ARE ABLE TO DISTINGUISH SIGNAL FROM BACKGROUND.

4.3.5 Data fits and results

Having tested and validated the fit on simulated events samples, we now use our background model to fit the real data and estimate the background from Standard Model processes. These fit results will later be used in the CLs procedure to place observed limits on possible signal strengths in the data.

Figure 68 shows the fit result for the data in the BJetHP box, including the 1D projections for the razor variables and the 2D French flag plot. The functional form for the background models well the R^2 and M_R distributions, and the French flag plot is quite smooth. Hence the background model seems to describe the data well.

Figure 69 shows the fit result for data in the Ele box, including the 1D projections for the razor variables and the 2D French flag plot. The fit projections suggest that the background functional form models the data well, and the data is seen to fall within the fit error bands. The French flag plot is not

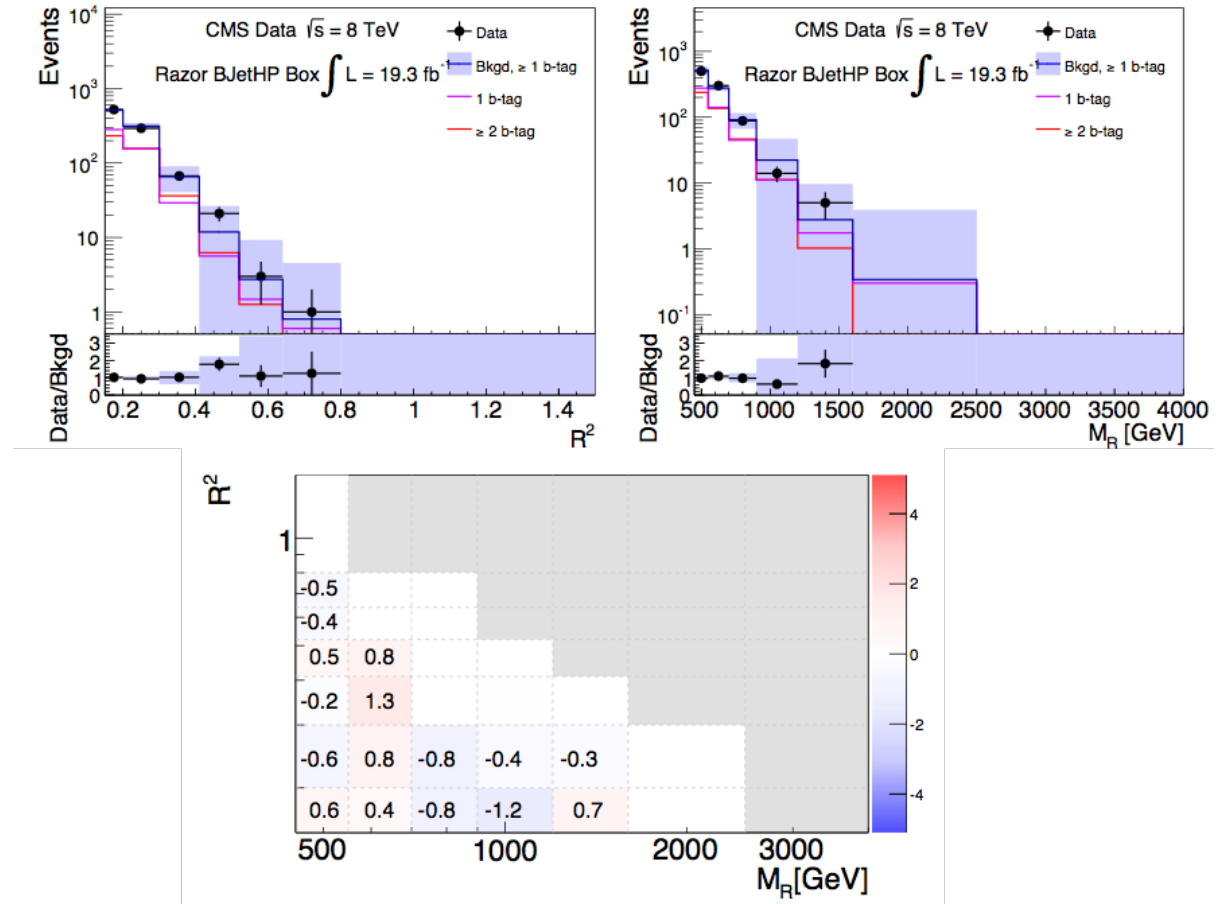


Figure 68. **Fitting our data sample after the BJetHP box selection with the background only model** - Fit result for the data in the BJetHP box, including the 1D projections for the razor variables (upper row) and the 2D French flag plot (lower row). The functional form for the background models well the R^2 and M_R distributions, and the French flag plot is quite smooth. Hence the background model seems to describe the data well.

quite as smooth as in the BJetHP case, but overall the background model seems to describe the data reasonably well.

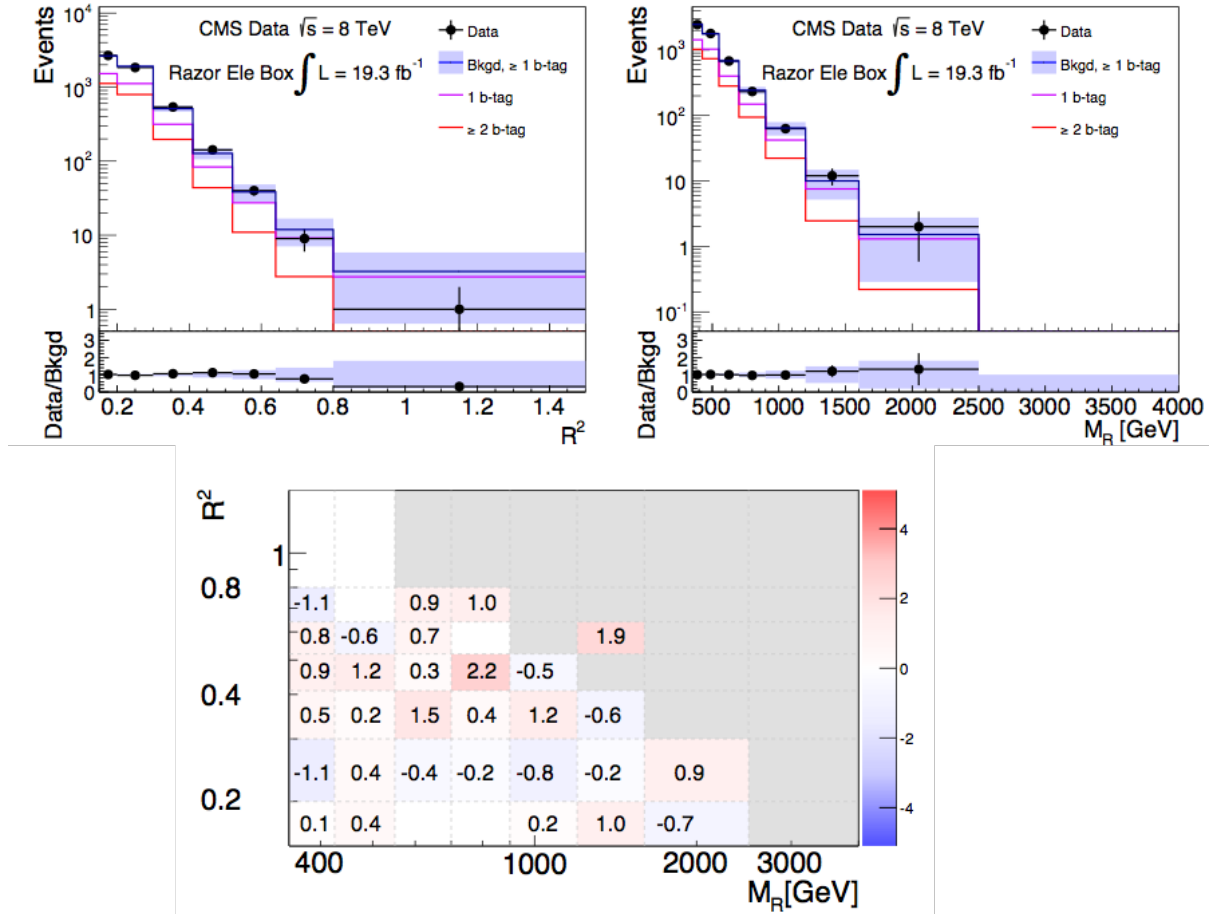
Figure 70 shows the fit result for data in the Mu box, including the 1D projections for the razor variables and the 2D French flag plot. The fit projections suggest that the background functional form models the data well, and the data is seen to lie within the fit error bands. The fit projections accommodate the data well. While some fluctuations can be seen in the French flag plot, they are within expectations from statistical fluctuations and the background model seems to describe the data well.

4.3.5.1 Limits on T2tt with our samples and analysis

Since Figure 68, Figure 69 and Figure 70 indicate that our background model fits the data well, we proceed to derive exclusion limits on allowed upper cross sections for the T2tt model.

Figure 71 shows the observed limit (red line) from data for the BJetHP box, together with the expected limit (black dotted line with green and yellow uncertainty bands) from simulation, for an LSP mass of 25 GeV. The observed exclusion limit is lower than the theoretically predicted value for stop masses, ranging from 200 GeV up to 700 GeV. Yet, we observe an $\sim 2\sigma$ excess around 500 GeV compared with the expected limit. This excess is small enough to keep the allowed cross section below the theoretically expected value for small LSP masses, but as we scan to higher LSP masses, where our efficiency to select signal becomes worse, we will not be able to exclude some models for which we expect to have sensitivity based on simulations.

For the leptonic boxes, we obtain the exclusions displayed in Figure 72. The upper left plot shows the results from the Mu box, which excludes stop masses up to about 400 GeV, for an LSP mass of 25 GeV. The results from Ele box, shown in the upper right plot, alone are too weak to bring much exclusion, but combined with the Mu box, stop masses up to 500 GeV are excluded. In all plots, the observed limit from data is shown as a red line, and the expected limit from simulation is shown as



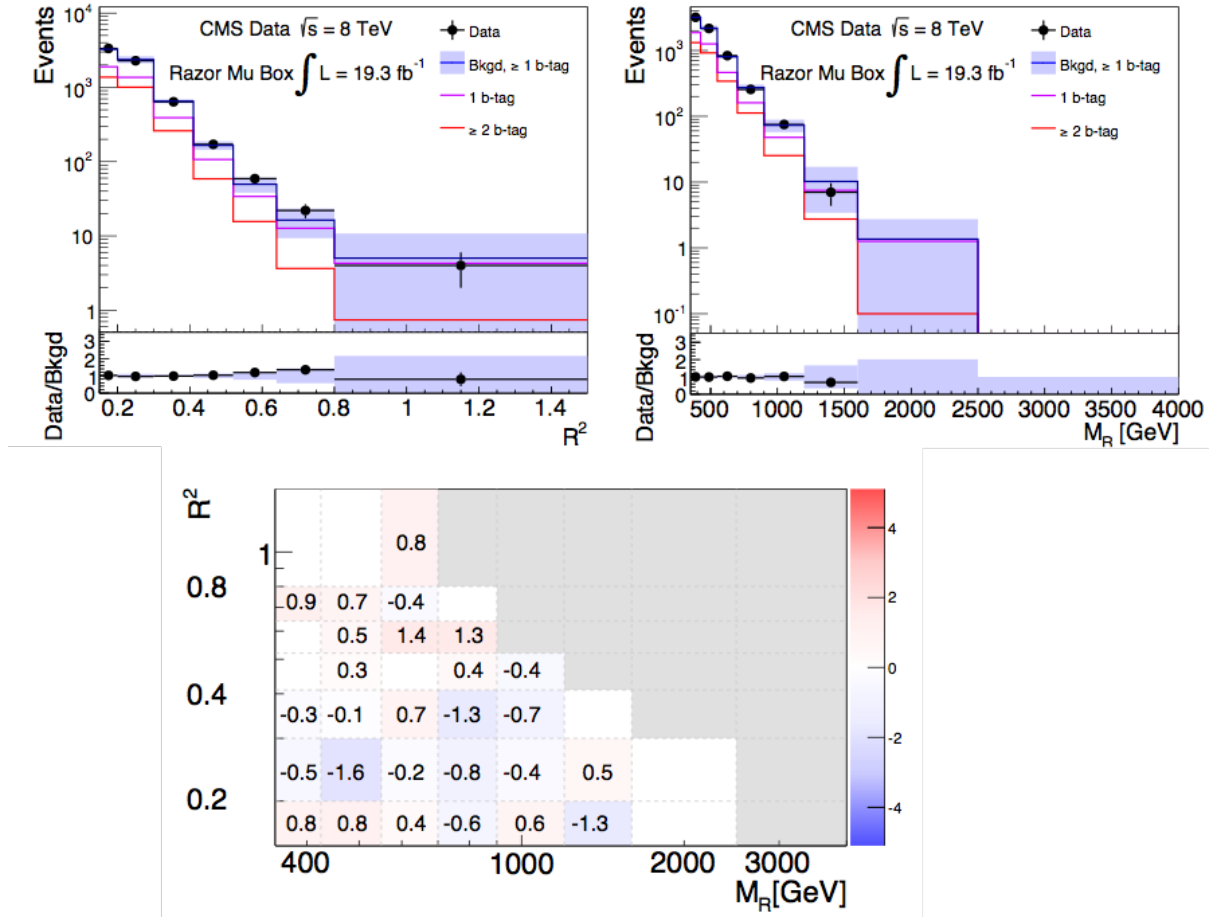


Figure 70. **Fitting our data sample after the Mu box selection with the background only model** - Fit result for data in the Mu box, including the 1D projections for the razor variables (upper row) and the 2D French flag plot (lower row). The fit projections suggest that the background functional form models the data well, and the data is seen to lie within the fit error bands. Overall the background model seems to describe the data well.

