

SEARCH FOR THE HIGGS BOSON PRODUCED IN
ASSOCIATION WITH W^\pm BOSON, IN $WH \rightarrow \tau\nu b\bar{b}$
FINAL STATE, IN PROTON ANTI-PROTON
COLLISIONS AT 1.96 TeV CENTER OF MASS ENERGY

Elisabetta Pianori

A DISSERTATION

in

Physics and Astronomy

Presented to the Faculties of the University of Pennsylvania in Partial
Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2011

Nigel Lockyer, Adjunct Professor, Physics and Astronomy
Supervisor of Dissertation

A.T. Charlie Johnson, Professor, Physics and Astronomy
Graduate Group Chairperson

Dissertation Committee

Joseph I. Kroll, Professor, Physics and Astronomy

Nigel S. Lockyer, Adjunct Professor, Physics and Astronomy

Hugh H. Williams, Professor, Physics and Astronomy

Josh Klein, Assistant Professor, Physics and Astronomy

Mark Trodden, Professor, Physics and Astronomy

The most beautiful thing we can experience is the mysterious.

It is the source of all true art and all science.

*He to whom this emotion is a stranger, who can no longer
pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed*

- A. Einstein

Acknowledgements

I could have not reached the end of my PhD if not for the help I received from the many people that have been part of my American adventure. I want to take this opportunity to acknowledge at least some of them.

I want to express here all my gratitude to Prof. Nigel Lockyer, my advisor. His wide horizon and prospective helped me broaden my own horizon, his trust in me was a motivation to do my best, his questions provoked me to go deeper into the important issues of my research. I hope to one day have the same breadth of knowledge and prospective that he always brought in our conversations.

Most of what I now know of the day to day work of an experimentalist I learned it from Dr. Pavel Murat: I really appreciate his patience and all the time that he spent with me. His honest search and love for truth, and his view of the research work is something that will always stay with me.

For the very hard questions, the constant provocation to explain the reasons of anything I was doing, and the support throughout all these last 6 years, I need to thank Prof. Joe Kroll.

Although the time we had together was short, I deeply enjoyed working with Dr. Anadi Canepa. I learned the rope of our work from her, and her enthusiasm and dedication made working with her a highly productive and fun experience. I hope there will be other times for us to work together!

I also hope that in my future experience I will be able to work again with a group of talented, friendly people as the one that were part of the Penn group at CDF when I started: Prof. Chris Neu, Dr. Aart Heijboer, Dr. Justin Keung, Dr. Tatiana Rodriguez, Dr. Yanjun Tu and Dr. Joel Heinrich. I've learn a lot from each one of them. I will always remember the sense of being welcomed in a group, the relationships between the post-docs and the students, and the many lunches together. And, in general, I found the Particle Physics group at Penn a very alive and stimulating environment, and I consider myself privileged for the opportunity to interact and learn from Prof. Brig Williams, Prof. Elliot Lipeles, Prof. Evelyn Thomson, Prof. Mark Trodden and Prof. Josh Klein.

I must also acknowledge the leaders of the Higgs Discovery Group and the WH subgroup, Dr. Eric James, Prof. Craig Group, Dr. Homer Wolfe, Dr. Ben Kilmister and Dr. Wei-Ming Yao, as well as Dr. Tom Junk and Dr. Tom Schwarz for the useful suggestions and constructive discussions.

On a more personal note, I need to thank all the friends that shared my life, with all the struggles (and the many complains) as well as the successes, during my years in Chicago. First of all Paola, who become a sister for me, Leo, Dado, Giulia, Mayela, Nidia, Giulia and Viga, Benjo and Gaia, Cathy, my roommate for many years, Leslie, Katie, Cathy, Doug and Sarah, Ula, Paolo, Maria, Giovanni, Manu, and the rest of the CL community. Because of them, Chicago is my home now.

Also I'm grateful to Cristobal Cuenca Almenar and Barbara Alvarez, who welcomed me as soon as I arrived at Fermilab, and suggested me to become GSA at Fermilab. While doing that I met four incredible guys that become real friends throughout the years: Jorge Benitez, Emanuel Stauss, Tyler Dorland and Rob Forrest. Being a GSA representative was fun, but getting to know them and spending time together has been a great adventure! My experience in Chicago would have not

been the same without them! I also have to mention the friends with whom I carpooled and spent time exploring Chicago during my last year of grad school: Bjoern Penning, Marcelle Soares-Santos, and Hugh Lippincott (plus Rob Forrest!). During those moments of great difficulty, driving to Fermilab with them was the highlight of my day! I will always be grateful for all the laughs and interesting conversations we had about both work and life.

I cannot forget where everything started, in Philadelphia. I will always remember Luca Salvi, who, on the first day I landed in the United States, welcomed me into his home, and from then on, has always been a faithful friend, as few others in my life. I owe deep gratitude to my classmates, who patiently taught me English, explained American culture to me and shared with me the unending working hours of the first year of grad school, especially Joey Hide, Meeri Kim and Mike Herbst. I also need to thank the other friends that shared my time in Philly: Marco Grazzi, Federico Badala', Enea Simonato, Rebecca and Colin Chierico, Marcello D'Ascenzo, Lindsay Murray, Sarah Holtz, Michelle Bernardini, and Federica Sgariglia.

A special thanks goes to Dr. Giorgio Ambrosio: his friendship in these years has become more and more essential. Every conversation we had was a provocation, a reminder of a wide horizon in life able to embrace everything.

Last but not least, I have an immense gratitude to my parents Stefano and Giovanna, my brothers Davide and Raffaele and my grandmother Adriana, who have always believed in me, have always been proud of me, and have pushed me to aim high in life. Without their constant support, I could not have done it!

ABSTRACT

SEARCH FOR THE HIGGS BOSON PRODUCED IN ASSOCIATION WITH
 W^\pm BOSON, IN $WH \rightarrow \tau\nu b\bar{b}$ FINAL STATE, IN PROTON ANTI-PROTON
COLLISIONS AT 1.96 TeV CENTER OF MASS ENERGY

Elisabetta Pianori

Supervisor: Nigel Lockyer

The search for the associated production of the Standard Model Higgs boson and a W boson is motivated and discussed here. It is performed using data corresponding to an integrated luminosity of 5.7 fb^{-1} , collected by the CDF detector during $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$ at the Tevatron. $WH \rightarrow \tau\nu b\bar{b}$ candidate events are selected requiring two jets, a hadronically decaying τ lepton and large missing transverse energy. To increase the signal-to-background ratio, at least one of the jets must be consistent with originating from a bottom quark. A binned likelihood fit of the dijet invariant mass distribution is performed to test for a potential Higgs boson signal. In the absence of an observed excess, we set a 95% Confidence Level (C.L.) upper limit on the production rate times branching ratio for a potential Higgs boson as a function of its mass. For a test mass of $115 \text{ GeV}/c^2$, the observed (expected) 95% C.L. upper limit is 28.7 (46.6) times the Standard Model expectation.

Contents

Acknowledgements	iii
1 Theory: The Standard Model	1
1.1 The Standard Model	2
1.1.1 Gauge Symmetries	3
1.1.2 Quantum Chromodynamics	4
1.1.3 Electroweak Sector	5
1.1.4 Electroweak Symmetry Breaking	7
1.2 Naturalness	10
1.3 Higgs production mechanisms and decay modes	11
1.4 Theoretical Constraints on the Higgs Mass	13
1.4.1 Unitarity	13
1.4.2 Triviality and Vacuum Stability	14
1.5 Experimental Limits on Higgs Mass	15
1.5.1 Indirect Limits	15
1.5.2 Direct searches at LEP	17
1.6 Current status of Higgs searches at Hadron Colliders	18
1.6.1 Direct searches at the Tevatron	18
1.6.2 Direct searches at the LHC	22
2 Experimental Apparatus	25

2.1	The Tevatron Accelerator Complex	25
2.2	The CDF II Detector	27
2.3	The Trigger System	28
2.3.1	Calorimeter information at trigger level	29
2.3.2	TAU_MET trigger and the Level-2 upgrade	31
2.3.3	TAU_MET efficiency in signal events	33
3	Object Identification	34
3.1	Tau Lepton	34
3.2	Electrons	39
3.3	Muons	40
3.4	Jet Reconstruction	40
3.5	Neutrinos and Missing Transverse Energy	41
3.6	b -tagging algorithm	42
3.6.1	SecVtx b -tagging efficiency for heavy flavor jets	44
3.6.2	False positive b -tagging rate	45
4	Signal and Background Processes	47
4.1	Signal	48
4.2	Background processes	48
4.2.1	W+jets	49
4.2.2	Top production	52
4.2.3	Single top production	52
4.2.4	Z + jets	53
4.2.5	Diboson production	54
4.2.6	QCD Multijet events	54
5	Signal and Background estimation	57
5.1	Trigger Selection	57
5.1.1	TAU_MET trigger emulation	60

5.2	Offline Event Selection	60
5.3	Sample Composition and Background Estimate	62
5.3.1	Monte Carlo Based Background Estimate	63
5.3.2	Non- W background	64
5.3.3	W +jets	68
5.4	Signal Expectation	69
5.5	Scale Factor Measurement	70
5.6	Event Yield	71
5.7	Event Selection Optimization	76
5.8	Systematic Uncertainties	82
6	Results	86
6.1	Limit Calculation	86
6.2	Dijet Invariant Mass	88
6.3	Expected and Observed Limit	89
7	Conclusions	94

List of Tables

3.1	τ lepton decay modes	35
3.2	τ -candidate identification cuts	37
4.1	NNLO production cross section for WH	49
4.2	Higgs boson decay branching fractions.	50
4.3	Background processes and their cross sections and errors.	51
5.1	TAU_MET Trigger Requirements, version 16	58
5.2	Event Level cuts	61
5.3	Pretag sample background composition	73
5.4	Single Tag sample background composition	74
5.5	Double Tag sample background composition	74
5.6	\cancel{E}_T^{perp} cut optimization in Single Tag sample	77
5.7	\cancel{E}_T^{perp} cut optimization in Double Tag sample	78
5.8	Single Tag signal sample	79
5.9	Double Tag signal sample	79
5.10	Systematic uncertainties, Single Tag sample	85
5.11	Systematic uncertainties, Double Tag sample	85
6.1	95% C.L. limit, Single and Double Tag samples combined	91
6.2	95% C.L. limit, Single Tag sample	92
6.3	95% C.L. limit, Double Tag sample	93

List of Figures

1.1	Higgs potential in $\phi_1 - \phi_2$ space.	8
1.2	Fermionic one loop correction to the Higgs boson mass	11
1.3	Feynman diagrams for Higgs (Φ) production at a hadron collider. . .	12
1.4	Higgs production cross section at hadron colliders.	13
1.5	Branching ratios of the Standard Model Higgs boson decay modes. . . .	14
1.6	Theoretical bounds to the Higgs mass	15
1.7	One loop corrections to the W propagator	16
1.8	m_W and m_{top} experimental values	17
1.9	Electroweak Precision Measurements constraint on Higgs boson mass .	18
1.10	Combined CDF and D0 Higgs boson mass limit	21
1.11	Higgs boson search channels at LHC	23
1.12	ATLAS Higgs boson mass limits	24
2.1	The Fermilab accelerator chain.	26
2.2	A cross-sectional view of the CDF detector.	27
2.3	The CDF trigger and data acquisition system.	30
2.4	Level-2 calorimeter based isolation region	32
2.5	Upgraded TAU_MET trigger efficiency	33
3.1	τ signal and isolation cone	36
3.2	Jet Definition	41
3.3	Secondary Vertex Algorithm	43

3.4	b -tagging efficiency	45
3.5	Positive and negative L_{xy}	46
3.6	b -tagging mistag rate	46
4.1	Feynman diagram for the signal process	48
4.2	W+jets Feynman diagram	51
4.3	$t\bar{t}$ production Feynman diagram	52
4.4	Single top production Feynman diagrams	53
4.5	One of the Feynman diagrams for Z +jets productions	54
4.6	Diboson production Feynman diagrams	54
4.7	Multijet production Feynman diagram	55
4.8	τ -candidate track multiplicity	56
5.1	Fit to \cancel{E}_T distribution, Pretag region	65
5.2	\cancel{E}_T^{perp} shape	66
5.3	Fit to \cancel{E}_T^{perp} distribution, Pretag region	67
5.4	Fit of the \cancel{E}_T^{perp} distribution in the Pretag sample	72
5.5	Fit of the \cancel{E}_T^{perp} distribution in the Single Tag sample	72
5.6	Fit of the \cancel{E}_T^{perp} distribution in the Double Tag sample	73
5.7	Data-Background Model comparison, Single Tag sample	75
5.8	Data-Background Model comparison, Double Tag sample	75
5.9	\cancel{E}_T^{perp} cut optimization	78
5.10	Data-Background Model comparison, Single Tag sample, signal region	80
5.11	Data-Background Model comparison, Double Tag sample, signal region	81
5.12	Effect of JES uncertainty on dijet mass distribution, Single Tag sample	82
5.13	Effect of JES uncertainty on dijet mass distribution, Double Tag sample	83
6.1	M_{jj} - \cancel{E}_T^{perp} correlation	89
6.2	M_{jj} distribution in Single and Double Tag samples	90
6.3	95% C.L. limit, Single and Double Tag samples combined	91

6.4	95% C.L. limit, Single Tag sample	92
6.5	95% C.L. limit, Double Tag sample	93
7.1	Tevatron projected sensitivity, $m_H = 115 \text{ GeV}/c^2$	96
7.2	ATLAS projected sensitivity	97

Chapter 1

Theory: The Standard Model

The modern understanding of the elementary particles and their interactions is summarized by the Standard Model (SM). Although it proved to be a very successful model, as its predictions have been tested and validated across a wide range of energies in many experiments, it is an incomplete model. Many are the experimental evidences that cannot be easily accommodated in its theoretical framework, for example the presence of neutrino masses and their values, matter anti-matter asymmetry, the presence of dark matter and dark energy in the universe. The shortcoming of the Standard Model considered in the present work is the prediction that all particles are massless, in contradiction with experimental data and our experience in the world.

A short introduction of the Standard Model and its underlying principle is proposed, and the motivation for a mechanism that could provide electroweak symmetry breaking is explained. The Higgs mechanism is then introduced, as the simplest solution to the generation of fermion and boson masses. The last section discusses the production and decay modes of a Higgs boson at a hadron collider. Finally, current theoretical and experimental limits on the Higgs boson mass are discussed.

1.1 The Standard Model

The Standard Model is an effective field theory that describes the properties of the elementary particles and their interactions [1].

Particles are classified depending on their spin. If it is half integer they are called fermions, as they follow the Fermi statistic. If it is integer, they follow the Bose-Einstein statistic, and they are called bosons. Ordinary matter is made of fermions, that are the building blocks of the Universe. Bosons, instead, mediate the interactions between particles.

Fermions in the Standard Model are classified as either leptons or quarks.

There are three types of charged leptons, the electron (e), the muon (μ) and the tau (τ), ordered by increasing mass. They interact electromagnetically, as well as through the weak force. Associated to each charged lepton, there are neutral particles, called neutrinos. They only interact weakly. The 3 lepton families are

$$\begin{pmatrix} e \\ \nu \end{pmatrix} \begin{pmatrix} \mu \\ \nu \end{pmatrix} \begin{pmatrix} \tau \\ \nu \end{pmatrix}.$$

Quarks as well are organized in 3 families:

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}.$$

Quarks, together with gluons, are the constituents of the protons and neutrons (and all other hadrons). The *up* (u), *charm* (c) and *top* (t) carry electrical charge $-1/3$, while the *down* (d), the *strange* (s) and the *bottom* (b) carry electrical charge $+2/3$. They interact both via the electromagnetic and the weak force. They also carry another charge, called *color* charge, that allows them to interact via the strong force. There are three *colors*: red, blue and green, but the bare charges have never been observed. The quarks, in fact, always group together in *color*-less objects, called mesons in case of a pair of quarks, or baryons in the case of 3 quarks combined. Each of the 12 elementary fermions has an anti-fermion associated with it, with the same mass but with the opposite spin-, baryon- and lepton- quantum numbers.

There are four fundamental forces present in the Universe: electromagnetic, weak, strong and gravitational. The electromagnetic, weak and strong interactions all arise as a consequence of the principle of gauge invariance, as we will discuss later. They are mediated by particles with integer spin called gauge bosons: the photons for the electromagnetic interaction, the W^\pm and Z for the weak interaction, and eight gluons for the strong force. It is still not clear what is the role that gravity plays, and a complete theory for quantum gravity has not been developed yet. It is also the weakest interaction, with a strength more than 20 orders of magnitude smaller than the other ones, so it is not expected to play a significant role in the physics processes described by the Standard Model.

1.1.1 Gauge Symmetries

Particles and their interaction can be described using group theory, and Noether's theorem provides the connection between symmetries and the description of the world. The theorem states that for every differentiable symmetry generated by local actions, there corresponds a conserved current.

One clear example of this is the conservation of the electric charge, that is a consequence of imposing local $U(1)$ gauge invariance. Let's consider the Lagrangian of a free spin-1/2 particle:

$$L = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi - m\bar{\Psi}\Psi \quad (1.1)$$

and the local gauge transformation

$$\bar{\Psi} \rightarrow e^{-i\alpha(x)}\bar{\Psi} \quad (1.2)$$

The second term is invariant under such transformations while the first term, that contains the derivative of Ψ , is not. To enforce invariance of the equation 1.1, it is necessary to introduce a "covariant derivative"

$$D_\mu \equiv \partial_\mu - ieA_\mu \quad (1.3)$$

where the field A_μ transforms as

$$A_\mu \rightarrow A_\mu + \frac{1}{e} \partial_\mu \alpha. \quad (1.4)$$

It is easy to verify that substituting ∂_μ with D_μ in equation 1.1, the Lagrangian is now invariant under the transformation 1.2, and takes form:

$$L = \bar{\Psi}(i\gamma^\mu \partial_\mu - m)\Psi + e\bar{\Psi}\gamma^\mu \Psi A_\mu \quad (1.5)$$

Demanding the invariance under a $U(1)$ local gauge transformation, we are forced to introduce a new vector field A_μ , that can be recognized as the electromagnetic potential, and its interaction with an electron in the second term in equation 1.5. The kinetic term for the photon field is missing in equation 1.5. For it to satisfy the gauge invariance it must take the form $F_{\mu\nu}F^{\mu\nu}$, where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (1.6)$$

is the electromagnetic field strength tensor. Equation 1.5 with the added term $\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ is the Lagrangian of QED. It is important to notice that the introduction of a mass term for the photon like $m^2 A_\mu A^\mu$ would violate the gauge invariance. This statement is true for all the gauge fields present in the Standard Model. The symmetry group that describes the Standard Model is:

$$SU(3)_C \oplus SU(2)_L \oplus U(1)_Y$$

where the subscripts correspond to the charges generating the symmetry groups.

1.1.2 Quantum Chromodynamics

The $SU(3)_C$ group corresponds to the strong force, generated by the *color* charge. Requiring the invariance of the Lagrangian 1.1 under a local color phase transformation introduces eight gauge vector fields, called gluons.

Because of the non-Abelian nature of the group, new terms appear in the Lagrangian, that describe the self-interaction of three and four gluons. These self-interacting terms are the cause of the two properties that differentiate the strong force from the electroweak one: confinement and asymptotic freedom. The first one describes the observation that quarks are never observed "naked", that is isolated, but only in *color*-less combinations: the coupling strength increases as the distance between the *colored* objects. On the other side, the coupling becomes small for short distance interactions, and quarks and gluons are quasi-free: this is called asymptotic freedom.

1.1.3 Electroweak Sector

For a long time the electromagnetic and weak interaction were thought to be independent from each other. One of the greatest successes of the Standard Model is the unification of the description of these two forces as arising from a $SU(2)_L \oplus U(1)_Y$ symmetry group [2]. The index L indicates that the weak isospin current couples only to left handed fermions, that are represented as a $SU(2)$ doublets, with weak isospin $T=1/2$ and eigenvalue associated to the third component $T^3 = \pm 1/2$. The right-handed fermions are instead weak isospin singlets, with quantum numbers $T=T^3=0$. The left handed and right handed components of the fields are given by:

$$\Psi_L = \frac{1 - \gamma^5}{2} \Psi, \quad \Psi_R = \frac{1 + \gamma^5}{2} \Psi \quad (1.7)$$

where γ^5 is the given by the product of the four Dirac matrices. For the first lepton family, $\Psi_L = (\nu_L, e_L)$ and $\Psi_R = e_R$.

The weak hypercharge Y , that generates the $U(1)_Y$ group is defined by the Gell-Mann-Nishijima relation

$$Q = T^3 + \frac{Y}{2} \quad (1.8)$$

where Q is the electric charge.

The Lagrangian of the electroweak interaction is:

$$L_{EWK} = -\frac{1}{4}W_{\mu\nu}W^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \bar{\Psi}\gamma^\mu D_\mu \Psi \quad (1.9)$$

where we use

$$W_{\mu\nu} = \partial_\mu W_{\nu\alpha} - \partial_\nu W_{\mu\alpha} - g W_\mu \times W_\nu \quad B_{\mu\nu} = \partial_\mu W_{\nu\alpha} - \partial_\nu W_{\mu\alpha} \quad (1.10)$$

and the covariant derivative is written as:

$$D_\mu \equiv i\partial_\mu - g\frac{1}{2}T \cdot W - g'\frac{Y}{2}B_\mu \quad (1.11)$$

The linear combination

$$W_\mu^\pm = \sqrt{\frac{1}{2}}(W_\mu^1 \mp iW_\mu^2) \quad (1.12)$$

describe the physical states mediating the charged weak interaction. The other two fields, W_μ^3 and B_μ , mix in such a way that the physical states are:

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \quad (1.13)$$

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W \quad (1.14)$$

where θ_W is the Weinberg angle, defined as $\tan\theta_W = g/g'$. The fields A_μ and Z_μ can be identified with the photon and the Z boson.

Since the neutral current must represent the electromagnetic interaction and the weak neutral current, it can be proven that the strength of the coupling of the electromagnetic interaction to charge fermions e (the electric charge of the positron) is related to the coupling to $SU(2)_L$ and $U(1)_Y$ symmetry group by:

$$e = g \sin\theta_W = g' \cos\theta_W \quad (1.15)$$

. The W^\pm bosons couples only with left-handed fermions, with a vertex factor:

$$-i\frac{g}{\sqrt{2}}\gamma_\mu\frac{1}{2}(1 - \gamma^5) \quad (1.16)$$

while the Z^0 boson couples with both right- and left-handed fermions, with coupling

$$-i\frac{g}{\sqrt{\cos\theta_W}}\gamma_\mu\frac{1}{2}(c_V^f - c_A^f\gamma^5) \quad (1.17)$$

where c_V^f and c_A^f are the vector and axial-vector couplings, given by

$$c_V^f = T^3 - 2\sin^2\theta_w Q \quad (1.18)$$

$$c_A^f = T^3. \quad (1.19)$$

The local gauge symmetry, though, requires all the particles of the Standard Model, fermions and boson, to be massless. So the symmetry must be broken for the particles to acquire mass.

1.1.4 Electroweak Symmetry Breaking

If the Lagrangian is required to be symmetric under a local gauge transformation, then the gauge boson fields appearing in the theory have to be massless. If this is true for the photon field, and the gluons, it is not true for the weak force, mediated by the W and Z bosons, that have masses of 80.4 and 91.2 GeV/ c^2 each. A possible solution to this puzzle is provided by the Higgs mechanism, proposed by Peter Higgs, in 1964 [3], but its validity still has to be proven experimentally.

Spontaneous symmetry breaking of a transformation group occurs when the Lagrangian exhibit a symmetry to the transformation, but the physical configuration of the field in the ground state no longer holds the same symmetry. According to Goldstone's theorem, if a continuous symmetry is broken, a new massless scalar boson appears, but this is in contrast with experiment. The new boson is equivalent to an additional degree of freedom in the theory, but with a careful choice of gauge it is possible to assign the extra Goldstone boson degree of freedom to the longitudinal polarization of the gauge boson, that acquires mass. This mechanism is not new in physics: in fact ferromagnetism and superconductivity are described by spontaneous symmetry breaking.

The electroweak interaction is represented by a $SU(2) \oplus U(1)$ gauge symmetry, so the Higgs field has to belong to $SU(2) \oplus U(1)$ multiplet [4], so it will have the form

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} \quad (1.20)$$

with

$$\Phi^+ \equiv \frac{(\Phi_1 + i\Phi_2)}{\sqrt{2}} \quad (1.21)$$

$$\Phi^0 \equiv \frac{(\Phi_3 + i\Phi_4)}{\sqrt{2}}. \quad (1.22)$$

This field is characterized by a potential V such that

$$V = \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2. \quad (1.23)$$

If $\mu^2 \leq 0$ and $\lambda > 0$ the potential have the form of a Mexican hat (see figure 1.1) and

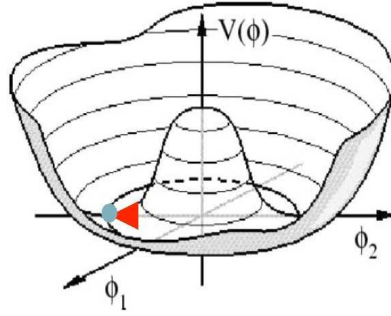


Figure 1.1: Higgs potential in $\phi_1 - \phi_2$ space.

any choice of a ground state will apparently break the symmetry. The ground state for the Higgs field must break the $SU(2)$ and $U(1)_Y$ symmetries, but not the $U(1)_{em}$, since the Higgs mechanism needs to provide mass for the W and Z boson, but must not affect the photon. The appropriate choice for the ground state that guarantees this is:

$$\langle \phi_0 \rangle = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (1.24)$$

with $v = \sqrt{-\mu^2/\lambda}$.

A perturbation around the minimum can be written as

$$\phi_0 = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (1.25)$$

When it is inserted in the Lagrangian for the electroweak interaction, three of the four degree of freedom associated with the Higgs field are absorbed by the W and Z boson as longitudinal polarization. In this way, the W and Z bosons acquire mass, while the photon stays massless, as required by construction of the choice of ground state.

$$M_W = \frac{1}{2}vg \quad M_Z = \frac{1}{2}v\sqrt{g'^2 + g^2} \quad M_\gamma = 0 \quad (1.26)$$

The only component left of the Higgs field is the scalar particle h, electrically neutral, described by

$$L = \frac{1}{2}\partial_\mu h \partial^\mu h - \lambda v^2 h - \lambda v h^3 - \frac{\lambda}{4}h^4. \quad (1.27)$$

and with mass:

$$M_h = \sqrt{2v^2\lambda}. \quad (1.28)$$

However, the theory does not yield direct information on the Higgs mass. The vacuum expectation value $v = 246$ GeV can be calculated from the W and Z mass, but λ remains a free parameter. Being able to produce and study the Higgs boson, or whatever mechanism is driving the spontaneous symmetry breaking, is one of the most exciting prospects for particle physics right now.

We should notice that in the Standard Model, the coupling of the Higgs boson to the gauge bosons are known, allowing us to predict of the production cross section for the Higgs boson at hadron colliders.

The presence of a Higgs boson, introduced to explain the electroweak symmetry breaking and the mass of W^\pm and Z^0 gauge boson, allows the introduction of fermion masses in the Lagrangian. Since left-handed fermions are a SU(2) doublet, while right-handed are a SU(2) singlet, a mass term of the form

$$m(\bar{f}_L f_R + \bar{f}_R f_L) \quad (1.29)$$

is not $SU(2)$ invariant. The presence of the Higgs doublet allows a Yukawa-like term like:

$$L_m = G_l((\bar{f}_L\phi)f_R) + \bar{f}_R(f_L\phi) \quad (1.30)$$

When the Higgs field is in its ground state, there are two new terms appearing in the Lagrangian. One is the mass term for the fermions, that depends on the parameter G_l , that is not predicted by the theory; the second term is the interaction between the Higgs boson and the fermion, and it is proportional to the mass of the fermion itself. This last feature, and the prediction on the coupling of the Higgs boson with the W and Z fix all the properties of the Higgs in the context of the Standard Model, except its mass.

1.2 Naturalness

The theoretical paradigm described until now shows serious shortcomings when computing the quantum correction to the Higgs boson mass. A one-loop correction, as in figure 1.2, contributes with a term that diverges quadratically with the cut-off energy Λ at which the Standard Model ceases to work,

$$m_h = (m_h^2)_0 + \frac{g^2 k}{16\pi^2} \Lambda^2, \quad (1.31)$$

making the Higgs boson the element of the Standard Model most sensitive to the presence of new physics, and hence even more interesting.

If the scale of new physics Λ is much larger than the electroweak scale, some unnatural cancellations must occur for the Higgs boson mass to be of the order of a few GeV/c^2 , as the unitary constraints and the direct and indirect experimental limits suggest.

It is worthwhile to note that the corrections to m_H have opposite signs if they are due to boson or fermion loops. This observation suggests that there could be a

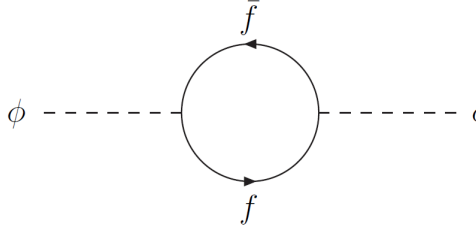


Figure 1.2: Fermionic one loop correction to the Higgs boson mass

new symmetry that relates bosons and fermions such that for every fermion in the theory a new boson is introduced, and for every boson a new fermion: this is referred to as SUSY. In this case, the corrections to the Higgs mass for a particle and its super-partners will cancel out, resolving the so called hierarchy problem.