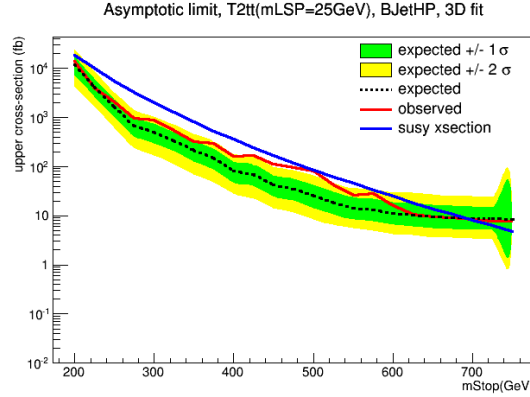


Figure 71. **Exclusion limits for $m_{\text{LSP}}=25\text{GeV}$, T2tt, BJetHP box** - Cross-section lower limits at 95% CL as function of stop mass, for fixed LSP mass. The red line represents the observed exclusion limit from data. The dotted black line, together with the green (1σ) and yellow (2σ) uncertainty bands, represent the expected exclusion limits from simulation. The blue line represents the theoretical prediction from the T2tt model.

a black dotted line with green and yellow bands, corresponding to the 1σ and 2σ uncertainties on the expected limit.

Comparing the BJetHP and leptonic boxes, the leptonic boxes have sensitivity at low stop masses, but the hadronic box is more useful at high stop masses. Hence, combining all boxes together still provides an exclusion limit on all stop masses up to 700 GeV, as shown in Figure 73. In the combination, the excess around 500 GeV is attenuated somewhat, thanks to the leptonic boxes, but does not disappear.

Repeating the same procedure as above, but varying the LSP mass, we can draw exclusion limits in the (stop, LSP) 2D mass plane. Figure 74 shows the combined exclusion range in the 2D (stop, LSP) mass plane, for all boxes together. The observed limit is displayed with a continuous black line, while the expected limit is shown in dashed line together with the 1σ and 2σ uncertainties on the expected limit. The colored temperature axis corresponds to the excluded upper limit cross-section at 95% CL

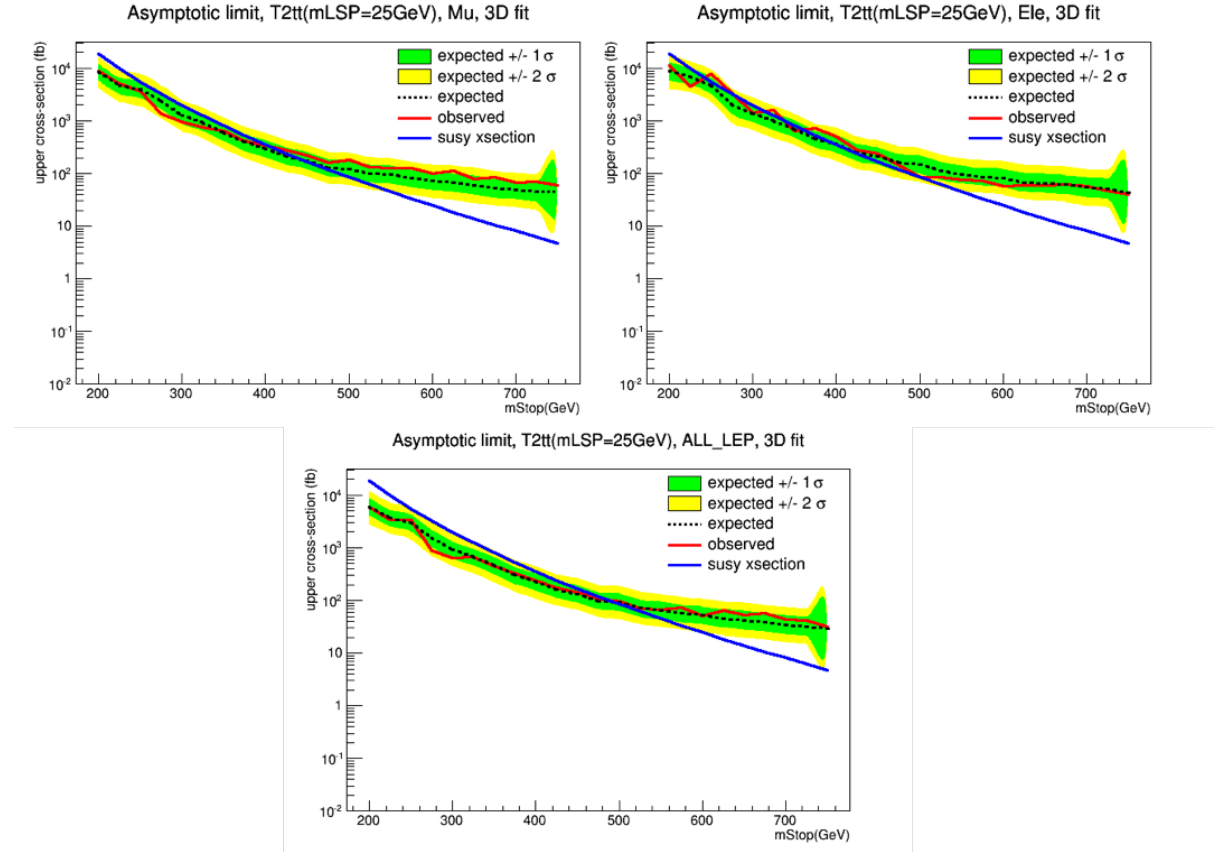


Figure 72. **Exclusion limits for $mLSP=25GeV$, $T2tt$. Mu and Ele boxes are shown separately and combined** - Cross-section lower limits at 95% CL as function of stop mass, for fixed LSP mass. Upper left: Mu box. Upper right: Ele box. Lower center: Combined results from Mu and Ele boxes. The red line represents the observed exclusion limit from data. The dotted black line, together with the green (1σ) and yellow (2σ) uncertainty bands, represent the expected exclusion limits from simulation. The blue line represents the theoretical prediction from the $T2tt$ model.

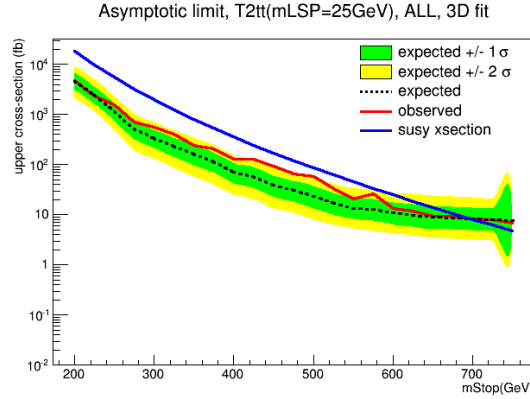


Figure 73. **Exclusion limits for $m_{\text{LSP}}=25\text{GeV}$, T2tt, BJetHP, Mu and Ele boxes combined** - Cross-section lower limits at 95% CL as function of stop mass, for fixed LSP mass. The red line represents the observed exclusion limit from data. The dotted black line, together with the green (1σ) and yellow (2σ) uncertainty bands, represent the expected exclusion limits from simulation. The blue line represents the theoretical prediction from the T2tt model.

(in fb). The analysis is sensitive to models with an LSP mass up to 250 GeV, but as anticipated in the discussion of BJetHP results of Figure 71, the analysis is unable to exclude models with an LSP mass above ~ 150 GeV, with a statistical significance of about 2σ .

4.3.5.2 Comparison with existing results

Figure 75 shows a summary of exclusion limits for top quark pair production obtained by CMS and ATLAS searches for direct stop production. If we compare to Figure 74, we see that the expected sensitivity for all of the boxes combined is similar to the best CMS analysis for the on-shell top-quark region (when the difference in the stop mass and the LSP mass is equal to or greater than the top-quark mass): our exclusion range goes from 200 GeV up to 700 GeV for stop masses, and up to 250 GeV for LSP masses, while CMS searches reach at most 800 GeV for stop masses and barely 300 GeV for the upper LSP mass. Moreover, this analysis is also close to the ATLAS sensitivity, which goes up to an

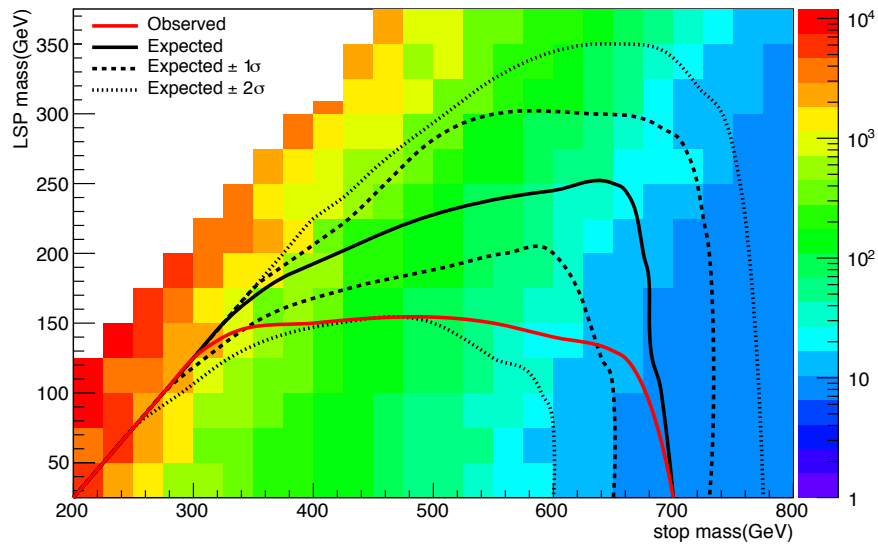


Figure 74. **Combined Exclusion Limits in the 2D stop mass - LSP mass plane for all boxes** - The observed limit is displayed with a continuous black line, while the expected limit is shown in dashed line together with the 1σ and 2σ uncertainties on the expected limit.

The temperature axis corresponds to the excluded upper limit cross-section at 95% CL (in fb). The analysis is sensitive to models with an LSP mass up to 250 GeV, but as anticipated (see text) the analysis is unable to exclude models with an LSP mass above ~ 150 GeV, with a statistical significance of about 2σ

LSP mass of about 300 GeV and a stop mass of 720 GeV. However, due to the small excess seen in the BJetHP box near stop masses of about 500 GeV, the observed exclusion limits on the LSP mass die out near 150 GeV for this analysis. The significance of the observed excess, and hence the compatibility of this result with other results from ATLAS and CMS, is about $\sim 2\sigma$ and is consistent with a statistical fluctuation of Standard Model backgrounds.

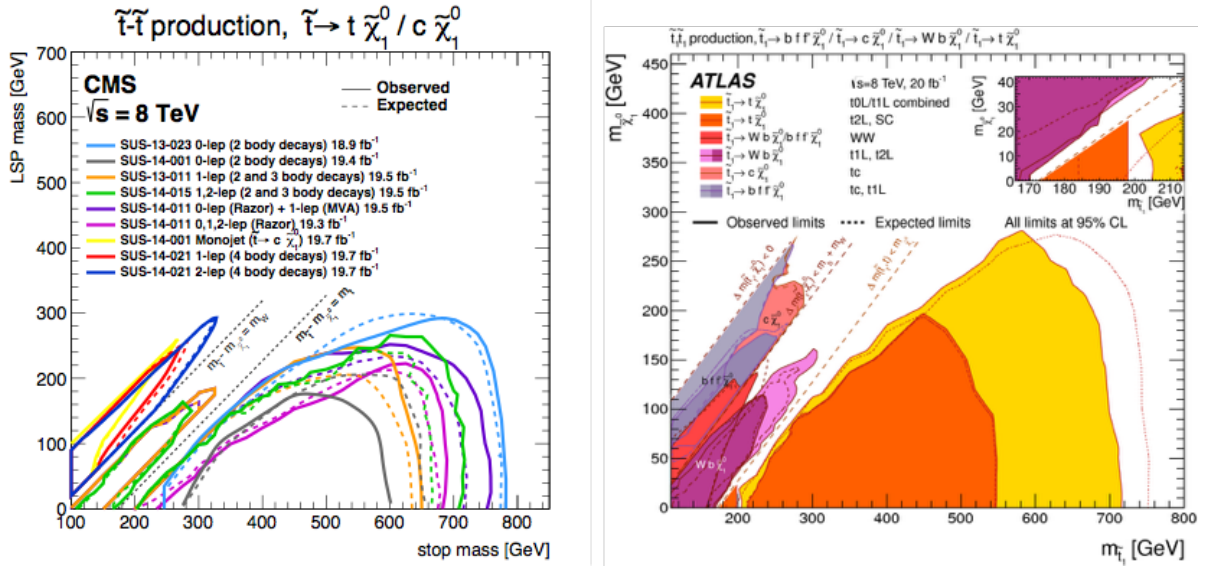


Figure 75. **Existing CMS and ATLAS results with 8TeV data** - Exclusion limits for stop production, obtained with 8TeV data, by CMS searches (left plot, (73)) and ATLAS searches (right plot, (74)). Both experiments have similar ranges, excluding stop masses up to 700-800GeV, and LSP masses up to 300GeV. Our expected limits are not so far from these, but our observed limits are lower.

CHAPTER 5

CONCLUSIONS

5.1 Summary

A search for supersymmetric particles has been developed. Focusing on third generation spartners, expected to be lighter compared to the first and second generations squarks, the present analysis excludes their existence for a wide mass range. The available phase space for such particles is narrowing, but they still have a chance.

The analysis work performed with the CMS detector relies on the large amount of data collected. To do so, we need a data selection, or trigger, system, which has been detailed in chapter 3. This system required careful monitoring and development during the first run of the LHC to maintain an efficient data taking, making the most of the computing resources to search for new physics.

5.2 A human adventure - final word

In the small world of particle physics, CERN and the Large Hadron Collider have grown as a *the place to be* in the recent years. After decades of development, the start of the LHC brought lots of excitement. An international collaboration, an unprecedented energy, gigantic machines...to answer questions about the tiniest scale, but also on the much farther, remote galaxies and cosmological objects, our Universe: all the ingredients are present to make a lively and buoyant cocktail.

The discovery of the Higgs boson announced in 2012 brought a key piece of understanding of the particles' world, finally providing evidence for the missing link in the Standard Model. Yet, many questions remain. While supersymmetry is not yet ruled out, this elegant theory keeps evading us. The hope is that the new energy frontier reached by the LHC during its second run will allow us to give a final word, thus paving the way for what physics beyond the Standard Model might actually be.

Although scientists at CERN are dealing with highly fundamental topics, connections with the outside world have grown along with the LHC. Visitors come by thousands each year, as young as 8-9 years old. Outreach activities are developing, as a natural spinoff to the public funding us. As a CMS guide, my goal was to find a balance between providing pieces of understanding and fueling enthusiasm for this extraordinary world.

On a daily basis, a machine remains a machine: no matter how big the beast is, it has to work. The data taking at P5, the CMS experimental site, easily keeps awake at night: I never got bored about seeing fireworks on the screens, enjoying the team work, in all possible languages, *just to see it work....*

Six years in the making. A fantastic experience. A chapter in my life...that continues elsewhere.



Figure 76. **Refuge du Couvercle -**

L'absurdité des mythes, des généralisations et des systèmes quels qu'ils soient, c'est la rupture qu'ils supposent avec le monde vivant.

J-M.G. Le Clézio.

APPENDICES

Appendix A

WELCOME TO CMS!

[DETECTOR CHAPTER - A CMS GUIDE VERSION...]

*When asked about why I chose particle physics, the true story usually sells well : “When I was 15, I read a novel, *Alice in Quantumland*, by Robert Gilmore, and I thought : wow, that’s what I wanna do! And here I am.” What that meant was unclear : roughly, something about quantum mechanics, although I had a very cloudy idea about what it could refer to. Years after, already quite advanced in my PhD, I would often think that there was a mismatch between that and my daily activities: boring papers, endless meetings, poorly designed code, inertia of a 4000 people collaboration...so I started to write my detector chapter with Alice in mind, Alice exploring and questioning and complaining that grown-ups are really weird persons. Taking Alice for another stroll maybe was the way to find back the excitement, and this is probably what I do when I guide people through the CMS experimental site. Hope it works...for the writer, for the reader.*

In the following section, I invite you to follow suit a group of visitors, among whom a young girl named Alice, and join a tour to CMS...

Upon her arrival at CERN, Alice joins a group of French high school students¹ ready to leave for a visit to some place called *CMS*. After a twenty minutes drive through the French countryside, their bus arrives at a gate. “Welcome to P5!”, says their CERN guide. *P5* sounds just as romantic as the names nowadays astrophysicists give to remote stars. Some warehouse-like buildings, a big tent...nothing exciting at first glance. They are met with a Christmas tree-looking guide, decorated with colorful badges,

¹Who should know a bit about particle physics ;)... Things evolved since I was a high school student.

Appendix A (Continued)

a helmet, some sort of necklace carrying an unknown device, an ID badge, a not-so-normal either key and wearing an oversized jacket. After distributing the badges to the visitors, she walks them into a warehouse and starts:

“It’s slightly warmer and less windy inside. Welcome to CMS, my name is Lucie, I’m a PhD student and I’ve been based at CERN for about three years now. Please feel free to take pictures, ask questions, even if it’s to say ‘*I understood nothing please start from scratch*’. Actually, *I* am going to ask you lots of questions... and, to begin with: where are we?”

“Somewhere in the middle of nowhere?”

“No, that’s Fermilab... ”, the guide winks.

“In France”,

“Yes, true, most of the LHC is underneath France, not Switzerland.”

“We’re at the LHC...”

“...hmmm not exactly (but this is what I wanted to hear, to make one point clear). The LHC, which stands for Large Hadron Collider, is the apparatus providing us collisions of hadrons. Most of the time the hadrons circulating within the LHC are protons: the LHC accelerates them to very high energies, extremely close to the speed of light, and collides them. There are only four points along the LHC where we get collisions: at these points, we put some sort of cameras (or detectors, experiments) to observe the collisions. What you’re about to visit is one of these cameras, called CMS.”

“In fact, the LHC is only the last step of acceleration. It all begins within a small bottle of hydrogen - by the way, what is hydrogen?”

Alice remembers from her very first journey into particles’ world:

“It is the simplest atom: one proton, one electron.”

“Yes! Now, what we want to accelerate at the LHC are the protons, so, we *just* need to strip off the electrons from the hydrogen atoms, and we obtain the protons.”

Appendix A (Continued)

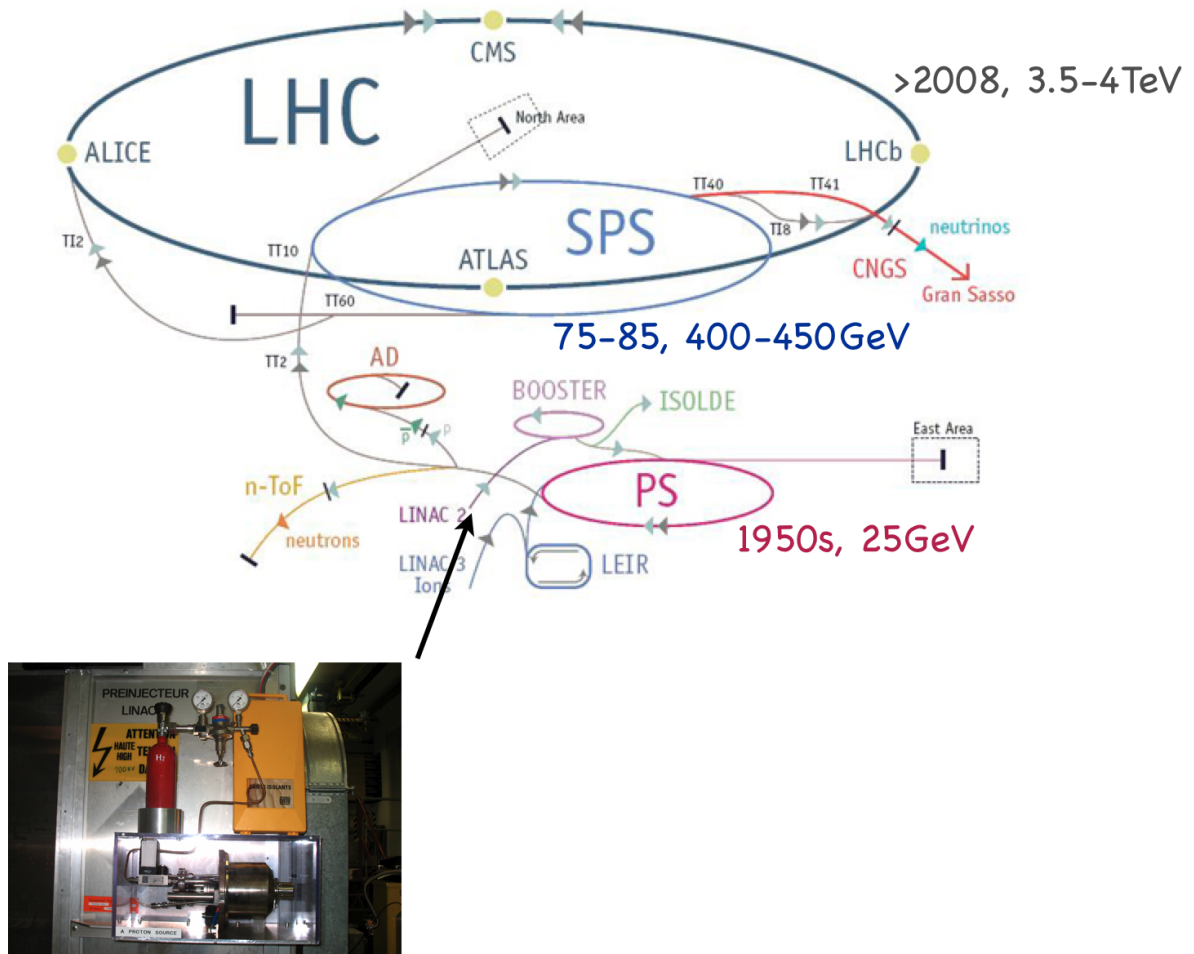


Figure 77. **Recycling & developing: the LHC accelerator complex** - The LHC is only the last step of acceleration! Sustainable development @CERN. ©CERN 2008.

Appendix A (Continued)

“Afterwards... could you please start the LHC?”, she asks Alice. Alice frowns, the guide smiles:

“Please just go and press the grey button over there. Thanks, here we are: you now see the protons on the poster (Figure 77...which is not actually reproducing the poster, for I never took pictures at P5...).” Indeed, on a large picture displaying an aerial view of the area, extending from the Jura mountains to Geneva, they can now see red dots moving, materializing (bunches of) protons.

“After extraction, we send the protons to a linear accelerator, where they are brought to an energy of 50MeV¹ before they enter the Proton Synchrotron (PS).

The PS is an accelerator from the 50s, one of the very first accelerator built at CERN, and you see, it is still in use as an injector for today’s experiments. The protons reach an energy of 26GeV in the PS, after what they continue their journey to the Super PS (SPS). The SPS, built in the 70s, increases even more the protons’ energy, up to 450GeV. At this energy, they are now allowed to jump on the real highway, the LHC, a twenty-seven kilometers long ring, where they acquire an energy as high as 4TeV. ”

How fast is this?

Defining:

¹an Electron Volt or eV corresponds to the amount of energy acquired by an electron moving through a potential difference of 1V. MeV stands for Mega eV = 1 000 000 eV. It is the most commonly used unit in particle physics, for energies ... and mass, thanks to the equivalence between mass and energy. A proton weights about 1GeV.

Appendix A (Continued)

E : energy of a proton, m : proton mass, c : speed of light, p : proton momentum, γ : Lorentz factor, v : proton velocity.

$$\begin{aligned}
 E^2 &= m^2 c^4 + p^2 c^2 \\
 \gamma^2 m^2 v^2 c^2 &= E^2 - m^2 c^4 \\
 \gamma^2 v^2 / c^2 &= \frac{E^2 - m^2 c^4}{m^2 c^4} \\
 \frac{v^2 / c^2}{1 - v^2 / c^2} &= \frac{E^2 - m^2 c^4}{m^2 c^4} \\
 v^2 / c^2 &= \frac{1}{1 + m^2 c^4 / (E^2 - m^2 c^4)} \\
 v / c &= 0.99999993750000.....
 \end{aligned}$$

...so extremely very close to the speed of light...

“Two beams circulate in the LHC, in opposite directions. And now...collisions!” They can now see the red dots moving and *colliding*, as if exploding to blue and orange dots at four points along the LHC circle.

“At four points along the LHC both beams cross and some protons collide in: ATLAS (A Toroidal LHC ApparatuS) next to the Globe, ALICE (A Lead Ion Collider Experiment) towards the Jura mountains, CMS, at the very top, and LHCb near Ferney-Voltaire. At the moment, we are standing on very thick concrete blocks, below which dives a huge shaft, a hundred meters deep, and all the way down there is our camera, CMS. Question: why has the LHC been built so deep? (see Figure 78). There are three reasons, or advantages, to construct the LHC all the way down there...any idea?”

“To avoid radiations (?)”

“Sort of...which kind of radiations?”

“For people around...”