1.3 Higgs production mechanisms and decay modes

Since the couplings of the Higgs boson with the Standard Model particles are known, it is possible to calculate the Higgs boson production cross section and the branching ratios of its decay modes.

Figure 1.3 shows the Feynman diagrams for some of the most important production mechanisms of the Higgs boson at a hadron collider. Figure 1.4 shows the expected cross sections as a function of mass for the $p\bar{p}$ collision at the Tevatron, and pp collision at the LHC.

The most important modes are:

- $gg \rightarrow H$: gluon fusion has the highest cross section both at the Tevatron (1.2 pb for $m_H = 115 \text{ GeV}/c^2$) and LHC (18.2 pb for $m_H = 115 \text{ GeV}/c^2$) across the whole mass range of interest. Gluon fusion happens mostly through top-loop, since top is the heaviest quark.
- $q\bar{q} \rightarrow V^* \rightarrow VH$ where $V = W$ or Z boson. Higgs-strahlung or associated

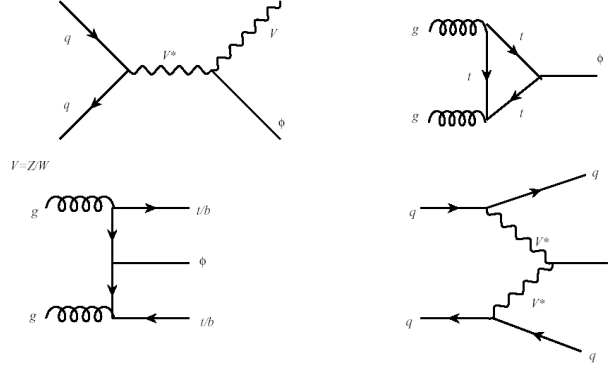


Figure 1.3: Feynman diagrams for Higgs (Φ) production at a hadron collider. Top left, associated production with a gauge boson (WH and ZH). Top right gluon fusion through top loop ($gg \rightarrow H$). Bottom left: production in association with top pair. Bottom right: Vector Boson Fusion.

production with a vector boson is the second leading production mechanism at the Tevatron for $m_H \leq 180 \text{ GeV}/c^2$, with a cross section of about 0.28 pb ($m_H = 115 \text{ GeV}/c^2$). It contributes at LHC as well (cross section about 1.2 pb), although it is not as enhanced in pp collision as gluon fusion is.

- $q\bar{q} \rightarrow q\bar{q}H$: vector boson fusion has the second largest cross section for $m_H \geq 180 \text{ GeV}/c^2$ at the Tevatron ($\sim 0.2 \text{ pb}$ for Higgs masses above 180 GeV/c^2). It is the second highest cross section at the LHC on the all mass region, with a value similar to the sum of $\sigma(pp \rightarrow WH) + \sigma(pp \rightarrow ZH)$ for $m_H = 115 \text{ GeV}/c^2$ ($\sigma(pp \rightarrow q\bar{q}H) = 1.3 \text{ pb}$). In this process, the quark and anti-quark radiate a vector boson pair which annihilates into a Higgs boson, and it shows the distinctive signature of two forward jets.
- $t\bar{t}H$: Higgs associated production with a $t\bar{t}$ pair has the smallest cross section, both at the Tevatron and LHC. Because the cross section is proportional to the top Yukawa coupling, it is an important production mechanism to study the Higgs boson properties.

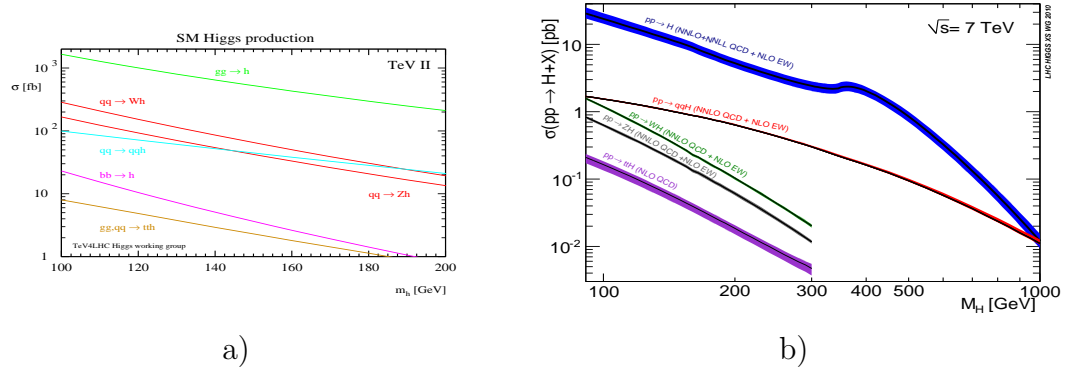


Figure 1.4: Production cross sections for the Standard Model Higgs in $p\bar{p}$ collisions with $\sqrt{s} = 1.96$ TeV center-of-mass energy at the Tevatron [5]. b) Production cross sections for the Standard Model Higgs in pp collisions with $\sqrt{s} = 7$ TeV center-of-mass energy at the LHC [6].

Figure 1.5 shows the branching fractions of the Higgs boson. If $m_H \leq 135$ GeV/ c^2 the Higgs decays mostly to b-quarks pairs. For $m_H \geq 135$ GeV/ c^2 , the decay $H \rightarrow W^+W^-$ is the dominant one.

1.4 Theoretical Constraints on the Higgs Mass

The mass of the Higgs boson, as we said earlier, is a free parameter in the theory, but it is possible to set constraints on it using theoretical considerations [7].

1.4.1 Unitarity

A limit on the Higgs boson mass can be set studying the scattering of longitudinal gauge bosons. At very high energy, where $s \gg m_H$, the gauge boson scattering cross section grows with m_H . For unitarity to be preserved

$$m_H < 760 \text{ GeV}/c^2. \quad (1.32)$$

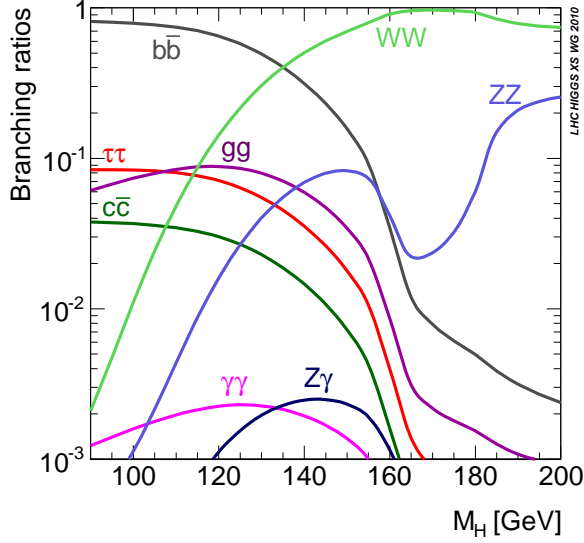


Figure 1.5: Branching ratios of the Standard Model Higgs boson decay modes.

In case the Higgs boson is heavier than the energy scale, the unitary condition set the critical scale, s_c at which perturbative unitarity is violated

$$\sqrt{s_c} < 1.2 \text{ TeV} \quad (1.33)$$

This is one of the reasons why the scientific community expects new physics at the TeV scale, if the Higgs boson does not exist.

1.4.2 Triviality and Vacuum Stability

If m_h is too large, the Higgs self-coupling λ increases, and the theory would cease to be perturbative. Requiring that this does not happen, and that the Higgs quartic coupling is finite at a scale Λ defines an upper bound on its mass. This bound is shown as the upper band in figure 1.6, above which perturbation theory does not work anymore.

A lower bound on the Higgs boson mass can be set requiring that the spontaneous symmetry breaking occurs, and that λ remains positive at all scale of Λ . This bound

is shown as the lower band in figure 1.6.

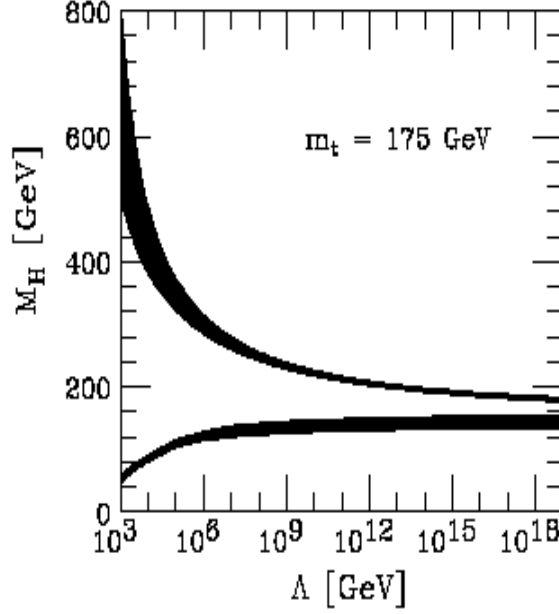


Figure 1.6: Theoretical bounds to the Higgs mass from triviality and vacuum stability, as a function of the energy scale at which new physics enters. The allowed region is the area between the two bands. See [8] and reference within for more details.

1.5 Experimental Limits on Higgs Mass

In this section we present the constraints on the Higgs mass due to direct and indirect experimental evidences.

1.5.1 Indirect Limits

The Higgs boson contributes to the propagator of the W^\pm and Z^0 boson, as in figure 1.7. Hence, the Higgs boson mass can be deduced from the precision measurements of the top and W mass, assuming that the Standard Model is the correct theory of nature. Since the corrections are logarithmic functions of the Higgs boson mass, the

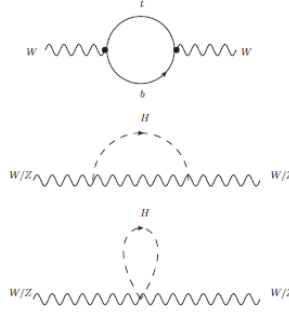


Figure 1.7: One loop correction to the W propagator from Higgs boson and top quark. The corrections depend quadratically on the top mass, and logarithmically on the Higgs boson mass.

constraints on the Higgs mass achieved in this fashion are not very stringent, but a comparison between a direct measurement of the Higgs boson with the electroweak precision measurements will be a powerful test of the validity of the Standard Model. Figure 1.8 shows that the current best measurements for the W^\pm and top quark prefer a light Higgs.

The most stringent indirect limit on the Higgs mass is set by a global fit of all the high- Q^2 precision electroweak measurements done at the Tevatron and LEP. Figure 1.9 shows $\Delta\chi^2 = \chi^2 - \chi^2_{\text{minimum}}$ of the fit as a function of the Higgs mass, while the bands represent an estimate of the theoretical uncertainties. The yellow area shows the values of the Higgs mass already excluded at 95 % C.L. by direct searches at LEP and Tevatron, as of June 2011 [9]. The valued preferred by the fit is:

$$m_H = 92^{+34}_{-26} \text{ GeV}/c^2 \text{ at } 68\% \text{ C.L.} \quad (1.34)$$

The χ^2 distribution can also provide a one-sided 95% C.L. upper limit on the Higgs boson mass, and, considering also the LEP exclusion:

$$m_H < 185 \text{ GeV}/c^2 \text{ one sided } 95\% \text{ C.L.} \quad (1.35)$$

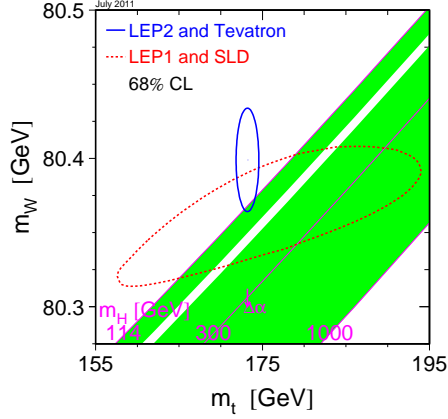


Figure 1.8: Values of m_W and m_{top} directly measured at LEP2 and the Tevatron (solid line) and indirectly inferred in LEP1 and SLD data (dashed line). The green area shows the relationship between the masses of the W boson and top quark for Higgs mass not excluded as of June 2011 (direct searches at the Tevatron excluded $156 \text{ GeV}/c^2 < m_H < 177 \text{ GeV}/c^2$ at 95 % C.L., see [9]). See [10] for more details.

1.5.2 Direct searches at LEP

Before the Tevatron, the most stringent limit on the Higgs boson mass was set by the experiments collecting data at the LEP collider at CERN. LEP was an e^+e^- collider located at CERN, that started taking data at a center-of-mass energy near the Z^0 pole (LEP1). It was then upgraded to reach center of mass energy up to 209 GeV. The main production mechanism at LEP is the Higgs-strahlung from a Z boson (see fig 1.3), and the Higgs boson production cross section depends on the energy of the center-of-mass. LEP1 was able to set a lower limit of $65 \text{ GeV}/c^2$ in Higgs mass, while the LEP2 experiment ALEPH reported an excess of 3 standard deviations for a Higgs mass of $115 \text{ GeV}/c^2$ [11]. The other CERN experiments DELPHI, L3 and OPAL did not confirm the excess, and combining their data it was not possible to exclude at 95% confident level a Higgs mass of $115 \text{ GeV}/c^2$ [12]. In LEP1, the search was performed only in events where the Z decays to charged leptons or neutrinos, and the Higgs boson to $b\bar{b}$ pairs ($ZH \rightarrow llb\bar{b}$ and $ZH \rightarrow \nu\bar{\nu}b\bar{b}$), while during LEP2

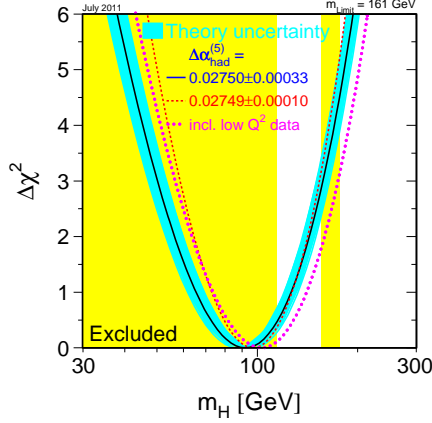


Figure 1.9: $\Delta\chi^2$ as a function of the Higgs mass, for the fit including all high Q^2 data. The band represents an estimate of the theoretical error due to missing higher order corrections. The vertical band shows the 95% C.L. exclusion limit on m_H from the direct searches at LEP2 (up to 114 GeV/ c^2) and the Tevatron (158 GeV/ c^2 to 175 GeV/ c^2).

the Higgs boson was searched in more channels, exploring all the possible decays mode of the Z boson.