“Hmmm....Not really. We do not produce long-lived radioactive elements in the collisions per se, we are not playing with heavy nuclei. A few parts of the detector have been made with heavy materials that may be activated by particles emitted in the collisions. At the moment, these pieces of the detector

Appendix A (Continued)

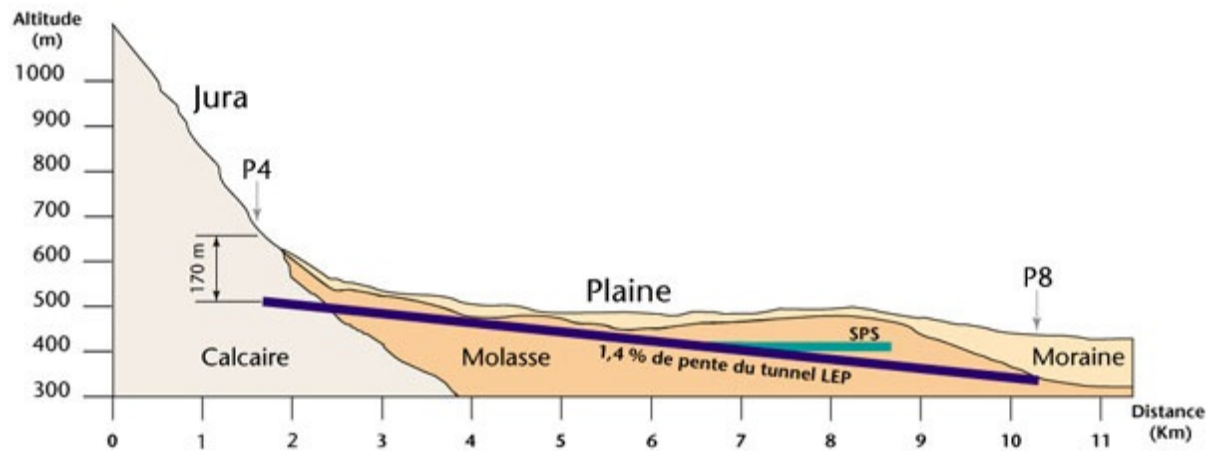


Figure 78. **Where is the LHC?** - The LHC tunnel lies between fifty and a hundred and seventy-five meters below the surface. It is tilted, to avoid a “hole” towards Ferney-Voltaire, where the molasse is buried much deeper. ©CERN.

have been removed and stored into sealed tanks, so that we can access the cavern - not only for visits, but also & mostly because there are people working in the experimental cavern. Besides, between this experimental cavern and what we call the service cavern (and, in fact, all around the experiment), there is a seven meters thick concrete wall which allows us to go down in the service cavern even when there is beam in the machine. So, with a lot of concrete, we can isolate the environment from radiations from the machine. Yet, being about a hundred meters below may help with *cosmic* radiations. Indeed, we are constantly bombarded with cosmic rays, but, what we want to study here are LHC collisions, not cosmic rays. Fortunately, we receive much fewer of them down in the cavern.

Appendix A (Continued)

More importantly, the ground down there is much more suitable to build a tunnel: homogeneous, neither too soft nor too hard¹.

And, last, look at the picture of the area: villages and fields, there is no room to build the LHC on the surface!

I started by asking ‘where are we?’, but I barely answered the question. When asking this, the first point I want to make clear is that the LHC is the device accelerating and colliding particles, while CMS is one of the cameras taking pictures of the collisions. Having said that, I can now refine: we are in the hall where CMS has been assembled. Think about it: fifteen years ago, here, there was nothing but fields! Pieces were built in the various institutes participating in the collaboration and then brought here, while workers were still building the shaft. However, our detector is more than twenty meters long, about fifteen meters in diameter (the picture you can see here is a real size one - Figure 79), and weighs fourteen thousands tons: there is no crane strong enough to lower down such a monster. Thus, the detector has been sliced into eleven parts, some of which still weighting up to one thousand tons. Then, the detector was lowered down slice by slice. Another advantage of this design is that now we can open the detector both for maintenance...and visits: we actually get to see the guts of the beast. Questions?”

She leads them out of the hall, into a tent next to it, stopping by a model of the CMS detector (Figure 80 for a view of the model, but Figure 81 provides a more reader-friendly view.). “Here we have a small version of our more than twenty meters long, fifteen meters in diameter camera, which makes it easier for me to describe you the structure of the detector. The collisions take place at the very center of the experiment (note that I’m using equivalently the words ‘camera’, ‘detector’, ‘CMS’ and

¹Between the Jura mountains and the lake, the ground contains a layer of *molasse* provided we dig deep enough below the soft loam & sand. This molasse, resulting from the erosion of the Alps, is more suitable to build a tunnel than the upper layers, which also tend to be quite watery.

Appendix A (Continued)

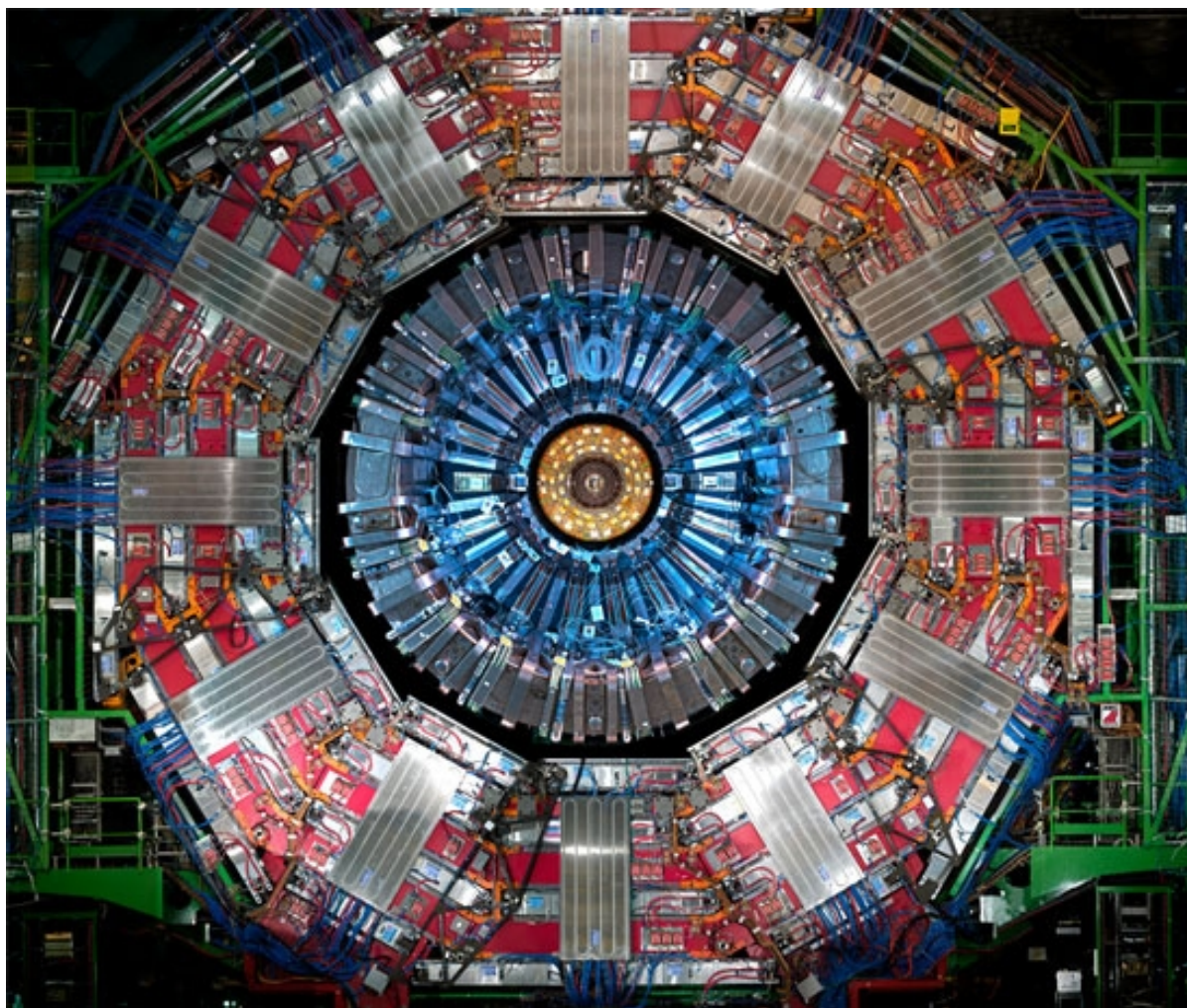


Figure 79. **A cut view of the CMS detector (Image Michael Hoch/Maximilien Brice)** - The detector is a 30m long, 15m in diameter cylinder.

Appendix A (Continued)

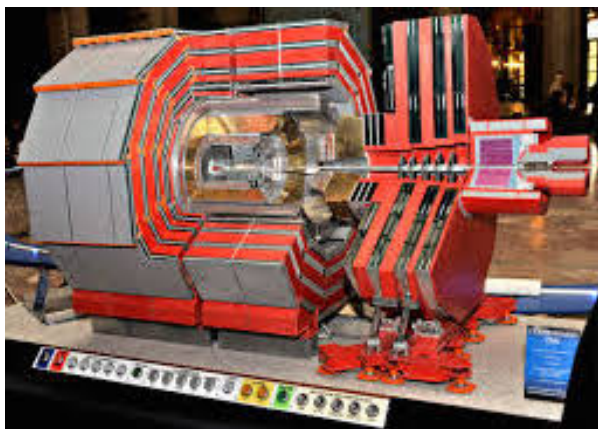


Figure 80. **A model of the CMS detector.** - With this model, we can visualize the different layers of our *onion-like* detector.

‘experiment’). By the way, why do we collide particles? What is the principle behind collisions?”

“To break them...”, “for the energy...”

“I prefer hearing about energy: the principle behind a collision is Einstein’s mass-energy equivalence, you know, $E = mc^2$, mass is energy, energy is mass. If we provide a lot of energy, as we do when colliding protons, we do not really break them into sub-pieces. Rather, we attempt to provide the energy needed to produce new particles, that were not there initially. However, these new particles that we may produce are extremely unstable - their lifetime may be as short as ten to the minus twenty-something. Thus, as soon as they are produced, they decay right away into more stable particles: we cannot directly see, say, a Higgs boson, we see its decay products. We collect those and out of them we try to reconstruct the puzzle back to the mother particle.”

“To observe the decay products themselves, and identify them, the detector relies on several layers, or subdetectors, each with a different purpose. The innermost subdetector is the pixel: only ninety centimeters long, extending to a radius of ten centimeters, but containing sixty-six millions channels!

Appendix A (Continued)

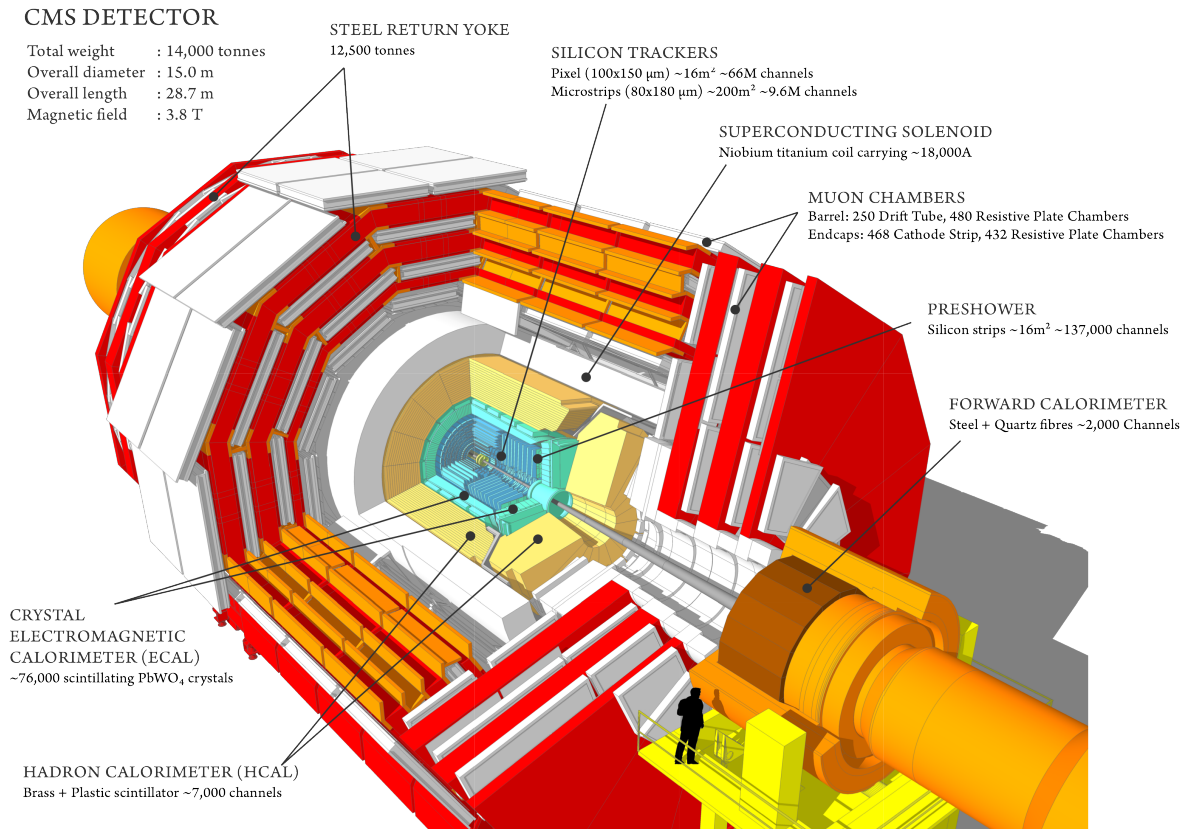


Figure 81. **A schematic view of the CMS detector.** - An easier way to visualize CMS layers, starting from the tracking system (pixel and strips), followed by the calorimeters (ECAL and HCAL), the magnet or solenoid, and the muon chambers.

Appendix A (Continued)

Indeed, since it is the innermost part, this is where the density of particles is the highest, where we need a very high detector granularity so as to avoid that two particles from the same collision hit the same pixel.”

“Outerwards, we find the different parts of the silicon strips tracker: inner barrel, inner disks, outer barrel, outer disks (different parts of the model lighten up as the guide presses buttons while mentioning these names). Instead of pixels, it is made of silicon strips. Pixel & tracker constitute the tracking system: whenever a charged particle goes through one layer of these detectors, it leaves a signal¹. Since we have several layers - three for the pixel, ten for the silicon strips tracker - we have several points with which to reconstruct the track of the particle. We then measure the curvature of the track to derive two things, can you guess what?”

“...”

“First, depending on whether the bending is clockwise or counter-clockwise, we may determine the electric charge. Then, the curvature also allows to measure the momentum of the particle - thanks to Lorentz’s force, in case you know it...Actually, what is momentum? ”

“...mass times velocity...”

“yes, in classical physics, it’s mass times velocity. Since our particles are essentially relativistic there is an extra gamma factor...but why do we measure the momentum?”

“It’s conserved...”

“Indeed, very good, that’s somehow the idea. You remember Einstein’s equation?”

“ $E = mc^2$ ”

“Yes, mass is energy, energy is mass - up to a c^2 constant that we happily forget about. In fact, the complete version is a bit longer: $E^2 = p^2 + m^2$, energy is mass plus momentum. Therefore, when we

¹creating electron-hole pairs, hence a current, in the semiconductors.

Appendix A (Continued)

want to identify a particle - and, in first approximation, for massive particles, there is only one particle at a given mass - so, when we want to identify a particle, we need some information in energy, and some information in momentum.

In practice, it is even more refined: what I just explained applies to heavy particles, which we do not directly see but attempt to “reconstruct” via their decay products. The light particles that we do see in our detector are highly relativistic though: 99.999...% of their energy is kinetic energy, their mass accounts for nothing. For these, momentum and energy are pretty much the same. Yet, those having a high energy but on the lower edge exhibits nicely curved tracks, while having more troubles reaching the calorimeters. For these, it is easier to measure their momentum, rather than their energy. On the higher edge, the tracks are mostly straight and it’s easier to measure the energy. In between, both measurements are complementary....”

Momentum measurement.

In a magnetic field of strength B , Newton’s second law yields:

$$\frac{mv^2}{R} = qvB$$

$$R = \frac{mv}{qB}$$

$$R = \frac{p}{qB}$$

$$p = qRB$$

“So, to recap: the first layers of the detector, pixel & tracker, are meant to measure the charge and the momentum. I already hinted at the next ones: calorimeters. Any idea about what are these calorimeters measuring?”

“...”

“Calori-meters are devices meant for energy measurement. We have two calorimeters, the innermost being the electromagnetic calorimeter. For which kind of particles do we use this calorimeter?” The

Appendix A (Continued)

guide repeats, distinctly splitting the syllables : “Electromagnetic calorimeter?”

“Electrons (?)”

“Yes! What else?”

“Anti-electrons?”

“Which we more commonly call positrons, yes, what else?”

“... ”

“What are the carriers of the electromagnetic force? Have you ever heard about the light particles?”

“Photons!”

“Yes, that’s it: the electromagnetic calorimeter allows to measure the energies of photons and electrons. It is made of nearly eighty thousands crystals of lead tungstate (and oxygen: PbWO_4). When electrons & photons enter one of these crystals, they emit light - that is, photons-, which we collect at the other end with a photodetector. Here is one of these crystals.”

Alice takes the glass-looking crystal, about twenty (23) centimeters long, two by two (2.2 by 2.2) centimeters:

“Wow, that’s heavy! And quite transparent as well for something made of lead and tungstene.”

“Heavy yes, but more important than heavy?”

“Transparent!”

“True, this matters since we measure the energy with the amount of light emitted in the crystals, but back to heavy. It’s heavy, and not so big, so it is...”

“...it is dense...”

“Exactly, we need something which is dense because to measure the energy, we absorb all of it (in principle), so we need some *dense* material with which the incoming particles have a high probability of interaction. By the way, how would you make the difference between electrons and photons?”

“... ”

Appendix A (Continued)

“Is the photon charged?”

“Hmmm...no...”

“No, it’s not, so it does not have a track, only a deposit in the electromagnetic calorimeter. And what about the electron, is it charged?”

“Yes, it is, so it shows a track!”

“...plus a deposit in the electromagnetic calorimeter: having different sub-detectors allows us to measure different quantities, but also helps with particle identification.”

“After the electromagnetic calorimeter, we have another calorimeter: the hadronic calorimeter. Which particles does it look at?”

“...”

“hadrons (?)”

“Yes! and what are hadrons?”

“hadrons...heavy... quarks?” Alice tries.

“Yes! hadrons are particles made of quarks, like protons, neutrons, kaons, lambdas, pions....there are many many different kinds of hadrons (too many of them, and that’s why physicists from the sixties designed the more elementary quark model). And why do we talk about hadrons and not quarks?”

“because quarks cannot be free (?)”

“That’s it, quarks are always confined within pairs or triplets, so what we see in our detector are hadrons, not individual quarks. The hadronic calorimeter is what we call a sampling calorimeter: it consists in brass plates, in which we absorb some energy of the particles, interspersed with scintillator, to measure the energy.”

“After the calorimeters, we have...”, another part of the model lightens, “the S of CMS. What does it stand for?”

Appendix A (Continued)

A student, looking at his badge:

“Compact ... Muon ...Solenoid!”

“Indeed - what is a solenoid?”

“Hmmm...a magnet?”

“True, a solenoid is a like a coil, but making a tube (thirteen meters long, seven meter in diameter here), which we feed with a current to produce a magnetic field. What is the use of a magnetic field here?”

“To accelerate particles (?)”

“Ah no... Of two things one: first, you do not accelerate particles with magnets. To accelerate charged particles, you need an electric field. A magnet deflects them...recall, I told you that to measure particles' momenta, we need to measure their track curvature. The magnetic field acts on charged particles and bends their trajectory, otherwise, they would go straight and we would not be able to measure particles' momenta. Does it make sense?”

(They're nice and politely agree.)

“Our magnet is a bit special though. Since the particles produced in the collisions are quite energetic, we need a very strong magnetic field: 3.8 Tesla, about a hundred thousand times higher than the Earth magnetic field. Now, when we are outside of the detector, we don't want to walk around in such a high magnetic field, that's why there is what we call the *iron yoke*: all these red layers are actually made of steel, creating a counter field to isolate the environment from the CMS high magnetic field. The iron yoke makes up for most of the weight of the CMS detector, twelve thousands tons out of its fourteen thousands.

In between, you see the M of CMS: muon chambers. Any idea about what are muons?”

“Heavy electrons! But then, why don't you see them with the electromagnetic calorimeter, like the electrons?”

“Indeed, muons behave similarly to electrons in general, but since they are two hundreds times heavier,

Appendix A (Continued)

they emit much less bremsstrahlung radiation - the way deflected charged particles lose energy - and they can penetrate further into matter.

One question at this point: could we measure first the energy, and then the momentum?"

Some hesitations...

"Hmm...no...."

"Why? What's left after calorimeters? Since we absorb all of the energy of the particles?"

"Well, nothing then."

"Yes, in principle, there is not much left after the calorimeters, so we must measure first the momentum then the energy. So, if you go & see ATLAS some day, you will observe the same structure: a tracking system for the innermost layers, followed by a calorimetric system, some magnets (possibly in between), and, last, muon chambers. The technologies are different, but the structure remains quite similar. "

"Why are there two detectors then?"

"Well, we have one LHC around the globe, if we want to trust the results that we see, we need two experiments to cross-check each other. Had ATLAS observed a Higgs boson and not CMS, or if one of these experiment had seen it at, say, $130 \pm 1 \text{ GeV}$ and the other at $120 \pm 1 \text{ GeV}$, we would have been quite embarrassed. The physics goals are the same, the structure is similar but the technologies used are different - mainly driven by different choices in terms of magnetic field. I told you earlier that the S of CMS is its solenoid, providing an incredibly strong - nearly 4Tesla- magnetic field. In fact, it is also the C of CMS: Compact. Our detector, with its 30 meters long, 15 meters in diameter, is much more compact than ATLAS - 40 meters long, 25 meters in diameter. The reason behind is that they have a magnetic field which is roughly twice lower, so that the bending power is roughly twice less, and when they want to measure the curvature they need a longer range."

"Why didn't they use a stronger magnetic field then?"

Appendix A (Continued)

“Well, at the time these detectors were design, it was not obvious that it would be feasible to produce 13 meters long, 7 meters in diameter superconducting solenoid, nor was it deemed realistic to fit detectors, and their electronics, within such a high magnetic field. The LHC, its experiments, are often their own prototypes...At some point, people designing them trusted that between the design and the actual construction, we’d have the necessary knowledge to go beyond ‘realistic’.

Talking about the magnet, here we have a piece of it . You have to imagine that the circle is complete, and the length is not 70 centimeters but about 13 meters. To generate such a field, we use a current close to 18000 Amperes. How much do you have at home?”

“10”, “20”, “32”...

“Yes, not much more than a few dozens Amperes at most: 18000 Amperes is a lot, way too much for copper for instance. To sustain such a high current, we need superconducting technology: our solenoid is made of Niobium-Titanium wires which, once cooled down to 4K, have no longer any electrical resistance. Questions?”

“How do you cool down the magnet?”

“The first step is done with liquid nitrogen, then we use liquid Helium. By the way, do you know how Helium has been discovered?”

“...”

“By observing light coming from stars: it is rather rare on Earth. In fact, it is mostly produced by radioactive materials liberating alpha particles, that is Helium nuclei. Another question? No? A quick recap then.” She moves towards a picture (Figure 82). “There aren’t so many particles that we can directly observe in our detector: can you recall them?”

“Electrons...”, “Positrons”, “muons”, “photons”...

“Yes, just a couple more missing?”

“Hmm...those made of quarks...”

Appendix A (Continued)

“Yes, hadrons, charged and neutral. So, what can you see here in red? There is a track, so it has to be charged, plus a deposit in the ECAL...”

“Electron or photon?”

“Close - which one is charged?”

“The electron.”

“Yes! Now, this green track, plus a deposit in the HCAL?”

“It’s charged”, “it’s a hadron.”

“Indeed, a (light) charged hadron. And this one, light blue going all the way until the muon chambers?”

“Muon”

“Yes, now, last but one: dashed green plus a deposit in the HCAL?”

“Hadron again”

“True, but since there is no track, it’s a (light) neutral one. And last, just a deposit in the ECAL?”

“A photon?”

“Yes, and that’s it, all the particles that we observe “directly” in the detector.”

“Ok, let’s move to the pit now. We’re going to pass by the control room, but I won’t stop there, I’ll do that on the way out if we have time. Lots of black screens at the moment anyway.” They follow their guide into a rather dim room, with screens all over the place.

“Ah, the MAD key is not back yet...hmmm...I’ll describe you the control room while we wait for it then”, says the guide.

“The control room is where we operate the detector from. We’re not switching it on and off directly playing with its cables in the pit! There are two parts in this control room.