1.6 Current status of Higgs searches at Hadron Colliders

1.6.1 Direct searches at the Tevatron

The Tevatron, the $p\bar{p}$ collider at Fermilab operating at $\sqrt{s} = 1.96$ TeV, ended its operation on September 30th 2011, after delivering to CDF and D0 $\sim 12 \text{ fb}^{-1}$ of data. Many have been the contributions of the Tevatron to our understanding of particle physics in the past 35 years, but one may think that its search program is completed, and the baton passed to the LHC, which is exploring a new energy regime. I think that the Tevatron data could still give an important contribution to the exclusion or discovery of the Higgs boson, and to the measurement of its properties, and in fact,

both CDF and D0 collaborations are devoting a substantial fraction of their effort to Higgs searches. If the Higgs boson has a mass just above the LEP limit, observing $H \rightarrow b\bar{b}$ final states and measuring its cross section would be important to test the nature of the Higgs boson and its couplings. Although studying this decay mode is not impossible at LHC, it will be very challenging, while this is the most sensitive final state at the Tevatron.

The Higgs program at the Tevatron involves studying many different samples, each one optimized to target one specific final state, mostly classified by the Higgs decay mode. It is important to highlight that there is no channel that is, by itself, sensitive to the Higgs boson, and it has become clear that combining multiple independent channels is the best strategy.

The Higgs boson production mechanism with the highest cross section at the Tevatron is gluon fusion, but, depending on the Higgs mass, different search strategies provide the best sensitivity.

If $m_H \geq 135 \text{ GeV}/c^2$, the highest branching fraction is for Higgs decaying to W^+W^- : in the case where both the W bosons decay into e or μ , the dominant background is the irreducible W^+W^- electroweak production, plus pair produced leptons from photon or Z decay (Drell-Yan) or W production in association with jets. The signal-to-background ratio (S:B) in this final state is 1:50, but the angular separation between the leptons gives some separating power that improves the sensitivity of the channel. The other relevant branching fraction is $H \rightarrow ZZ$: in case where both the Z bosons decay into leptons, the channel has a S/B of $\sim 1:30$ for $m_H = 150 \text{ GeV}/c^2$, where the background is mostly non-resonant ZZ production.

If $m_H \leq 135 \text{ GeV}/c^2$, the main decay mode is $H \rightarrow b\bar{b}$, but a final state of only two b -jets is indistinguishable from the QCD production of b -quarks pairs, whose cross section is $\sim 10^9$ higher than the Higgs boson. The best S/B is achieved in the channel where the Higgs is produced in association with a W/Z boson. Because of this, the low mass Higgs boson is searched for in different final states:

- $WH \rightarrow l\nu b\bar{b}$: this final state considers events with an isolated high p_T lepton (either e or μ), in association with \cancel{E}_T , and two high p_T jets. In case both the jets are identified as originating from a b -quark, the S/B is $\sim 1 : 100$ so a counting experiment is not possible. Multivariate techniques are used to distinguish the signal from the backgrounds, and are used to test the presence of the Higgs boson. Since no signal seems present in data, a 95 % C.L. upper limit is set on the the Higgs boson cross section times $\text{Br}(H \rightarrow b\bar{b})$. The observed limit for $m_H = 115 \text{ GeV}/c^2$ is $\sigma(95\% \text{C.L.})/\sigma_{SM} = 3.64$ and it is in good agreement with the expected limit. When a gluon is emitted from the initial or the final state (initial- or final-state radiation), WH events can be characterized by three jets in the final state. When this class of events is analyzed as well, and its result combined with the one above, the 95 % C.L. upper limit on the Higgs production rate times $\text{Br}(H \rightarrow b\bar{b})$ is 2.65 times the Standard Model expectation.
- $VH \rightarrow \cancel{E}_T b\bar{b}$: events with high transverse missing energy and two jets could result from $ZH \rightarrow \nu\bar{\nu} b\bar{b}$ and $WH \rightarrow (l)\nu b\bar{b}$ decays, the latter in case the lepton escapes identification and contributes to the \cancel{E}_T in the event. This analysis suffers from a high QCD multijet background, where the \cancel{E}_T arises from jet energy mis-measurement: a Neural Network is used to reduce its contribution. The other powerful handle against background is b -tagging: when both the jets are consistent with originating from a b -quark, the S/B $\sim 1 : 60$. The 95 % C.L. upper limit on the Higgs mass times $\text{Br}(H \rightarrow b\bar{b})$ is 2.3 time the Standard Model expectation.
- $ZH \rightarrow ll b\bar{b}$ although this process has a small branching ratio, it is a powerful channel because of the low backgrounds. It has a S/B $\sim 1 : 50$.

Many other channels are explored in the Higgs quest at the Tevatron: associated production of the Higgs where all particles decay into quarks, channels where the

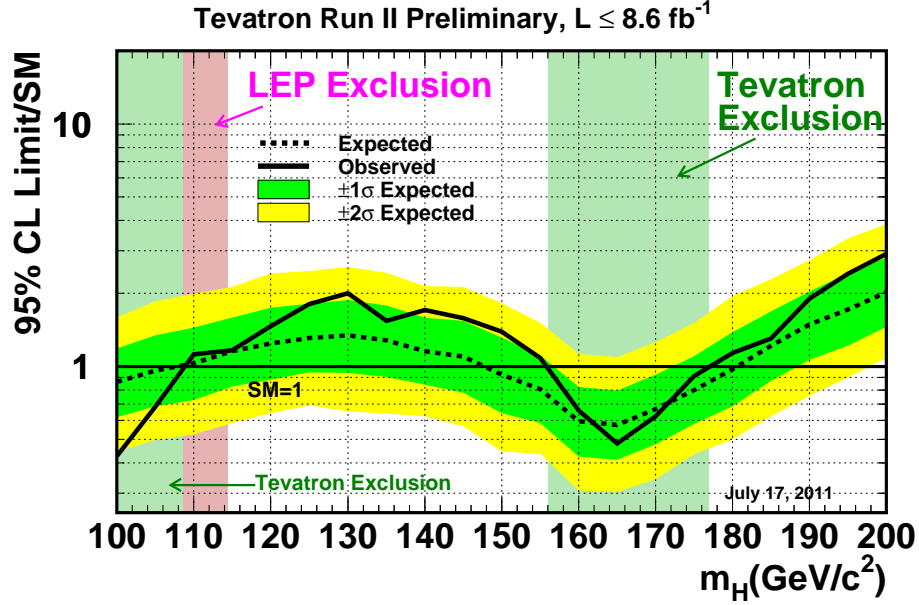


Figure 1.10: The observed and expected 95% C.L. limits on σ/σ_{SM} . The expected limit is the median for the background-only hypothesis. The bands indicate the 68% and 95% probability regions where the limit can fluctuate, in absence of signal [15].

Higgs decays into τ lepton or γ pairs or associated production of the Higgs with top pairs. If considered individually, each one of these analysis reaches a sensitivity of 5 to 10 times worse than the primary channels that were just briefly mentioned, but, once combined, they contribute as one of the primary channels. They are also more sensitive at a intermediate mass range $m_H > 130 \text{ GeV}/c^2$.

It is also clear that CDF nor D0 alone will be able to achieve by itself the sensitivity in the full Higgs mass range, and the collaborations have agreed to combine their results. Figure 1.10 shows the latest combination produced for the conferences during summer 2011: the dashed line is the expected limit, and the solid one the observed. In combination, the Tevatron experiments excluded a Higgs boson with

mass:

$$100 < m_H < 108 \text{ GeV}/c^2, \quad 156 < m_H < 177 \text{ GeV}/c^2 \text{ at } 95\% \text{ C.L.}$$

1.6.2 Direct searches at the LHC

The LHC, located in Geneva, Switzerland, is a pp collider designed to reach center-of-mass energy of $\sqrt{s} = 14 \text{ TeV}$. Because of technical problems, high energy collisions started in March 2010, at $\sqrt{s} = 7 \text{ TeV}$. Up to now, the collider has delivered more than 4 fb^{-1} of integrated luminosity each to the ATLAS and CMS multi-purpose detectors.

The choice of having pp colliding was dictated by two considerations: anti-protons production and storage is challenging, and the highest Higgs boson production cross section is gluon fusion. The technical challenges associated with anti-protons have always been a limiting factor for the Tevatron instantaneous luminosity, but one of the main goal to the Tevatron program was to discover the top quark produced via quark-anti-quark annihilation, so it was necessary to use anti-protons. The LHC main goal is to discover the mechanism causing electroweak symmetry breaking. The minimal extension of the Standard Model that provides such a mechanism is the inclusion of a Higgs boson and its non-zero vacuum expectation value. If this is the correct theory of Nature, the Higgs boson highest production cross section is gluon fusion through a top loop, so there is no need to produce anti-protons, since the gluon content in protons and anti-protons is the same.

Figure 1.11 shows the product of the Higgs boson production cross section times its branching ratio as a function of m_H . Each line corresponds to a different final state used to search for the Higgs boson at LHC. As at the Tevatron, the best sensitivity to the presence of the Higgs boson is achieved by studying data samples associated with each of the individual final states shown in figure 1.11 , and then combining them. Figure 1.12 shows the 95 % C.L. upper limit on the Higgs production cross section times branching ratios divided by the Standard Model expectation for the

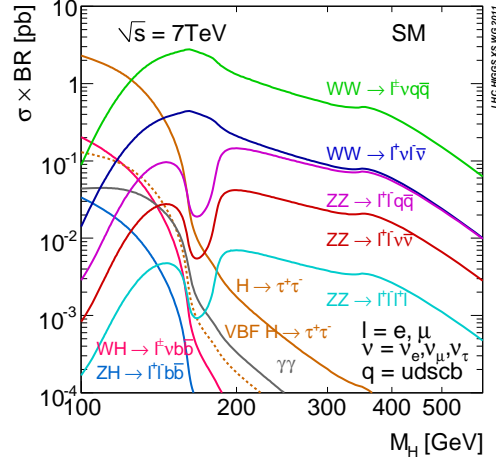


Figure 1.11: Higgs boson production cross section times branching ratios. Each line corresponds to a final state used at LHC for the Higgs search, from [16].

individual channels studied at ATLAS, and the combined results. ATLAS alone excludes at 95% C.L. the existence of the Higgs boson in a mass range:

$$146 < m_H < 232 \text{ GeV}/c^2, \quad 256 < m_H < 282 \text{ GeV}/c^2, \quad 296 < m_H < 466 \text{ GeV}/c^2$$

Similar search strategies are used by the CMS collaboration, that achieves comparable sensitivities [18], and excludes a Higgs boson with mass:

$$145 < m_H < 216 \text{ GeV}/c^2, \quad 226 < m_H < 288 \text{ GeV}/c^2, \quad 310 < m_H < 400 \text{ GeV}/c^2$$

All the direct searches performed so far bound the Higgs boson mass to

$$114 < m_H < 145 \text{ GeV}/c^2. \quad (1.36)$$

It is interesting to note that it is also the region preferred by the global fit to the Standard Model precision measurements.

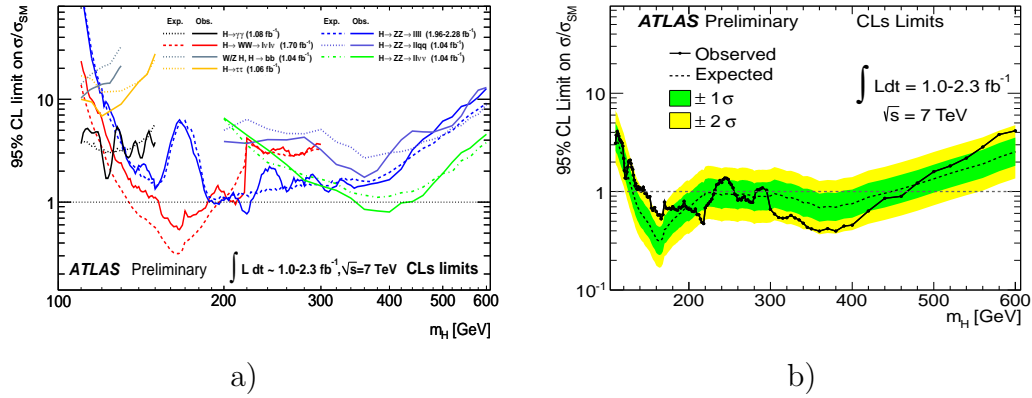


Figure 1.12: a) Expected and observed 95 % C.L. upper limits for the individual channels set at the ATLAS experiment. At very high mass $m_H \geq 200 \text{ GeV}/c^2$, $H \rightarrow ZZ$ is the dominant channel. At intermediate masses, $135 \text{ GeV}/c^2 < m_H < 200 \text{ GeV}/c^2$, the most sensitive channel is $H \rightarrow WW \rightarrow l\nu l\nu$. For $m_H < 135 \text{ GeV}/c^2$, several channels exhibit comparable sensitivity. b) ATLAS 95% C.L. combined upper limit on Higgs production divided by Standard Model expectation. [17].

The Standard Model is a very successful effective theory, but it necessitates a mechanism to explain electroweak symmetry breaking. There are many different extensions of the Standard Model that try to address this crucial issue, such as as Little Higgs model [19], or the strong symmetry breaking of technicolor models [20]. The simplest one, the Higgs mechanism, is what we will consider in this work. We perform a search for the Higgs boson in data acquired by the CDF detector at the Tevatron collider. Since the Higgs boson is most likely to be light, we concentrate on $H \rightarrow b\bar{b}$ final state. Many analysis are already performed at CDF in final states where the Higgs boson is produced in association with a W or Z boson. This works differs because it consider events where the W boson decays into τ leptons.

Chapter 2

Experimental Apparatus

The Higgs Boson is searched for in high energy protons and anti-protons collisions produced at the Fermi National Laboratory (Fermilab) in Batavia, Illinois, USA. Protons and anti-protons are accelerated through a series of accelerators, and collide at a final center-of-mass energy of $\sqrt{s} = 1.96$ TeV in two locations around the Tevatron. To investigate the products of the collisions, two experiments are located at the interaction points: D0 and CDF. The CDF II detector is used for this analysis. In this chapter, the Fermilab accelerator chain and the CDF II detector are briefly described.

2.1 The Tevatron Accelerator Complex

The Tevatron was the world's highest energy hadron collider until December 2009, when the Large Hadron Collider, located at CERN, in Geneva, Switzerland, produced the first proton-proton collisions at center of mass energy of 2.25 TeV.

At Fermilab, the proton and anti-proton beams are accelerated through a chain of particle accelerators, see figure 2.1, before being injected in the Tevatron, a synchrotron of radius of ~ 1 km, where they are accelerated in opposite directions to an energy of 980 GeV.