Behind you is the subdetectors area. When I talk about subdetectors I mean pixel, tracker, ECAL, HCAL, muon systems: the various layers of CMS. Subdetectors experts don’t have to be here all the

Appendix A (Continued)

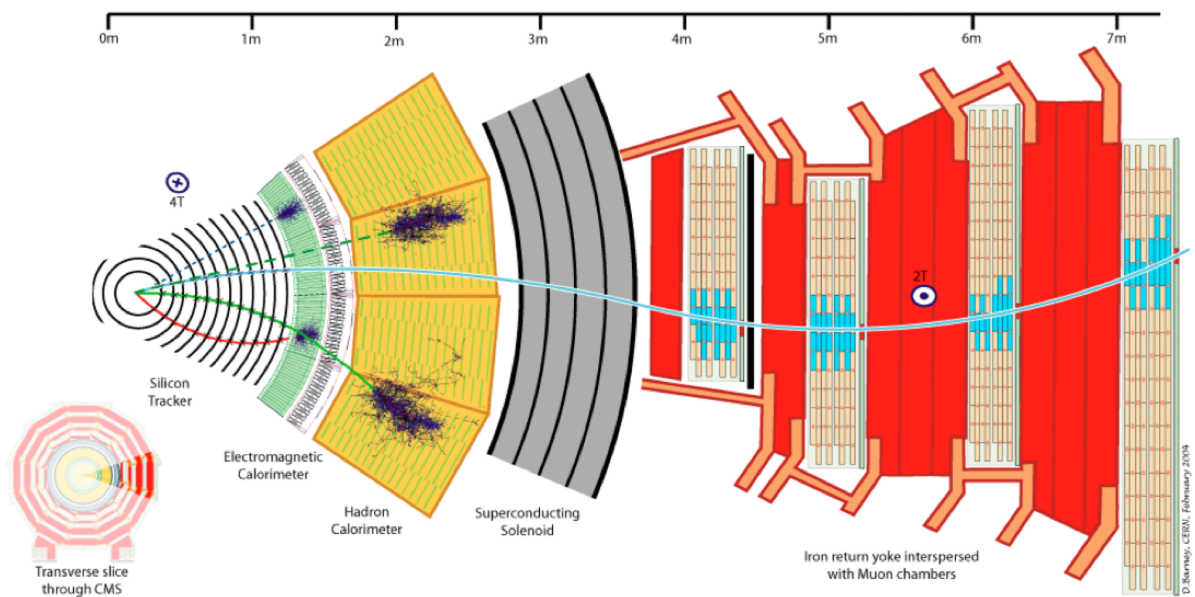


Figure 82. **Schematic cut view of CMS** - The picture shows where each kind of particles interacts: electrons have a track plus a deposit in the ECAL, photons only deposit energy in the ECAL, muons have a track plus deposits in the muon chambers, light hadrons leave energy deposits in the HCAL and have a track if they are charged.

Appendix A (Continued)

time: they come if they want to change or have to fix something, otherwise, they have a phone with them and we call them if needed.

Around us, this is the area of the control room where we usually have people 24/7 when we take data: one *shift* crew in the morning, another one in the afternoon, and again another one overnight. At the moment, since the LHC has entered its first long shutdown, you don't see so many people though.

Here usually sits the shift leader: his/her role consists in supervising the shift crew, interacting with the CCC (the LHC control room) to say "ok, we're ready to take collisions", or "what are your plans, could we have access to the cavern?".

Then, in the corner, we have the DQM, or Data Quality Monitoring, shifter, who checks that the data we are collecting looks alright. For instance, if we are collecting electrons, over thousands of collisions we should see them coming from everywhere in the detector; if we see electrons only from one side, there is an issue with the machine.

In the back, we have twelve screens for the trigger shifter. The trigger system is a key part of LHC's experiments: the LHC gives us collisions every fifty nanoseconds, which means that we get about seven millions bunch crossings per second (there are gaps in the beam due to injection steps, plus a $3\mu\text{s}$ beam abort gap). Of two things one: first, this is way more data than what we can record, second, not all of the collisions are interesting. As I just hinted at, the beam in the LHC is not a continuous beam, rather, it is made of bunches of protons: about one thousand bunches, each containing a hundred billions protons. When two bunches collide, most of them do not see anything: on average we get only fifteen interactions, but soft ones, protons glancing at each other. What we are interested in are highly energetic, head-on collisions, liberating a lot of energy to create new particles. The goal of the trigger system is thus to analyze each crossing very quickly to keep what looks interesting and throw away most of the collisions. This operation is carried out in two steps: first, what we call the Level 1 system selects about a 100kHz, within $3.2\mu\text{s}$ per event. This is done in the service cavern, we don't have the time to move the data up

Appendix A (Continued)

to the surface to do that (it would already costs about half a microsecond to move the information up and down at the speed of light). The second step, which we call the High Level Trigger is performed just one floor above our heads, and allows to further refine the selection, down to about 350Hz, within a hundred and sixty milliseconds on average per event.”

“But how can you be sure that you don’t miss anything?”

“Well...we need to study carefully beforehand how efficient our selection criteria are. It might not be 100%, it has to be a compromise between the physics (priorities) & what we can afford. By the way, when I say ‘selection criteria’ these are typically requiring a set of objects above a given threshold in energy, within some area of the detector. For instance, people who want to study the Higgs boson may ask for a couple of photons above a given amount of energy, or a couple of leptons, people studying B-physics typically ask for a couple of muons with some specific kinematics, etc...All these selection criteria are running in parallel: the data we collect is meant to feed hundreds of analyses, thanks to about five hundred selection criteria.”

“Behind me, usually sits the DAQ, for Data AcQuisition, shifter. (S)he is responsible for starting the runs, stopping them, ensuring that we are writing the data to disk.... Nobody reacts. The guide realizes that she’s expanding too much about the details. Of course, she’s been doing shifts here, she sees the room alive!

Last, we still have a couple of shifters these days at the DCS desks: they take care of all the infrastructure, powering, cooling, air conditionning,...we still go downstairs and we still want to be able to breath so we still need them.

“And what about the Champagne bottles?”

“Ah, that’s whenever the LHC achieves a new step: first collisions, first collisions at higher energy, first heavy ions collisions because sometimes they accelerate lead ions. One of the last that was open was for the 25ns run: in 2011 and 2012, we had collisions every fifty nanoseconds, but, in the design of the

Appendix A (Continued)

LHC, it should be twenty five nanoseconds. It's trickier because then two consecutive bunches starts to talk to each other, so, they did some tests at the end of the 2013 run, in preparation for the restart in 2015. But I've never been here when they opened a bottle", she laughs.

"Oh, the MAD key is coming back! Thanks XXX! Let's go down now!"

They walk by a set of pictures.

"This is only a very small subset of the collaboration, we are over 4000 people working for CMS", "here you can get a glimpse of the service shaft"...

They stop in front of a large blue door, with a sign picturing a barred heart.

"I am going to open you the MAD, please walk through and then towards your left to pick up helmets."

"What is this sign?"

"When the magnet is on, I usually ask whether anybody is pregnant or wear a pacemaker or any kind of electrical implant. These persons are not allowed to go underground, it's a measure of safety. Since the magnet is off these days, it is safe for everybody."

"But, you said that there is some iron to isolate us from the magnetic field?"

"Yes... mostly. It happens sometimes that people go downstairs with a laptop and come back because it broke...not always, but sometimes it does, because of small "leaks"...I'll walk through the PAD and meet you on the other side."

After getting helmets, they step in the elevator. Meanwhile, the guide continues:

"We're going to take the elevator down to about eighty meters. One point about safety: normally, what should you NOT do in case of fire in a building?"

"Run" , "Scream" , "Panic" , "Take the elevator"

"Yes! Why should you not take the elevator?"

"Because it can break ... because the power could be cut..."

"Hmm, no, that's not the main reason: the biggest issue is that the shaft acts as a chimney, all the

Appendix A (Continued)

fumes go there and you get suffocating even before there is any mechanical or electrical issue. Yet, here, the elevator is our entrance and exit way, no other choice. So, what has been done is that the area of the elevator, and the underground platforms next to it, are over-pressurized: the pressure is higher than elsewhere in the pit, so that in case of fire, the fumes would be pushed away. To sum up: in case of a fire, this is the safe place where we should move to. I should say, in case of any issue down there: stay with me, if half of the group rushes towards the elevator and the other half lags behind...I'm in trouble! Ok, here we are. ”

“You can take a look at the shaft from below, we were all the way up there before!”

They walk into a rather noisy room, filled with blue cupboards, the front size of which showing some sort of computing devices, the back side revealing a lot of cabling (or the other way round...).

“Here we are in one of our two counting rooms - basically a lot of computers. For instance, this is where the first level of selection (trigger) takes place. It is only a very small part of all the computers that we use though: we also have thousands of computers above the control room for the High Level Trigger, and then, since we have too much data to store all of it in a single place, we distribute it all over the world. As a result, when we want to analyze it, we may be writing our analysis code here at CERN, or somewhere in France, in the US, in India...anywhere, and then we send it to run where the data is stored and we get back the output. This is what we call *the Grid*.”

Alice looks perplex, “what do you mean by code?”

The guide looks embarrassed...

“Hmmm...how to explain it...wow, I'm spending most of my day writing code and I'm having troubles explaining what code is... First example: to analyze the data, we have some *code* to reconstruct the particles from the raw electrical signals that we collect from each of the sub-detectors. The *code* corresponds to a set of algorithms that can link relevant pieces together, discard the others, etc...but it is not only about algorithms, it's algorithms written in a language that can be understood by a computer...Yes,

Appendix A (Continued)

maybe that's a way to phrase it: a piece of *code* is a way to tell a computing device which operations to perform (possibly with what). Another example: any application that you use on a computer is *coded* in some way, there is some sort of text saying 'if Alice clicks here, a window should open, if she types 'Hello', please answer back (kindly),...if she asks for coffee...well, you may answer 'sorry, I don't know how to make coffee (yet)''.

"What are these orange stripes on the cupboard?"

"Oh, this means that if somebody pushes the AUG (emergency current switch off button), the power will not be switched off in these racks. For these racks, one would have to use the specific rack button.

Any other question?"

"Are we going to see the real thing?"

"Yes, off we go :)."

A flight of stairs, a picture of the LHC tunnel...

"Oh, a quick stop here. The picture that you see here is a real size one, so you see, the tunnel is big enough so that somebody may walk next to the accelerator, even bike actually." Some people smile... "We had an evacuation exercise once, with about two hundred people underground. Most of us exited using the elevator, but a few people were close enough to the tunnel so that it was "easier" for them to bike to the next shafts, P4 and P6, and exit through these. However, we are not going to see the tunnel. Here is a piece of explanation: for its superconducting equipment, the LHC uses materials such as liquid nitrogen and liquid helium. If, for some reason, there is a leak, these are going to vaporise. As a result, the percentage of oxygen in the air decreases. Do you know what is the normal oxygen percentage in the air that we breath?"

"21%"

"Yes, now, when it decreases, when does it start to get dangerous? I mean below which percentage?"

"10?", "15?"

Appendix A (Continued)

“Higher than that: below 18%, we are already not so well, and we may not even realize it, because it disturbs the reasoning. And below 16% it’s death in eight minutes. This is why we have this sign over there: in case of *oxygen deficiency hazard*, the light above would start flashing, and we would hear an alarm - better evacuate! On top of this, people who are working next to the detector and in the tunnel must have self-rescue masks with them, and they should have been trained to know how to use it. It is not complicated, but we cannot train visitors for that...Questions?”

“Well, you are saying that when the amount of oxygen in the air decreases from 21% down to 18%, it is already bad for us, but, when people go high up in altitude, say, at 4000m, we say that they have 40% less oxygen than usual, so, how come they can survive there? ”

“It took me a while to get this indeed: my understanding is that, as you go higher, the pressure decreases, hence, the air gets more diluted, so that in a single breath, you get 40% less oxygen, say, in terms of the numbers of molecules. The ratio between oxygen and other elements remains similar though.”

They turn towards some sort of complicated door.

“This is the door to access the experimental cavern. Normally, when someone wants to go there, (s)he badges, this drawer opens to give access to a set of keys, (s)he picks up one to open the door, then (s)he looks into the iris scan and if (s)he has the proper access rights, the second door opens and (s)he may proceed to the cavern. As long as there is one key missing from the set of keys, nobody can switch on the LHC: we don’t want to have beam in the machine when there is someone next to the machine.”

“Why is that? Is it dangerous?”

“Well, there are some radiations. Neutrons mostly, quickly absorbed by the concrete around, but still, you don’t want to stay in that flux.”

“For visits though, we’re not going to take one key per person, you are not in the database. I have a special key. In any case, the control room is watching us - see these cameras on each side!”

Appendix A (Continued)

Another door, and they're in. For the first time since the beginning of the visit it seems, the guide remains quiet for a while. They're standing on a balcony next to...next to...? It doesn't look like anything they've seen before, a lot of cables, something big, bright, colorful...and people working down there...

"As you can see, the detector is open at the moment: we have three slices to the right, one in front of us, the remainder on our left. Normally, when we take data, all this is closed, which means that these four slices move to the left, towards the central slice which holds the magnet and doesn't move. The slices are moved thanks to these orange feet -air cushions in fact- that you see down there. They allow us to lift the structure, while we draw some cables in between the slices to pull them. There are rails in the ground to guide them as well. When it happens, it is very slow. One day, I came here at 9a.m., and only two slices had been moved apart; when I came again at 11a.m., they were opening the third one, and I stared for a few seconds before being sure it was moving."

"Can you identify some parts of the detector I described earlier? Which detectors constitute this slice on the left?"

"Hmm, the red is...iron...."

"Yes, this is the iron yoke, and in between ?"

"Hmmm....muon things...."

"Indeed, the grey layers are muon chambers. It's a barrel slice, so these are DT, drift tubes. Then, any idea about what constitutes the slice on our right? It's harder, because it's an endcap slice, and I haven't described these..."

"Again muon chambers?"

"Yes, everything that reads 'ME&RE' are muon chambers. Endcap muon chambers are CSCs, for cathode strip chambers. We use three kinds of muon detectors in CMS, but they all rely on the same basic principle of an incoming muon, ionizing some gas, that is, freeing some electrons that are thereafter collected to generate an electrical current."

Appendix A (Continued)

“You may notice some sort of black and white stickers on these, similar to those that you can see also on the green structure. What we do is that we flash some laser light, which is reflected by the photosensitive whitish part. This enables us to determine the position of the detectors: think about it, with a 30 meters long, 15 meters in diameter monster, we can reach a precision of a few dozens of microns. It’s all the trickier than when we switch on the magnetic field, the detector shrinks by about one centimeter! It’s a lot compared to a few dozens of microns...In practice, we inject the information ‘position of the detector’, measured thanks to the lasers, in the data that we take, rather than aligning super-precisely the apparatus.

Here, where it’s written HE, this is the endcap part of the hadronic calorimeter, in front of which we have the endcap part of the electromagnetic calorimeter (and the preshower). When the detector is closed, these come fit within the (currently) ‘lonely’ slice, within the solenoid.”

“If we walk further on the balcony, what can you see? Can you see the solenoid?”

Alice points out towards the slice, sketching a circle in the air: “this”.

“Yes, here it is! Seven meters in diameter - could you guess why?”, she asks with a smile. “Some pieces of the magnet were built in different institutes, then brought to Prévessin, where it was assembled. Afterwards though, it had to be brought here...by truck, so it couldn’t be wider than the roads around here! A seven meter diameter, thirteen meters long superconducting solenoid is already quite big anyway. Within the solenoid, from outside to inside, are the barrel parts of the calorimeters, the tracker, and usually the pixel but at the moment it has been removed and brought up upstairs for fixing and developing. Shutdown for the LHC is not rest time for us, we’re upgrading as well!”

Indeed, Alice sees scaffolding and workers getting inside the detector - she’d be curious to go there too!

“There are so many cables everywhere! What happens if they get messed up? Who is working on this?”

“Technicians, engineers - to be honest, I’m not able to detail the different qualifications involved here.

Appendix A (Continued)

Nevertheless, I'm always impressed by the work done with this machine: there is no mistake possible with the cabling for instance, because once the detector is closed, it's closed and if there is something that doesn't work we cannot go and change it. It happened that some of the ECAL crystals (the electronics associated to them) were not responsive from the very start of the run. We had to look at the energy of the surrounding crystals to extract some information about the energy deposited in the faulty ones. Talking about cables, you see many of them...but we make our best to minimize the amount of cables: it is not active material, not what we use to see the particles - it's like our neurons if you wish, while the crystals, the silicon, the brass, the gas, etc... are our eyes. Another question for you: why do we put our detectors inside the solenoid and not outside?"

"Hmm...not to disturb the measurement (?)"

"That's it, we don't want particles to lose some of their energy before we measure it."

"But, where are the protons?"

"They are circulating in a beam pipe, which you cannot see at the moment: it has been removed. If you go and look at the magnet in Microcosm¹, you will observe that there are two pipes, one for each beam since they are circulating in opposite direction. When they arrive at the experiments, they are brought together in the same pipe, and focalized so as to collide at the very center of the detector. If the LHC gives us collisions over there in the endcap part, we're not happy, CMS is not made for that: we would not see the collisions with adequate glasses, and it would damage the detector². That's also why there are some devices which are called 'Beam Condition Monitor', placed very close to the beam and highly radiation hard (for ATLAS, the BCM is made of diamond sensors) to check whether the beam is

¹A permanent exhibition at CERN.

²There is some inertia in the beam: it cannot be only one bunch hitting the detector: if one does so, the whole beam is misaligned and is going to hit the detector. The whole beam energy is equivalent to a high speed train...

Appendix A (Continued)

deriving from its axis. If this occurs, the BCM sends an abort beam signal.”

“What is the beam pipe made of?”

“Well, you need a beam pipe because you want to maintain a high vacuum within, so that you study proton-proton collisions, not ‘proton-dust’ - so you’d better have something strong enough to sustain the vacuum. Yet, you also want that particles produced during the collisions go through the beam pipe without interacting to avoid energy loss: as a result, you need some light material. In practice, different materials have been used at different places, beryllium close to the interaction point, stainless steel further away, all this coated with other elements helping with the cleaning...”

“What is the temperature of the experiment? You said that the solenoid is very cold, so as to be superconducting, so everything here is very cold?”

“No, the different subparts of the detector may have different temperature. As you recall, the solenoid is cooled down to about 4K, that is, about -269 Celsius degrees. To reach this temperature, we use liquid Helium. Now, think about it, we want to cool down the solenoid, but we may not need such a low temperature for the detector around. In fact, we don’t want to cool down other subparts that much: not only because Helium is expensive and we should restrain its use to what is really needed, but, also, because when cooling down a system, you may have troubles like...condensation. Thus, the Helium circulating around the solenoid must be used only to adjust the solenoid temperature: the magnet is surrounded by a cylindrical tank made with an insulating material.

As for the subdetectors:

- the electromagnetic calorimeter operates at a temperature precisely regulated at 18 Celsius degrees.

The Avalanche Photo Diodes that are used to convert the light into an electrical signal are very sensitive to the temperature, such that if it changes, the amount of noise varies, and we have

Appendix A (Continued)

troubles interpreting the signal we collect. That's why the temperature should remain quite constant;

- the pixel & tracker are made with silicon, and when silicon is irradiated, some atoms may be kicked off their crystalline sites, and these defects tend to cluster. In fact, we try to operate Pixel and the Tracker at -10 Celsius degrees, so that these defects are less mobile and stay where they've been ejected to, rather than clustering.
- muon chambers electronics are cooled down with water.

"How long does an experiment last? How long does it take to do a measurement?"

"It's a bit special: when you study physics at school, when you have lab session, you come, setup your experiment, perform a measurement, analyze the results, and then move to something else. Typically, it lasts a few hours, days, or maybe weeks. With the LHC experiments, it's quite different: since we cannot decide of what happens in a collision, we carry on hundreds of experiments in parallel, via the trigger, that is, the selection criteria that we apply. For instance, people looking for the Higgs boson may be looking for collisions with at least two energetic leptons, people studying B physics may ask for two muons with peculiar kinematic properties, people looking for supersymmetry usually require a lot of hadrons and some *missing energy*. We collect all these types of data, and each group takes the part it is interested in analyzing."

"So, it's running all the time?"

"Let's say it's running on a 24/7 basis. In practice, there are cycles: first, the LHC operators inject a probe beam. If it goes well, they inject both full beams. Then, they ramp their energies, squeeze the beams, focus them and look for *the collisions point*. After a while, when they feel they have things well under control, they declare *stable beam* and we start to take pictures. The setup cycle takes about one-two hours (in streaming mode -for a restart, it usually takes longer)."

Appendix A (Continued)

“Why?”

“Well, after you change some pieces in a machine, you may need a bit of time to tame it again. For instance, once, they changed a *magnet kicker inject*: a magnet responsible for transferring the beam from the SPS to the LHC. When they restarted, anytime they were trying to inject, they had troubles with the vacuum, because the new piece of equipment was not yet used to its environment, and when the vacuum was getting higher and higher, a lot of dust would be released by the material.”

“Once we have collisions in the experiments, we take pictures for approximately up to ten-twelve hours, after what the beams are dumped: the luminosity, that is, the rate of collisions we can expect, gets too low to be interesting.”

“The LHC runs as continuously as possible, but, from time to time, there are few days shutdowns, for maintenance. Additionally, during the wintertime, there is usually a break because the electricity is more expensive, we are competing with Geneva: the LHC consumes as much as a 180 000 inhabitants city. This does not mean rest time, it’s always of some use not to have beam in the machine to be able to partially access it. We haven’t been running for a while though. In February 2013, the LHC entered what we call *LS1*, for Long Shutdown number 1. Any idea why? Have you heard about what happened when the LHC first started back in 2008?”

“It broke.”

“Hmm, why? What was the issue?”

“ ...”

“The LHC is using superconducting technology, for the magnets, the RF cavities, and, also, the soldering (to put it short) in between these elements. Unfortunately, one of these soldering was a bit too resistive, and there were several kiloAmps in there, which it couldn’t stand. Electric arc, lots of energy released and destroying some magnets...they had to stop for another year, to check each junction (after cleaning the area), to see at which energy we could run: the higher the proton energy, the stronger the magnetic

Appendix A (Continued)

field you need to keep the particles along the circle, hence the higher the current you need to inject within the magnets. For the first run of the LHC, we got collisions at 7TeV, then 8TeV. This was already 7-8 times higher than the previous accelerator (Tevatron, at Fermilab)! Yet, the nominal energy of the LHC is 14TeV, and we hope to get closer to this value for the second run, which requires to check again all the system, possibly changing some pieces, and training the magnets. Any other question? No? Then, let's go back up to the surface."

They walk again through the sophisticated door, and the guide finds something else to stop by: "Why is there a radiation sign here? In the room behind this door there is a Strontium source used for calibration. Talking about radiations, what I'm wearing here is a passive dosimeter: it is accumulating the radiations received, and, once per month, I check how much it got. Honestly, it is mostly measuring the radioactivity in the Pays de Gex: during the Open Days, each guide also had an active dosimeter, which beeps any time some particle goes through. Mine did not beep even once during the five hours I spent here."

"I point out safety measures, however, accidents are extremely rare. Yet, we are on a building site, eighty meters below ground, with potentially dangerous elements: if something happens, we have to be prepared to react adequately because it could get much more complicated than at ground level."

They walk again through a counting room, then towards the elevator. Some pictures on the wall.

"That's how it looked like fifteen years ago", says the guide, pointing to a picture with a few people walking on some rocks in the middle of fields. While they wait for the elevator, she continues: "The detectors are state-of-the-art technology, but it was already a serious challenge to build the shafts. When they started to dig the ground, they found an underground river. Can you guess what they did with it?" When asking questions, there's often something malicious in her tone. Probably all this is fun to her!

"Divert it?"

"Sort of: they put some pipes, filled them with liquid nitrogen, which froze the ground, hence they could

Appendix A (Continued)

dig the shaft, build the walls, and now the river is flowing around.”

Meanwhile, the elevator has arrived. A minute or so later, they're back up. “You can get rid of the helmets now, please put them back where you took them. I'm going to open you the door and I'll meet you on the other side.” Walking back again through the control room, back to the entrance....Any last question?”

“What is the Higgs bosons?”

“Ha! Well. Let's give it a try. In particle physics - probably I should have explained that earlier...well, you didn't ask...In particle physics, we try to describe the world around us in terms of its tiniest pieces. If you want to build a house, you will look at things around saying: here is wood, metal, concrete...I can use these. If you want to make chemicals like drugs, cosmetics, you may start playing at the molecular or atomic level. In nuclear physics, they play with the atoms' nuclei. We look at an even smaller scale, within protons & neutrons, when we reach the quark level. We also have electrons. So, quarks, electrons (and their heavier cousins muons & taus) constitute our set of elementary (unbreakable) matter particles. But this is not enough, we also want to explain how they interact, how they talk to each other if you wish. Us, humans, we interact mostly by exchanging words. Particles interact by exchanging...particles, the force carriers: photons, for the electromagnetic force, gluons, for the strong interaction gluing quarks together, W & Z bosons for the weak interaction (radioactivity). The graviton...is another story. Now, explaining in physics is not only describing with words what is happening. At some point, we want to fit our particles into equations in order to predict what they will do, how they behave. When casting particles into equations, there are some rules, like our grammar for words. And following these, initially, particles in our equations do not have mass. That is very embarrassing, because in physics experiments, they do have a mass. A workaround consists in introducing a new particle in the equations which is going to interact with the known ones to give them a mass. This is the Higgs boson. Does it make a bit of sense?”

Appendix A (Continued)

“Yes yes...but now that you’ve found the Higgs boson, what are you doing?”

“First, we want to measure better the properties of this new particle, to see to which extent it corresponds to what we call a standard model Higgs boson. Hopefully, at some point, it won’t match, and we’ll reach some new area of particle physics. There are also other things which we cannot really explain: dark matter, neutrino masses, the Higgs mass itself...” A honk sound...the bus driver is waiting for the visitors to get on the bus and go back to Meyrin, where they should be guided through a few other CERN installations. Time to leave...

“Bye!”