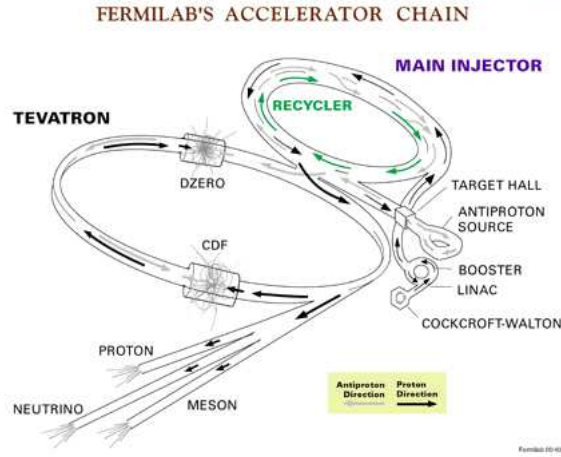


Figure 2.1: The Fermilab accelerator chain.

Producing protons is an easy task starting from Hydrogen atoms. Producing anti-protons is not that easy: 120 GeV protons from the Main Injector, when striking a nickel target, produce numerous secondary particles, anti-protons included. Magnets are used to select only the anti-protons, which are then collected and stored in a beam in the Recycler. Protons and anti-protons are injected in the Tevatron in 3 trains, with 12 bunches per train, each bunch separated by 396 ns. The proton beam has typically 10^{13} particles, and the anti-proton beam 2×10^{12} particles. They collide in designated locations on the ring, called the B0 and D0 points where the CDF and D0 detectors are located, to produce rare particles at a center-of-mass energy of 1.96 TeV. The amount of anti-protons per bunches is one of the main limiting factor to the number of primary interactions possible at the Tevatron. More details on the all accelerator chain and its individual components can be found in [21].

The number of collision per second is described by "luminosity", L . It depends on the collision frequency (f), the number of the protons and anti-protons in each beam ($N_P, N_{\bar{P}}$) and the average transverse beam size (A) as:

$$\mathcal{L} = f \frac{N_P N_{\bar{P}}}{4\pi A}. \quad (2.1)$$

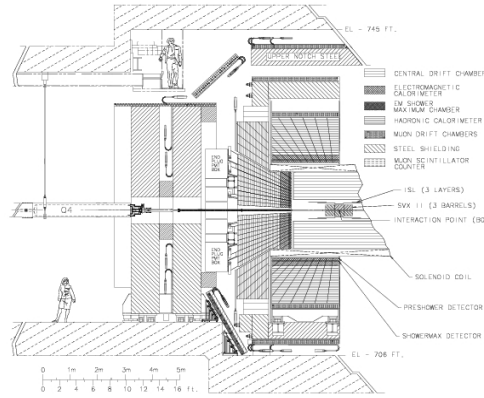


Figure 2.2: A cross-sectional view of the CDF detector.

The Tevatron ended its operations on September, the 30th. Since 2001, when RunII started, it delivered an integrated luminosity of 12 fb^{-1} , while 10 fb^{-1} of data were recorded at CDF. The highest instantaneous luminosity achieved at the Tevatron is $4.3 \cdot 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$.

2.2 The CDF II Detector

The CDF II detector is a magnetic spectrometer with cylindrical symmetry around the beam pipe, surrounded by calorimeters and muon detectors, see figure 2.2. It is described in details in [22] and [23]. Since it has a barrel-like shape, it uses a cylindrical coordinate system (r, ϕ, z) . The origin is located at the center of the detector, r is the radial distance from the beamline and the z -axis goes along the direction of the proton beam (towards the east). Spherical coordinates (ϕ, θ) are also commonly used, where ϕ is the azimuthal angle around the beam axis and θ is the polar angle with respect to the proton beam axis. The pseudorapidity η is defined as $\eta = -\ln \tan \theta/2$. The charge particle tracking system consists, in its innermost component, of six layers of double-sided silicon-microstrip sensors at radii between 2.5 and 22 cm from the beam, covering a region $|\eta| < 2$. It reconstructs displaced

tracks with a resolution of $15\ \mu m$, and it is an essential component of the B -hadron identification. It is surrounded by a 96-layer drift chamber, with coverage $|\eta| < 1$, immersed in a 1.4 T solenoidal magnetic field, parallel to the particle beam. Outside the solenoid there are sampling electromagnetic and hadronic calorimeters. A set of wire and strip chambers, embedded in the calorimeter at a depth of approximately 6 radiation lengths, where energy deposition of EM showers reaches its maximum, help to distinguish photons from electrons. A Cherenkov luminosity counters is used to measure the luminosity of $p\bar{p}$ collisions at CDF.

Most of the subdetectors are used for the current analyses, but only the trigger system will be discussed in details, as I was involved with it.

2.3 The Trigger System

CDF II uses a three-levels system to reduce the 1.7 MHz bunch crossing rate to 150 Hz written on tape [24]. Data are acquired at the collisions frequencies, that is at 7.6 MHz, and stored in a 42 clock cycle long pipeline. For a given collision, the Level-1 trigger forms a decision, based on a limited reconstruction of the muon, track and calorimeter information, in $5.5\ \mu s$. The maximum sustainable Level 1 output rate is 30 kHz.

When an event is accepted by the Level-1 trigger, the data are stored in one of the four Level-2 buffers, waiting for the Level-2 trigger decision. In the $20\ \mu s$, the Level-2 processor is performing track reconstruction in the silicon vertex detector and jet reconstruction in the calorimeter, and taking a decision. Each buffer is emptied when the Level-2 decision for the corresponding event has been asserted: if the event has been accepted, the buffer is read out, else it is simply cleared. If the Level-2 trigger decision takes too much time and the four buffers are all filled, the Level-1 accept is inhibited. This is a source of deadtime for the CDF II trigger. The maximum Level-2 output rate is 700 Hz.

The Level-3 trigger is made of a CPU farm and has a maximum output rate of 150 Hz. The events are fully reconstructed, and the trigger decision is based on the presence of electrons, photons and jets and on the total E_T and \cancel{E}_T . Figure 2.3 shows a schematic of the three levels triggering system. As a consequence of the increasing instantaneous luminosity achieved by the Tevatron in RunII, and thus of the higher occupancy of the detector, the Level-2 rate for single-jet triggers increased, and the efficiency of multijet triggers dropped. To address this, the Level-2 path for calorimeter information was upgraded in 2007-2008. In the next sections, I will describe briefly how the calorimeter information is used at trigger level, the impact of the upgrade, and my work on optimizing the TAU_MET trigger.

2.3.1 Calorimeter information at trigger level

The entire calorimeter is described, at trigger level, by a 24×24 trigger tower map, where each calorimeter trigger tower has a width $\Delta\eta=0.2$ and $\Delta\phi = 15^\circ$. The trigger tower energy data is sent to both the Level-1 and Level-2 trigger with a 10-bit energy resolution, with a least significant count of 125 MeV and a resulting full scale of 128 GeV. To reduce the complexity and the processing time, the Level-1 trigger uses only a 8-bit trigger tower energy information, by dropping the least significant bit and the most significant bit. Level-1 does not perform any clustering, and it simply selects events on the number of trigger towers above a transverse energy threshold or on the values of the computed total E_T and \cancel{E}_T . The Level-2 trigger uses the values of the E_T and \cancel{E}_T computed at low resolution by Level-1 and it has the main task to perform jet and electron clustering. This is done with a hardware-implemented algorithm which combines contiguous energetic trigger towers. Each cluster starts from a seed tower, which shows a total E_T typically above a few GeV, and includes all the contiguous lower energy towers. The cluster size expands until no more energetic adjacent towers are found. The cluster position is defined as the seed tower position. The jet clusters are also used to identify τ leptons.

Dataflow of CDF "Deadtimeless" Trigger and DAQ

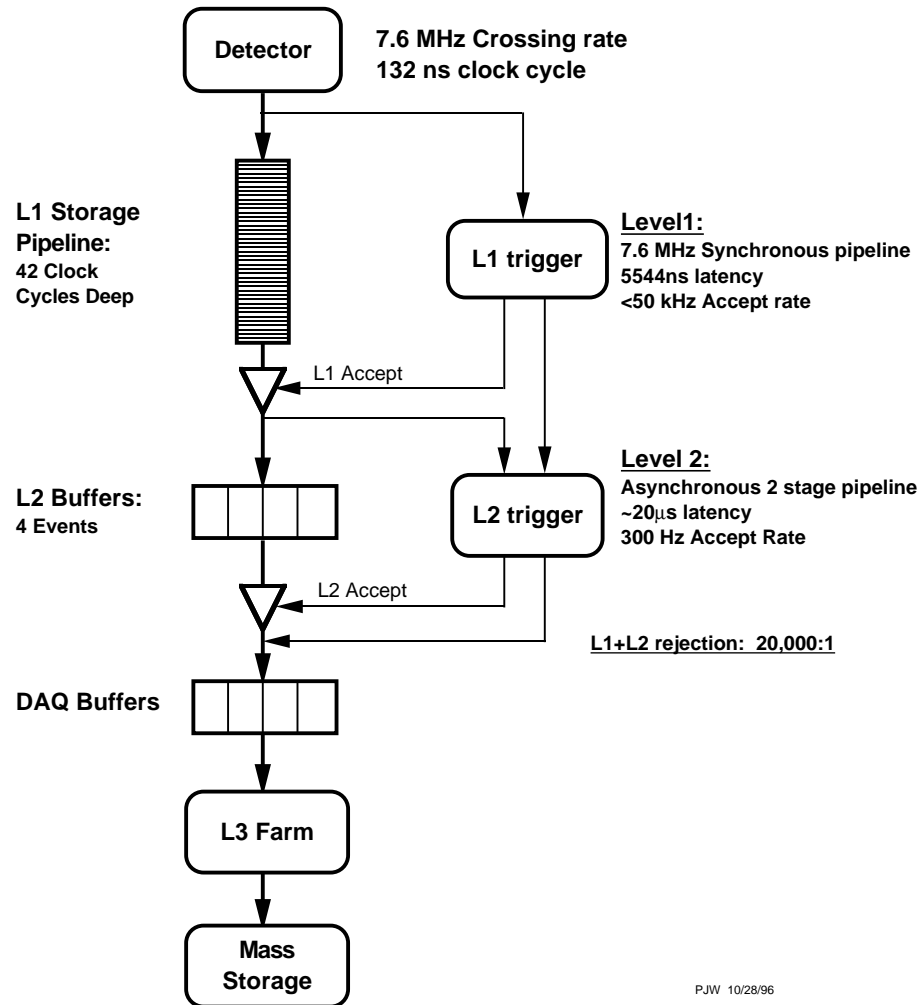


Figure 2.3: The CDF trigger and data acquisition system.

With the increase of instantaneous luminosity, the detector occupancy increases significantly as well as the presence of many energetic towers around the seed towers. As a consequence of the larger detector occupancy, large numbers of energetic towers could be erroneously clustered around a seed, into a single mis-identified jet, yielding an increase of the Level-2 rate for single-jet triggers and a significant efficiency loss for the multijet triggers due to the erroneous merging of separate lower energy jets into a smaller number of more energetic jets.

2.3.2 TAU_MET trigger and the Level-2 upgrade

The Level-2 calorimeter upgrade [25] introduced at trigger level the full 10-bit energy resolution, and the possibility of developing sophisticated software algorithms: thanks to this, jets and \cancel{E}_T reconstructed at Level-2 are almost equivalent to the offline variables.

Since we are interested in searching for $WH \rightarrow \tau\nu b\bar{b}$, we concentrated on the TAU_MET trigger path, that is designed to trigger on $W \rightarrow \tau\nu$ events. Before the upgrade, the trigger path requires, at Level-1, the presence of a calorimeter tower with $E_T \geq 10$ GeV, matched in ϕ only with a high p_T track ($p_T \geq 10$ GeV). At Level-2, a narrow cluster with $E_T \geq 10$ GeV must be matched with a high p_T isolated track, and $\cancel{E}_T \geq 20$ GeV. At Level-3, a reconstructed τ candidate with $E_T^{Cluster} \geq 20$ GeV, must be isolated. The invariant mass of the charged decay products must satisfy $m_{track} \leq 2$ GeV/ c^2 , and $\cancel{E}_T \geq 20$ GeV.

The improvement in \cancel{E}_T resolution, due to the full 10-bit energy information, increases the purity and the efficiency for the MET part of TAU_MET trigger. For the TAU part, we developed a new clustering algorithm that associates to a Level-2 τ candidate cluster a maximum of 9 tower, in a 3x3 matrix around the highest energy tower. This new clustering improves the energy resolution for τ candidate and, as a consequence, the efficiency of the trigger (see fig: 2.5a).

To improve the purity of the events selected, isolation requirements are applied at

Level-2. Before the upgrade, the isolation was based on the relative position of any track with $p_T \geq 1$ GeV/ c and the position of the high p_T track associated with the τ candidate: the τ is considered isolated if there are no tracks in the region $10 \leq \Delta\phi \leq 30$, where $\Delta\phi = \Delta\phi_{seed} - \Delta\phi_{track}$ in the transverse plane (since there is no stereo information, Level-2 tracks can only be projected into the transverse plane). This requirement does not necessarily reflect the real isolation of the τ object: in fact, tracks far from the τ can happen to fall in the isolation region and veto the trigger.

The flexibility of the upgraded Level-2 calorimeter system made it possible to develop a different kind of isolation requirement based on calorimetric energy: if the energy in the calorimeter is deposited by a jet, there is a substantial amount of energy just outside the tau cluster, that can be used to veto these objects. The simplest definition of an isolation region for a cluster is based on the neighboring towers for any of the cluster towers, as it is shown in Figure 2.4. Studies done with

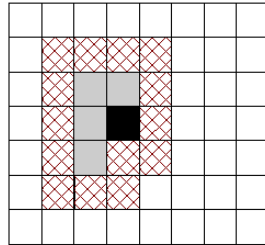


Figure 2.4: Isolation cartoon: the black tower corresponds to the seed, the gray ones are the shoulder towers of the cluster. The one crossed describe the isolation region. The size of the isolation area depends on the cluster geometry.

$W \rightarrow \tau\nu$ Monte Carlo and jets from data showed that a cut of 0.14 on the ratio of the energy in the isolation region over the energy inside the tau cluster (relative isolation energy) is sufficient to keep the trigger cross section under control up to an instantaneous luminosity of $300 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$.