CITED LITERATURE

1. CDF Collaboration: Observation of Top Quark Production in Antiproton-Proton Collisions with the Collider Detector at Fermilab. Physical Review Letters, 74:2626–2631, 1995.
2. Halzen, F.; Martin, A.: Quarks & Leptons. Wiley, 1984.
3. Noether, E.; Tavel, M.A. (translation): Invariant variation problem. Gott.Nachr.1918:235-257; Transp.Theory Statist.Phys.1:186-207,1971, arXiv:physics/0503066v1, 1918, 1971.
4. Yukawa, H.: On the interaction of elementary particles. Proc.Phys.Math.Soc.Jap. 17, 17:48–57, 1935.
5. ATLAS and CMS collaborations: Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. arXiv:1503.07589v1, 2015.
6. Englert, F.; Brout, R.: Broken Symmetry and the Mass of Gauge Vector Mesons. Physical Review Letters, 1964.
7. Higgs, P.: Broken Symmetries and the Masses of Gauge Bosons. Physical Review Letters, 1964.
8. Guralnik, G. S., Hagen, C. R., and Kibble, T. W. B.: Global conservation laws and massless particles. Physical Review Letters, 13:585–587, 1964.
9. Gell-Mann, M.: A Schematic Model of Baryons and Mesons. Physics Letters, 8:214–215, 1964.
10. Takaaki Kajitaa for the Super-Kamiokande and the Kamiokande collaborations: Atmospheric neutrino results from Super-Kamiokande and Kamiokande. Evidence for ν_μ oscillations . arXiv:hep-ex/9810001v1, 1998.
11. Garrett, K. and Duda, G.: Dark matter: A primer. arXiv:1006.2483v2, 2011.
12. Martin, S. P.: A supersymmetry primer. arXiv:hep-ph/9709356v6, 2011.
13. Ellis, J.: Supersymmetry for Alp Hikers. arXiv:hep-ph/0203114v1, 2002.
14. David Alves et al.: Simplified Models for LHC New Physics Searches. arXiv:1105.2838v1, 2011.
15. CMS collaboration: Interpretation of searches for supersymmetry with simplified models. arXiv:1301.2175v2, 2013.

16. CMS collaboration: CMS Physics - Technical Design Report. CERN/LHCC 2006-001 - CMS TDR 8.1, 02-02-2006.
17. CMS collaboration: The CMS Experiment at the CERN LHC. JINST 3 S08004, 2008.
18. CMS Collaboration: The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s}=7\text{TeV}$ at the LHC. J. Instrum., (arXiv:1306.6905. CMS-MUO-11-001. CERN-PH-EP-2013-072):P11002, Jun 2013.
19. CMS collaboration: Muon Technical Design Report. CERN/LHC 97-32, CMS TDR 3, 1997.
20. On behalf of the CMS collaboration, Florian Beaudette: The Particle Flow Algorithm. Proceedings of CHEF 2013, 2013.
21. CMS collaboration: Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. CMS-PAS-PFT-09-001, 2009.
22. CMS collaboration: Single Muon efficiencies in 2012 Data. CMS DP - 2013/009, 2013.
23. CMS collaboration: Electron reconstruction and identification at $\sqrt{s} = 7 \text{ tev}$. CMS PAS EGM-10-004, 2010.
24. CMS collaboration: Search for direct top squark pair production in the single lepton final state at $\sqrt{s} = 8 \text{ TeV}$. arXiv:1308.1586v2, EPJC 73 (2013) 2677, 2013.
25. Salam, G. P.: Towards Jetography. arXiv:0906.1833v2, 2010.
26. Cacciari, M. and Salam, G.: The anti- k_t jet clustering algorithm. arXiv:0802.1189v2, 2008.
27. CMS collaboration: Identification of b-quark jets with the CMS experiment. JINST 8 P04013, arXiv:1211.4462v2, 2013.
28. Olive, K.A. *et al.* (Particle Data Group): The Review of Particle Physics. Chin. Phys. C, 38, 090001, 2014.
29. Weiser, C.: A Combined Secondary Vertex Based B-Tagging Algorithm in CMS. CMS NOTE 2006/014, 2006.
30. Buckley, A, *et al.*: General-purpose event generators for LHC physics. arXiv:1101.2599v1, 2011.
31. Seymour, M.H. and Marx M.: Monte Carlo Event Generators. arXiv:1304.6677v1, 2013.
32. Sjostrand, T.; Mrenna, S. and Skands, P.: PYTHIA 6.4 Physics and Manual. arXiv:hep-ph/0603175, doi:10.1088/1126-6708/2006/05/026, JHEP 0605:026,, 2006.

33. Corcella, G., I.G.Knowles, G.Marchesini, S.Moretti, K.Odagiri, P.Richardson, M.H.Seymour, B.R.Webber: HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes). JHEP 0101:010, arXiv:hep-ph/0011363v3, 2001.
34. R. K. Ellis, W. J. Stirling and B. R. Webber,: QCD and Collider Physics. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, 1996.
35. Salam, G. P.: Elements of QCD for hadron colliders. European School of High-energy Physics, Bautzen, Germany, 14 - 27 Jun 2009, pp.45-100, arXiv:1011.5131, 2009.
36. GEANT4 Collaboration (S. Agostinelli *et al.*): GEANT4: A Simulation Toolkit. 10.1016/S0168-9002(03)01368-8, Nucl.Instrum.Meth. A506, pages 250–303, 2003.
37. CMS Collaboration: Fast simulation of the cms detector. J.Phys.Conf.Ser. 219 (2010) 032053, doi:10.1088/1742-6596/219/3/032053, 2010.
38. Maltoni, F. and Stelzer, T.: MadEvent: Automatic event generation with MadGraph. arXiv:hep-ph/0208156, JHEP 02 (2003) 027, 2003.
39. CMS collaboration: CMS, The TriDAS Project Technical Design Report, Volume 1: The Trigger Systems. CERN/LHCC 2000 - 38. CMS TDR 6.1, 2000.
40. CMS collaboration: CMS, The TriDAS Project Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CERN/LHCC 02-26, CMS TDR 6, 2002.
41. ROOT Home Page, <https://root.cern.ch>.
42. CMS collaboration: Search for Supersymmetry in events with b-quark jets and missing transverse energy in pp collisions at 7TeV. arXiv:1208.4859v2, PRD 86 072010 (2012), 2013.
43. CMS collaboration: Search for supersymmetry in hadronic final states with missing transverse energy using the variables α_T and b-quark multiplicity in pp collisions at $\sqrt{s} = 8$ TeV. arXiv:1303.2985v2, EPJC 73 (2013) 2568, 2013.
44. CMS collaboration: Searches for supersymmetry using the M_{T2} variable in hadronic events produced in pp collisions at 8 TeV. arXiv:1502.04358v2, JHEP 05 (2015) 078, 2015.
45. CMS collaboration: Searches for 3rd generation squark production in fully hadronic final states in pp collisions at 8 TeV. arXiv:1503.08037, 2014.
46. CMS collaboration: Search for New Physics in Multijets and Missing Momentum Final State in pp collisions at 8 TeV. arXiv:1402.4770, JHEP 06 (2014) 055, 2013.

47. CMS collaboration: Search for supersymmetry in pp collisions at $\sqrt{s} = 7$ TeV in events with a single lepton, jets, and missing transverse momentum. [arXiv:1212.6428v2](#), EPJC (2013) 73:2402, 2013.
48. CMS collaboration: Search for supersymmetry using events with a single lepton, multiple jets, and b-tags. [arXiv:1311.4937](#), PLB 733 328 (2014), 2013.
49. CMS collaboration: Search for physics beyond the standard model in events with two leptons, jets, and missing transverse momentum in pp collisions at 8 TeV. [arXiv:1502.06031](#), JHEP 04 (2015) 124, 2014.
50. CMS collaboration: Search for new physics in events with same-sign dileptons, b-tagged jets and missing energy. [arXiv:1205.3933](#), JHEP08(2012)110, 2011.
51. CMS collaboration: Search for new physics in events with same-sign dileptons and jets in pp collisions at 8 TeV. [arXiv:1311.6736](#), JHEP 01 (2014) 163, 2013.
52. CMS collaboration: Search for anomalous production of events with three or more leptons in pp collisions at 8 TeV. [arXiv:1404.5801](#), PRD 90, 032006 (2014), 2013.
53. CMS collaboration: Search for supersymmetry with photons in pp collisions at $\sqrt{s} = 8$ TeV. [arXiv:1507.02898v1](#), PLB 733 328 (2014), 2015.
54. CMS collaboration: Search for Supersymmetry at the LHC in Events with Jets and Missing Transverse Energy. [arXiv:1109.2352.v1](#), 2011.
55. CMS collaboration: Inclusive search for supersymmetry using the razor variables in pp collisions at $\sqrt{s} = 7$ tev. [arXiv:1212.6961v3](#), 2012.
56. Cacciari, M. and Salam, G. P.: Pileup subtraction using jet areas. [arXiv:0707.1378v2](#), 2008.
57. Cacciari, M.; Salam, G. S. G.: The catchment area of jets. JHEP0804:005, [arXiv:0802.1188v2](#), 2008.
58. CMS collaboration: Measurement of the top-quark mass in tt events with dilepton final states in pp collisions at $s = 7$ TeV. [arXiv:1209.2393v2](#), Eur. Phys. J. C 72 (2012) 2202, 2012.
59. Rogan, C.: Kinematical variables towards new dynamics at the LHC. [arXiv:1006.2727v2](#), 2011.
60. Rogan, C.: Searches for New Symmetries in pp Collisions with the Razor Kinematic Variables at $\sqrt{7}=7$ TeV. [FERMILAB-THESIS-2013-29](#), 2013.
61. CMS collaboration: A search for the decays of a new heavy particle in multijet events with the razor variables at CMS in pp collisions at $\sqrt{s} = 7$ TeV. [CMS-PAS-SUS-12-009](#), 2012.

62. CMS collaboration: Search for supersymmetry using razor variables in events with b-jets in pp collisions at 8TeV. CMS PAS SUS-13-004, 2013.
63. Verkerke, W. and Kirkby, D. P. : The RooFit toolkit for data modeling. arXiv:physics/0306116.
64. CMS collaboration: Technical Design Report Volume II: Physics Performance. CERN/LHCC 2006-021 CMS TDR 8.2, 2006.
65. Hoecker, A.; Speckmayer, P.; Stelzer, J; Herhaag, J.; von Toerne, E.; Voss, H.: TMVA 4, Toolkit for Multivariate Data Analysis with ROOT, Users Guide. arXiv:physics/0703039v5, 2009.
66. Kaplan, D.E.; Rehermann, K.; Schwartz, M.D. and Tweedie, B.: Top-tagging: A Method for Identifying Boosted Hadronic Tops. arXiv:0806.0848v2, 2008.
67. Gallicchio, J. and Schwartz, M. D.: Quark and Gluon Tagging at the LHC. arXiv:1106.3076v2, 2011.
68. CMS Collaboration: Measurement of the W-boson helicity in top-quark decays from tt production in lepton+jets events in pp collisions at $\sqrt{s} = 7$ TeV. arXiv:1308.3879v2, 10.1007/JHEP10(2013)167, 2013.
69. Chwalek, T.: Measurement of the W-boson helicity fractions in top-quark decays at CDF. arXiv:0705.2966, 2007.
70. The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group: Procedure for the LHC Higgs boson search combination in Summer 2011. ATL-PHYS-PUB-2011-11, CMS NOTE-2011/005, 2011.
71. Cowan, G.; Cranmer, K.; Gross, E.; Vitells, O.: Asymptotic formulae for likelihood-based tests of new physics. arXiv:1007.1727v3, 2013.
72. CMS collaboration: SUSY Cross Section Working Group. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections>.
73. CMS collaboration: CMS Supersymmetry Physics Results. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>, 2015.
74. ATLAS collaboration: ATLAS Run 1 searches for direct pair production of third-generation squarks at the Large Hadron Collider. arXiv:1506.08616v1, 2015.
75. CMS collaboration: Exclusion limits on gluino and top-squark pair production in natural SUSY scenarios with inclusive razor and exclusive single-lepton searches at $\sqrt{s} = 8$ TeV. CMS PAS SUS-14-011, 2014.

76. CMS collaboration: Electron performance with 19.6 fb¹ of data collected at $\sqrt{s} = 8$ TeV with the CMS detector. CMS DP - 2013/003, 2013.
77. CMS collaboration: SWGuideMuonId. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId>.

VITA

NAME	Lucie Gauthier
EDUCATION	B.Sc., Fundamental Physics, Université Paris XI, Orsay, France, 2007 M.Sc., Particle Physics, Université Paris XI, Orsay, France, 2009 Magistère, Fundamental Physics, Université Paris XI, Orsay, France, 2009 Graduate Student, University of Illinois at Chicago, 2009-2015
RESEARCH	Work with the Trigger Menu Development group, 2010-2011
ACTIVITIES	Susy trigger coordinator, 2011-2012
<i>within</i>	DAQ (Data AcQuisition System) shifter, 2010-2011
<i>the CMS</i>	HLT (High Level Trigger) on-call expert, 2012
<i>Collaboration</i>	DCS (Detector Control System) shifter, 2014 ARC member for the Physics Chapter of the L1 Upgrade TDR, 2014 analysis: search for supersymmetry with the Razor group, 2013-2015
SCHOOLS	CMSDAS at Fermilab, January 2011
ATTENDED	European School of High Energy Physics, Hungary, June 2013 Hadron Collider Physics Summer School, Fermilab, August 2014
TEACHING	Department of Physics, University of Illinois at Chicago,
& OUTREACH	Physics without Calculus for Undergraduate Students CMS visits CMS masterclasses
HONORS	CMS Achievement award, December 2012