2.3.3 TAU_MET efficiency in signal events

To quantify the improvement of these changes, we measured the trigger efficiency of the old and upgraded version of the level-2 trigger in the WH Monte Carlo simulated events. Figure 2.5b shows the trigger efficiency for the Level-2 part of TAU_MET in WH Monte Carlo events as a function of the visible energy of the τ candidate. The visible energy represents the energy of the τ lepton minus the energy carried by the neutrino. The red curve represents the efficiency for the old DCAS system, the black

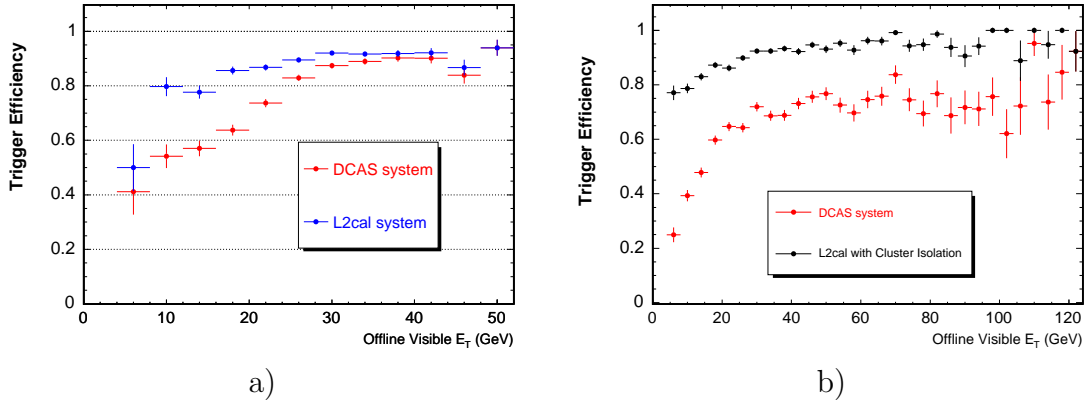


Figure 2.5: Level-2 trigger efficiency for simulated events that satisfy the TAU_MET Level-1 requirements, and that contain an identified τ -candidate. a) Level-2 efficiency as a function of τ -candidate visible E_T in $W \rightarrow \tau \nu$ simulated events. The red points represent the efficiency before the Level-2 calorimeter upgrade. The blue points shows the effect of introducing the upgraded \cancel{E}_T and τ clustering algorithm. b) Level-2 trigger efficiency as a function of τ -candidate visible E_T in $WH \rightarrow \tau \nu b \bar{b}$ simulated events. The red points represent the efficiency before the Level-2 calorimeter upgrade. The black points shows the effect of introducing the upgraded \cancel{E}_T , τ clustering algorithm, and the cluster based isolation.

one the improved trigger after the Level-2 calorimeter upgrade. The improvement in efficiency is 28%. Many are the trigger paths at CDF that use requirements on a τ at Level-2, and the algorithms we developed for the τ clustering and isolation applied to them as well.

Chapter 3

Object Identification

The particles produced in the $WH \rightarrow \tau\nu b\bar{b}$ final state are a τ lepton, a neutrino and two b -quarks. This chapter describes the algorithms used to identify the presence of these objects in a event. Since there is only one lepton in the final state, that is the τ lepton, events where an electron or muon are present are removed. Some details and references on electron and muon identification algorithms are presented here as well.

3.1 Tau Lepton

The τ lepton is characterized by a mass higher than the one of the leptons of the first two generations ($m_\tau = 1.777\text{GeV}/c^2$), a shorter lifetime (0.291 ps), and a variety of decay modes. In 35% of the case it decays leptonically to $l\nu_l\nu_\tau$, with $l = e, \mu$, and into hadrons the other $\sim 65\%$ of the time. Table 3.1 shows a summary of its decay modes. At CDF is not possible to distinguish electrons or muons from τ decays from the ones coming from a primary interaction, and so when talking about τ from now on we will consider only the ones that decay into hadrons.

For charge conservation, a τ decays to one or three (or an odd numbers of) charged hadrons, a neutrino, and, some of the time, neutral pions. The charged hadrons

Decay Mode	Final Particles	BR
Leptonic	$e^{-1}\bar{\nu}_e n u_\tau$	17.8%
	$\mu^{-1}\bar{\nu}_\mu n u_\tau$	17.4%
Hadronic 1-prong	$\pi^{-1} n u_\tau$	11.1%
	$\pi^{-1}\pi^0 n u_\tau$	25.4%
	$\pi^{-1}2\pi^0 n u_\tau$	9.2%
	$\pi^{-1}3\pi^0 n u_\tau$	1.1%
	$K^{-1} n u_\tau$	0.7%
	$K^{-1}\pi^0 n u_\tau$	0.5%
Hadronic 3-prongs	$2\pi^{-1}\pi^+\nu_\tau$	9.5%
	$2\pi^{-1}\pi^+\pi^0 n u_\tau$	4.4%

Table 3.1: τ lepton decay modes

π^\pm and K^\pm are reconstructed in CDF as high p_T tracks in the tracking chamber (COT, Central Outer Tracker) associated with energy deposition in the hadronic calorimeter. The neutral pion decays into photons pairs, that are detected with the central electromagnetic shower maximum detector (CES).

A new τ -candidate cluster is reconstructed starting from a seed tower, required to satisfy $E_T > 6$ GeV. In case any of the neighboring towers have $E_T > 1$ GeV, it is added to the cluster. Since $m_\tau/m_W \ll 1$, the τ is boosted and its decay products are collimated. Because of this and the value of m_τ , the size of the cluster is expected to be small: it is required to be ≤ 6 towers.

The charged hadrons are reconstructed as tracks in the COT, extrapolated to the calorimeter. The highest p_T track that falls into a cone of $\Delta R \leq 0.4$ around the direction of the cluster is chosen as the seed track. It is required to satisfy $p_T \geq 4.5$ GeV/ c . If other tracks with $p_T \geq 1$ GeV/ c fall into a cone of aperture $\Delta\alpha < 10^\circ$ around the seed track, they are associated to the τ -candidate as well (see figure 3.1 for a schematic representation of the τ -candidate cone).

The presence of a π^0 in the detector is inferred by its decay products, a $\gamma\gamma$ pair. A photon is reconstructed in the CDF II detector because of the energy deposition in the CEM and CES: the CES is used to determine its position, and the CEM for

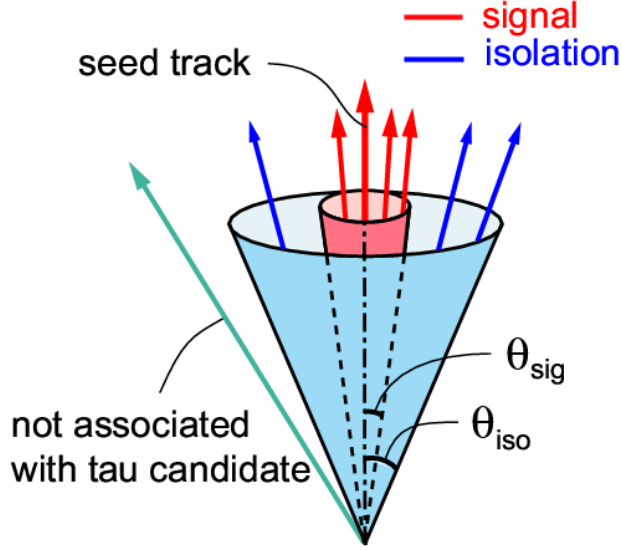


Figure 3.1: τ -candidate signal cone and isolation annulus for tracks and π^0 s.

the energy measurement. In τ decays, the π^0 is always produced in association with at least one π^\pm or a K^\pm , which sometime deposits energy in the CEM as well. The charged hadron is assumed to behave as a minimally ionizing particle(MIP) in the CEM, and its energy is subtracted from the energy in the CEM: this provides a more accurate measurement of the π^0 energy. In case the CES cluster is close to a track (distance ≤ 4 cm), it is likely to be caused by the electromagnetic shower of an electron, and the CDF algorithm fails to consider it as a π^0 . This increases the purity of the π^0 sample, but also affects the resolution of the τ energy, for which the decay products can be very collimated.

Since the neutrino escapes detection, only the energy of the visible τ decay products can be measured.

While the τ reconstruction algorithm is common to all τ -based analyses at CDF, the identification algorithm depends on the analysis and on the trigger used. In order to identify the τ in this analysis, three classes of cuts are applied: some are dictated by the need to select offline objects that would satisfy the trigger selection; others are

	Id Cut
Seed Track P_T	$> 10 \text{ GeV}$
Cluster E_T	$> 20 \text{ GeV}$
$ \eta^{seedtrk} $	< 1
Number of track with $P_T \geq 1 \text{ GeV}$ within isolation cone	0
Relative Calorimeter Isolation	≤ 0.1
Seed Track ZCes	$9 \leq \text{ZCes} \leq 230$
Seed Track $ Z_0 $	$\leq 60 \text{ cm}$
Seed Track $ D_0 $	$\leq 1 \text{ cm}$
Seed Track COT Ax. Seg.	$\geq 3(5)$
Seed Track COT St. Seg.	$\geq 2(5)$
E_T^{Vis}	$> 25 \text{ GeV}$
Track in signal cone	$= 1 \text{ or } =3$
m^{Vis}	$\leq 1.8 \text{ GeV}/c^2$
Electron rejection	$\text{Emfr} \leq 1 - 0.15/\text{EsumP}$
Number of tracks in signal cone	$=1 \text{ or } =3$

Table 3.2: τ -candidate identification cuts

cuts to ensure fiduciality and quality of the tracks; others are used to increase the purity of the τ -candidate. Table 3.2 lists the τ -candidate identification cuts used in the analysis.

In the first few lines of the table, we list the cuts that are driven by the trigger requirements. The offline cuts on the seed track (highest p_T track associated with the τ -candidate) and the cluster E_T confirm the requirements applied at trigger’s Level-2 and Level-3. Tau-candidates are only reconstructed if they are central, so the η cut would be redundant, if the trigger didn’t cut on the η of the seed track. So we added the requirement $|\eta^{seedtrk}| < 1$. The offline reconstruction algorithm at CDF uses a “shrinking cone” to define the signal cone and the isolation cone (see [28]), while the isolation of the τ -candidate at Level-3 is defined using a fixed size cone. At the trigger level, a τ -candidate is considered isolated if there is no track with $p_T > 1.0 \text{ GeV}$ reconstructed in a cone between 10° and 30° around the seed track. For this analysis, we use the fixed size cone of 10° to define the τ -candidate

quantities, so that the offline requirement on track isolation is as close as possible to the trigger one (see section 5.1 for the list of the trigger cuts).

As discussed in the previous section 2.3.2, after the Level-2 trigger calorimeter upgrade, the isolation requirement applied at Level-2 changed from track-based to calorimeter-based one. For this reason, we also apply a calorimeter-based isolation cut. The fraction of energy in a cone of $\Delta R \leq 0.4$ after subtracting the energy associated with the τ -candidate, must be less than 10% of the τ cluster energy.

The second set of cuts is used to ensure the quality of the seed track. We cut on the number of segments in the axial and stereo layers of the COT, to ensure that the track has been reconstructed correctly. The seed track must come from the primary vertex, so we require that $|Z_0^{seedtrk} - Z_0^{vtx}| \leq 1$ cm, and we require the track to have a small impact parameter. It is also important to make sure that the seed track is not reconstructed at the edge of the CES: this reduces the contribution of events with the τ -candidates reconstructed close to the central crack. This also improves the electron identification efficiency and thus reduces the electron contamination.

The last set of cuts is important to reduce the contamination of fake objects, like electrons and jets that mimic the τ signature.

Electrons are always identified as τ -candidates, since they also appear in the detector as narrow isolated clusters pointed to by a track. Although events containing an identified electron are not used in the analysis, there is still a residual contribution from $W \rightarrow e\nu$ events, where the electron fails the electron identification criteria (see 3.2). To remove these electrons, the CDF analyses typically require the fraction of electromagnetic energy of the cluster to be smaller as the fraction of energy from the neutral particle decreases. See [28] for more details.

The visible 4-momentum of the τ -candidate, P_{vis}^τ , is defined as the sum of the 4-momentum of the tracks and π^0 s in the signal cone (10° around the seed track). The visible momentum is the best estimate of the energy of the τ -candidate. Most of the time the τ decays to π^\pm , that can behave as a MIP in the EM compartment,

and deposits most of its energy in the hadronic compartment. The energy resolution for the tracks is better compared to one of the hadronic calorimeter, so using the track energy information improves the energy resolution. We require the $E_T^{Vis} > 25$ GeV to ensure that the visible energy and the cluster energy are consistent with each other.

In case the object reconstructed as a τ -candidate is a jet the visible mass, m_{vis} , would be larger than the one for real τ s, that have only small tails above 1.78 GeV/ c^2 . Hence, we require $m_{vis} \leq 1.8$ GeV/ c^2 . It should also be noted that the trigger applies a requirement to the τ mass. The track mass m_{trk} , calculated using only the tracks four-vectors, must satisfy $m_{trk} \leq 2$ GeV/ c^2 . This is always true when the previous cut on m_{vis} is true.

Since the τ is charged, it can only decay into odd numbers of charged particles: this motivates the cut on the number of tracks. Although there is a small fraction of real τ s reconstructed with two tracks (mostly due to an extra track from the conversion of one of the photons), adding them would not improve the analysis, since the signal over background ratio is worse than for the τ s with three tracks. We will exploit this fact in section 4.2.6, when we discuss the multijet background modeling.

3.2 Electrons

Electrons are reconstructed starting from a energy deposition in the central electromagnetic calorimeter, matched to a high p_T COT track and a CES cluster [29]. A set of identification criteria are then applied to the electron candidates, to improve their purity [30].

Electrons, if they have sufficient energy, are always identified as τ -candidates. If we were trying to achieve the best acceptance to the WH production, we would not be interested in reducing their contribution in the signal sample. In order to keep this analysis as minimally overlapping as possible to the $WH \rightarrow l\nu b\bar{b}$ analysis that has

been designed to be sensitive to $WH \rightarrow e\nu b\bar{b}$ (see [31]) , we veto events that contains identified electrons.

There are also other background processes that produce high p_T electrons, such as $Z \rightarrow ee$ and $Z \rightarrow \tau\tau \rightarrow \tau e \nu_\tau \nu_e$. To reduce their contribution we veto events that contain an identified τ -candidate and an electron candidate passing a loose identification criteria. See [32] and references therein.

3.3 Muons

Muons have a distinct signature of an high p_T track, that does not deposit much energy in the calorimeter, and leaves hits in the muon chambers located outside of the detector. Muons have a small probability of being identified as τ -candidates: they behave as MIP in the calorimeter, while the analysis requires the object identified as τ -candidate to deposit 20 GeV in the calorimeter. The main source of the muon contamination comes from $Z \rightarrow \tau\tau \rightarrow \tau\mu\nu_\tau\nu_\mu$. To reduce it, we veto events that, in addition to a τ -candidate, contains either a tightly identified μ or an isolated track with $p_T > 10$ GeV/ c . The standard muon identification requirements are described in [33], while the definition of an isolated track is described in [32].

3.4 Jet Reconstruction

When quarks and gluons are produced in the primary interaction, they fragment into colorless hadrons, that, if unstable, decay into stable particles. This process leaves in the detector the distinctive signature of a spray of particles called a jet (see Fig. 3.2).

At CDF, jets are reconstructed using the JETCLU algorithm [34], an iterative fixed cone algorithm, based on calorimeter information only, with cone size $R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$. The jet four-momentum is calculated as the sum of

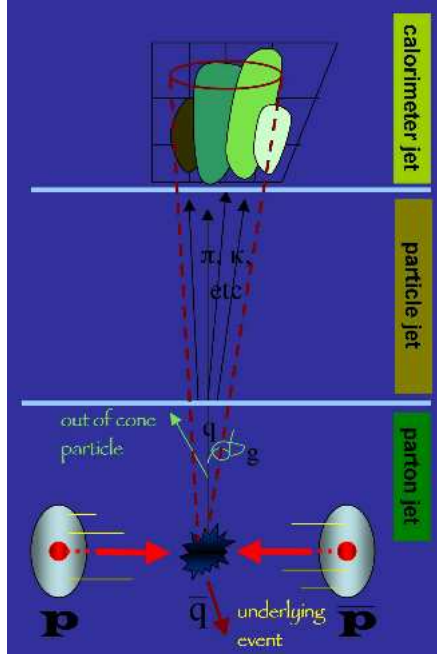


Figure 3.2: A parton from the primary interaction hadronizes and produce a spray of particles, that, when they reach the calorimeter, deposit their energy.

the massless four-vector momenta associated with the electromagnetic and hadronic tower compartments. The jet energy is then corrected for losses in the gaps, multiple interactions, and non-linearity of the calorimeter response [35] that have a bigger impact on the low energy particles often present in a jet. The jets used in this analysis satisfies $E_T^{corr} \geq 20 \text{ GeV}$, $|\eta| < 2$.

3.5 Neutrinos and Missing Transverse Energy

Neutrinos do not interact with the detector. The presence of energetic neutrinos in an event can be inferred from the significant momentum imbalance in a plane orthogonal to the beam direction. Calorimetric measurement of the transverse momentum imbalance, $\vec{\cancel{E}}_T$, is defined as a sum over the calorimeter towers

$$\vec{\cancel{E}}_T = \sum E_{T_i} \cdot \vec{n}_i,$$

where \vec{n}_i is a unit vector pointing from the event vertex to the i -th tower. E_T is further corrected to account for energy corrections applied to the jets with $E_T^{corr} > 12$ GeV and $|\eta| < 2.4$, except for the jet that is matched with the τ -candidate ($\Delta R < 0.4$). In principle the energy correction for a real τ and a jet could be different. As we didn't develop τ specific energy corrections, we don't correct the E_T for the τ -candidate energy.

3.6 b -tagging algorithm

Low mass Higgs bosons ($m_H \leq 135 \text{ GeV}/c^2$) decay mostly to $b\bar{b}$ pairs, so being able to distinguish jets originated by a b -quark from light flavor and gluon jets is a powerful tool to reduce the background to the search.

Two characteristics of b -hadrons that are exploited in the b -tagging algorithm at CDF are that they are long lived, with a mean lifetime of ~ 1.6 ps, and that they are massive ($m_{B^\pm} = 5.28 \text{ GeV}/c^2$). The first property allows them to travel a significant distance before decaying: for example, a b -quark with $p_T = 50 \text{ GeV}/c$, will travel on average almost half a centimeter. The mass affects the opening angle between the daughter particles: since it is large, the tracks associated with the charged decay products have a sizable impact parameters with respect to the interaction point. Both these aspects are used by the CDF secondary vertex b -tagging algorithm called SecVtx, [36], that attempts on a jet-by-jet basis to reconstruct a secondary vertex among the tracks with large impact parameter.

SecVtx uses a two-pass approach to find the secondary vertex. A set of cuts on transverse momentum, number of hits associated to the track and the χ^2/ndf of the final track fit are applied to reject poorly reconstructed tracks. Only if the jet contains at least two of these tracks can the secondary vertex be reconstructed: the jet is called "taggable". The algorithm starts considering tracks with $p_T \geq 0.5 \text{ GeV}/c$

and impact parameter significance $|d_0|/\sigma_{d_0} \geq 2.5$, and attempts to reconstruct a secondary vertex which includes at least three tracks, one of which has $p_T \geq 1 \text{ GeV}/c$. If the first pass is not successful, tighter cuts are applied to the tracks (such as $p_T \geq 1. \text{ GeV}/c$ and $|d_0|/\sigma_{d_0} \geq 3$) and the reconstruction of a secondary vertex is attempted considering only two tracks (of which one has $p_T \geq 1.5 \text{ GeV}/c$).

The algorithm uses the displacement of the secondary vertex L_{xy} to identify jets originating from b -quarks, where L_{xy} is the vector pointing from the primary to the secondary vertex projected onto the jet axis in the x-y plane. (see Figure 3.3)

The 2-dimensional displacement can be signed. If the secondary vertex is located

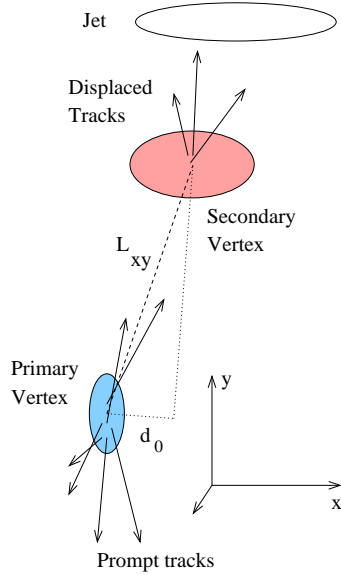


Figure 3.3: A diagram of the secondary vertex algorithm and the quantities involved.

on the side of the primary vertex where the jet is pointing, then L_{xy} is positive. In case the secondary vertex is on the opposite side compared to the direction of the jet, L_{xy} is negative. If the jet is originated by a b - or a c -quark, the 2-dimensional displacement is expected to be large and positive, while it can be negative only as an effect of the limited resolution of the silicon detector and mis-measured tracks. This

effect is exploited to calibrate the error rate of the tagger, see 3.6.2. To reduce the background from the false secondary vertices, a good secondary vertex is required to satisfy $L_{xy}/\sigma_{L_{xy}} \geq 3.$, where $\sigma_{L_{xy}}$ is the uncertainty on the displacement (on average $190 \mu m$), but calculated vertex-by-vertex. In case the secondary vertex associated with a jet is a good vertex (as stated above), the jet is called *b*-tagged.

The SecVtx algorithm and its ability to discriminate *b*-quark originated jets and those produced by a light quarks are very important in the search of a low mass Higgs. The two goals of the calibration of this algorithm are the measurement of the efficiency of *b*-tagging jets originated from heavy flavor jets, and the measurement of the rate at which false secondary vertices are reconstructed inside light flavor jets.

3.6.1 SecVtx *b*-tagging efficiency for heavy flavor jets

The signal and most of the backgrounds will be modeled using simulated events. Since Monte Carlo simulated data contains all the information necessary to the SecVtx algorithm to calculate L_{xy} , it is possible to determine if a jet is *b*-tagged in simulation. Problems arise, though, when comparing the efficiency of the algorithm applied to events with heavy flavor quarks in simulation and data: they are different. The ratio of the efficiency in data and in simulation is measured to account for this difference. This ratio is called scale factor (more details can be found in [36]). Figure 3.4 shows the efficiency of the SecVtx algorithm in tagging *b*-quark originated jets in top quark Monte Carlo sample. The band shows the systematic uncertainty on the scale factor, dominated by the uncertainty of *b*-tagging efficiency measured in data. The scale factor used in this analysis is:

$$SF^{b-tag} = 0.96 \pm 0.05. \quad (3.1)$$

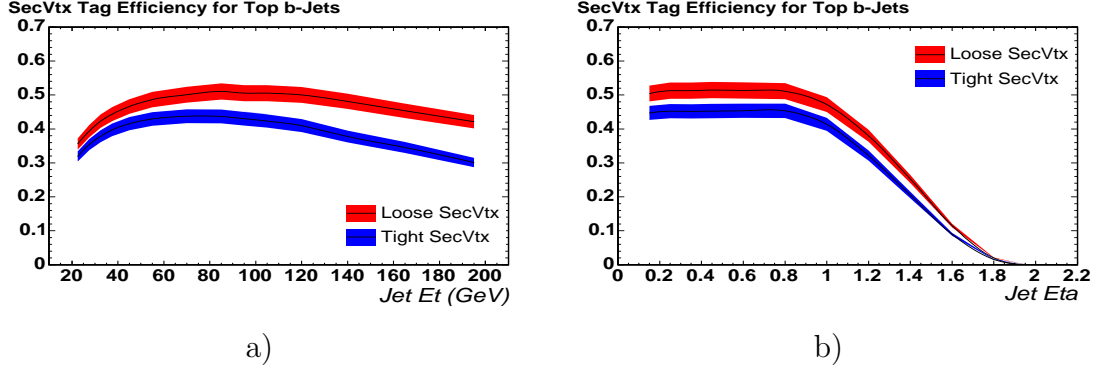


Figure 3.4: a) b -tagging efficiency for jets originated by heavy flavor quarks in simulated top quark pair events, as a function of E_T b) and as a function of η .

3.6.2 False positive b -tagging rate

It was observed that simulated events underestimate the probability of identifying a light jet as a heavy flavor jet ("mistag"). Because of this, we will discuss here the mistag matrix, that is used to calculate the probability that, for a given event, one or more jets are erroneously b -tagged. Figure 3.5 shows the difference between a jet with a negative L_{xy} , due to detector resolution and mis-reconstructed tracks, and a jet with a positive value for L_{xy} .

Detector effects are assumed to be independent of the sign of the L_{xy} , so, if they were the only source of mistag, the rate of the negative tags would be equal to the rate of the false positive tags. In reality, there is a small contribution from K 's and Λ 's and nuclear interactions with the detector material to the positive mistag rate, that is taken into account.

The probability of a jet to be falsely tagged has a dependence on jet kinematics. The mistag rate can be properly described if it is parametrized as a function of the following six variables: transverse energy of the jet (E_T); the number of tracks in the jet (N_{trk}); the sum of the transverse energy of all jets in the event ($\sum E_T^{jet}$); the pseudorapidity of the jet (η); the number of reconstructed vertices in the event (N_{vtx}); and the z -position of the primary vertex (z_{vtx}). Figure 3.6 shows the false

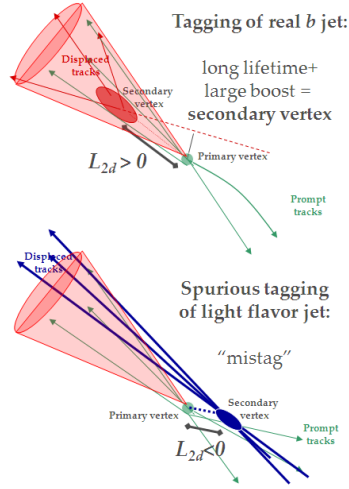


Figure 3.5: An schematic example of positive and negative L_{xy}

positive tag rate in inclusive jet data.

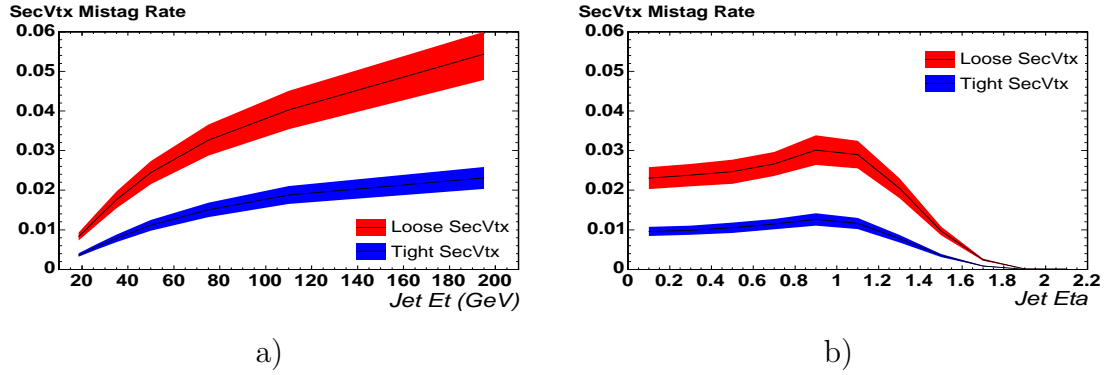


Figure 3.6: a) Mistag rate in jet inclusive data as a function of E_T , and b) as a function of η

Chapter 4

Signal and Background Processes

While searching for a new phenomenon it is crucial to have reliable methods to describe the backgrounds from known Standard Model processes, that can be compared with the observed data. It is necessary to be able to model the signal process as well. In high energy collider physics, the understanding of acceptances, efficiencies and backgrounds relies on detailed simulation of physics processes and detector response. Monte Carlo methods are used to simulate the complex processes that take place between the initial hard scattering and the final stable particles reconstructed in a event. Different Monte Carlo event generators are available to the experimentalists to generate simulated events. The most used ones at CDF are PYTHIA [37], ALPGEN [38], and MADEVENT [39]. Once generated, the τ leptons in the final state are decayed using TAUOLA [40], that correctly accounts for τ polarization. The CDFII detector response to particles is reproduced in Monte Carlo events using the GEANT [41] simulation.

In this chapter we will describe briefly the signal and background processes, and the Monte Carlo samples used to simulate them.

4.1 Signal

The signal process considered in this work is $WH \rightarrow \tau\nu b\bar{b}$. The Higgs boson, produced in association with a W boson, decays into a $b\bar{b}$ pair, while the W decays into a τ and a ν_τ . The τ subsequently decays into hadrons in the form $\tau \rightarrow X_h\nu_\tau$, where X_h is a system of hadrons, mainly composed by charge and neutral pions.

The leading order Feynman diagram is shown in figure 4.1.

Signal events are generated with PYTHIA [37], for $100 \leq m_H \leq 150 \text{ GeV}/c^2$ in

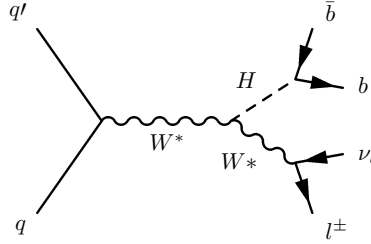


Figure 4.1: Feynman diagram for the signal process

$5 \text{ GeV}/c^2$ steps. The production cross section for $qq' \rightarrow WH$ is calculated at next-to-next-to-leading order [42], using the MSTW 2008 version of the parton distribution function (PDF) [44]. Table 4.1 lists the cross section as a function of the Higgs boson mass. The $b\bar{b}$ branching fractions, computed with HDECAY [43], are shown in table 4.2.

4.2 Background processes

The dominant background process to the $WH \rightarrow \tau\nu b\bar{b}$ signal is the QCD production of multijet events, where one of the jets mimics the signature of a τ , and the \cancel{E}_T arises from the jets energy mis-measurement. Among all the background processes it is the only one modeled using a data-based approach. The other background processes considered here are the production of a W boson in association with jets

m_H (GeV/ c^2)	σ_{WH} (fb)	scale (%)	PDF+ α_s^{exp} (%)	α_s^{th} (%)
100	291.90	+0.7 -0.8	+5.0 -5.0	+0.5 -0.4
105	248.40	+0.7 -0.8	+5.4 -5.4	+0.6 -0.4
110	212.00	+0.7 -0.9	+5.8 -5.8	+0.6 -0.5
115	174.50	+0.7 -0.9	+6.1 -6.1	+0.7 -0.5
120	150.10	+0.7 -0.9	+6.4 -6.3	+0.8 -0.6
125	129.50	+0.7 -1.0	+6.6 -6.7	+0.8 -0.6
130	112.00	+0.7 -1.0	+6.4 -6.7	+1.0 -0.7
135	97.20	+0.7 -1.0	+6.9 -6.8	+1.0 -0.7
140	84.60	+0.7 -1.1	+6.9 -6.7	+1.1 -0.8
145	73.70	+0.7 -1.1	+7.3 -7.1	+1.2 -0.9
150	64.40	+0.8 -1.1	+6.8 -6.7	+1.2 -0.9

Table 4.1: The NNLO production cross section for WH and the associated uncertainties.

("W+jets"), both in the case where the jets originate from a heavy flavor quark, such as b and c ("W+heavy flavor"), or when they are created by the fragmentation and hadronization of a light flavor quarks (u, d and s). Also top pair and single top productions are included, as well as production of Z +jets and gauge boson pairs (diboson production). All these processes are described using Monte Carlo simulation. Table 4.3 lists all the backgrounds, along with their theoretical production cross sections. No cross section value is quoted for W+jets and multijet QCD production, since the determination of their normalizations does not rely on it. More details can be found in the background estimate chapter 5.3

4.2.1 W+jets

The second most important background, after the overwhelming contribution of multijet QCD events is the W+jets production, which exhibits the same final state of the WH signal. All the leptonic decay modes of the W are considered, included the case where the identified τ -candidate is generated by an e or a μ .

m_H [GeV/ c^2]	$BR(H \rightarrow b\bar{b})$ (%)
90	81.2
95	80.4
100	79.1
105	77.3
110	74.5
115	70.5
120	64.9
125	57.8
130	49.4
135	40.4
140	31.4
145	23.1
150	15.7

Table 4.2: Higgs boson decay branching fractions.

ALPGEN, an event generator dedicated to multiparton hard process in hadronic collisions [38], is used to produce the W +jets samples. It performs exact LO matrix element calculations, that are completed by PYTHIA parton showering, that accounts for higher-order corrections that lead to the development of partonic cascades, and the formations of hadrons. After the shower evolution, the samples corresponding to different parton-level multiplicities can be combined together to obtain inclusive samples. For example, the production of a W associated with a $b\bar{b}$ pair ("W+ $b\bar{b}$ ") is described by the sum of the samples corresponding to the $W + b\bar{b}$ production without other partons present in the final state ($W + b\bar{b} + 0p$), $W + b\bar{b} + 1p$ and $W + b\bar{b} + 2p$. A matching scheme between partons and parton level jets, called MLM [45], is incorporated in ALPGEN, to avoid double counting: a configuration of a given ALPGEN process may, in fact, appears twice, once generated at the matrix level (for example, N+1 jets corresponding to N+1 partons), and another time generated by the parton showering (N jets corresponding to N partons, and an extra jet in the event due to a large angle emission during parton showering). Similarly, because heavy flavor can

Background Process	Cross Section [pb]
WW	11.3 ± 0.7
WZ	3.22 ± 0.23
ZZ	1.20 ± 0.06
Single top (t-chan)	1.05 ± 0.07
Single top (s-chan)	2.10 ± 0.19
Z+jets	253.1 ± 17.8
$t\bar{t}$	7.04 ± 0.41
W+jets	
Non-W	

Table 4.3: Background processes and their cross sections and errors.

be produced in the parton shower of W+light flavor events, and light flavor events can be produced in the evolution of $W + b\bar{b}$, $W + c\bar{c}$ or $W + c$, care is taken to remove this possible source of overlap [46].

To describe the inclusive W+jets sample, the different processes ($W + p$, $Wbb + np$, $Wcc + np$, $Wc + np$, where np is the number of extra parton generated at the matrix element level) are then added together according to their LO cross sections.

Some representative Feynman diagrams of the W+jets productions are shown in figure 4.2.

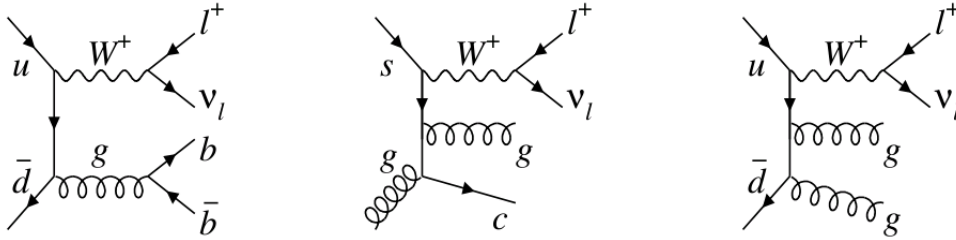


Figure 4.2: Some of the Feynman diagrams included in ALPGEN generation of W+jets events.

4.2.2 Top production

Top pair production is simulated using PYTHIA, assuming $m_{top} = 172.5 \text{ GeV}/c^2$. Its cross section is known to next-to-next-to-leading-order [47]. This process contributes to the background to the WH search when one of the W decays into a τ and a ν_τ , and the other one is lost due to detector acceptance. It can also contribute when one of the W decays into quarks, and one of the jets is mis-identified as a τ candidate, the other one is lost, and the decay products of the other W are lost as well. Figure 4.3 shows the Feynman diagram associated with $t\bar{t}$ production.

The contribution from this background increases when events with higher jets

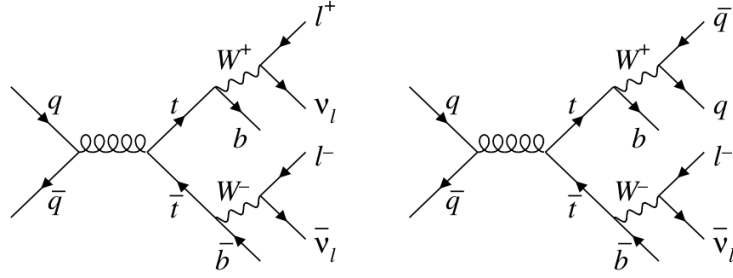


Figure 4.3: Feynman diagrams associated with $t\bar{t}$ production.

multiplicity are considered, as it will be evident in the background estimate table, tables 5.4 and 5.5.

4.2.3 Single top production

Single top production happens at hadron colliders via s - and t -channels. The leading order Feynman diagrams are shown in figure 4.4.

Although the cross section is small compared to most of the other backgrounds, its signature is very similar to the signal one, with the production of at least one b -quark, a W boson from the top-quark decay and one extra quark. The single-top events are simulated using MADEVENT, since this generator preserves information

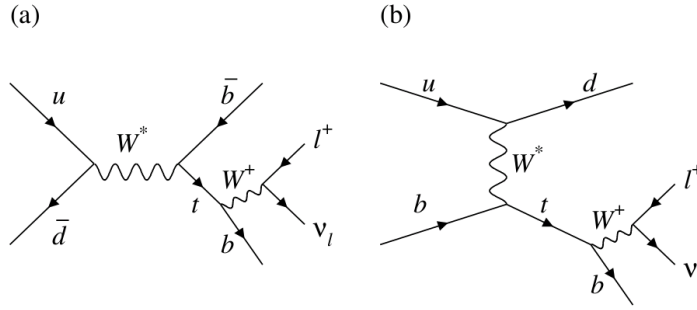


Figure 4.4: Feynman diagrams corresponding to single top production via s -channel (on the left) and t -channel (on the right).

from the polarization of the top quark, which PYTHIA does not include. PYTHIA is used instead to produce the parton showering in the events.

The cross section calculation for the s -channel process is performed at NNLL [48], while the t -channel one is evaluated at NLL [49]. Both the calculations use the MSTW2008 PDF set, and assume $m_{top} = 173 \pm 1.2 \text{ GeV}/c^2$.

4.2.4 $Z + \text{jets}$

The contribution of the Z +jets background in the final signal sample is very small. It is simulated using PYTHIA. Z +jets events contribute to the background of the WH only when Z/γ^* decays into $\tau - \tau$ pair, and one of the τ is lost due to detector acceptance. A Feynman diagram representative of the Z +jets production is shown in figure 4.5

The cross section used to estimate the Z +jets contribution in the signal sample is the one measured at CDF [50].

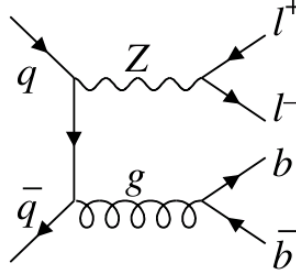


Figure 4.5: One of the Feynman diagrams for Z +jets productions

4.2.5 Diboson production

The samples describing diboson productions WW , WZ and ZZ are generated using PYTHIA. The WW and WZ processes contribute to the background when $W \rightarrow \tau\nu$, and the Z decays into quarks. In ZZ events, one of the τ must get lost, as in Z +jets. Feynman diagrams for these processes are shown in figure 4.6.

MCFM [51] is used to compute NLO cross section for WW , WZ and ZZ [52].

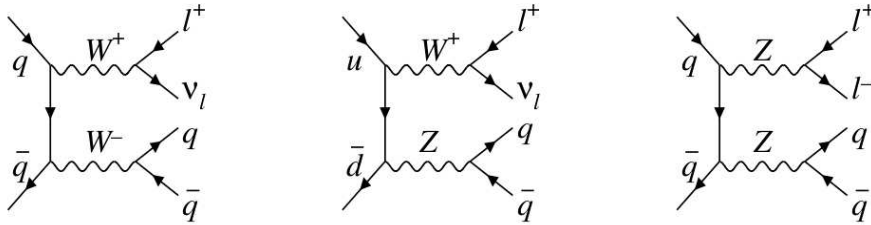


Figure 4.6: Feynman diagram associated with dibosons production.

4.2.6 QCD Multijet events

As it was said earlier, the dominant background to a $WH \rightarrow \tau\nu b\bar{b}$ search is the QCD production of multijet events. The lowest order Feynman diagram for multijet production is shown in figure 4.7.

In case one of the jets is identified as a τ candidate and the jet energy is mis-

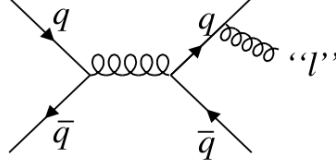


Figure 4.7: Lowest order multijet production Feynman diagram

measured enough for a non-negligible amount of \cancel{E}_T to arise in the event, the signature of a W boson is reproduced: this mechanism produces the multijet background, that for this reason is also referred to as “non- W ”.

Since the multijet QCD production cross section is many orders of magnitude higher than the W production cross section, the multijet background becomes the dominant one.

Multijet events have a small probability to satisfy the event selection, hence it is very difficult to rely on the Monte Carlo simulation to generate enough statistic to describe this process. Events from a multijet enriched data sample are used to model it. The rate is determined in data, as section 5.3.2 describes.

The multijet enriched data sample is collected with the same TAU-MET trigger used to acquire the events in the signal sample (see chapter 5.1), so that the effect of the online selection on the kinematics of non- W event is already described appropriately. It differs from the signal sample, though, by the number of tracks in the signal cone ($\alpha \leq 10^\circ$) of the τ -candidate: in the signal sample, τ -candidates are required to have exactly one or three tracks in signal cone (corresponding to the 1- and 3-prong decay of the τ), while in the multijet enriched sample τ -candidate have exactly two tracks in the cone.

Figure 4.8 shows the distribution for the number of tracks in the τ -candidate signal

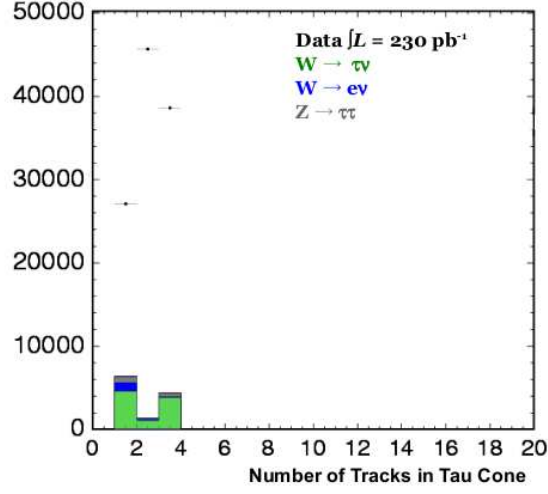


Figure 4.8: Tau candidate track multiplicity. The points represent the data, while the histograms show the expected contribution from W +jets and Z +jets event.

cone in a fraction of the data for an inclusive $W \rightarrow \tau\nu$ selection. The filled histograms represent the expected contributions of W +jets and Z +jets (modeled using PYTHIA, normalized to the cross sections measured at CDF [50]), while the points shows the observed data. It is reasonable to assume that there is no other significant contribution from Standard Model processes that was not included. Hence, the difference between the data and the background expectation is due to the multijet contribution. The first thing worth noting is that the inclusive $W \rightarrow \tau\nu$ +jets sample is dominated by the multijet production. The second one is that the contribution of electroweak processes to the events with 2 tracks in the τ signal cone very small. The validity and the limitations of the assumption that, in QCD multijet production, the track multiplicity of one of the jets does not affect the kinematics of the events, will be tested later.

Chapter 5

Signal and Background estimation

We present here how events consistent with $WH \rightarrow \tau\nu b\bar{b}$ signature are selected at CDF, both online by the triggering system, and offline, after the object reconstruction. The method used to determine the background composition of the data sample is then described. It is built on the method of background estimation detailed in [53], that was initially developed to predict the composition of a $l + \cancel{E}_T + \text{jets}$ ($l = e, \mu$) sample, dominated by W+jets. We adapted this method to the sample containing τ leptons, that presents a different sample composition.

5.1 Trigger Selection

The candidate events used for this search, were collected with the TAU_MET trigger. The TAU_MET trigger is designed to select $W \rightarrow \tau\nu$ decays. It requires the presence a narrow isolated cluster in the calorimeter, matched to a high p_T track and $\cancel{E}_T \geq 20\text{GeV}$. To limit the trigger rate as the Tevatron instantaneous luminosity was increasing, the trigger requirements have been evolving over time. One can count 11 different trigger versions, but only six of them show appreciable differences in the algorithm.

As an example, Table 5.1 shows the latest version of the trigger used to collect almost half of the analysed data. The most notable changes happened during the L2Cal up-

Name	Requirement
L1_TAU10_PT10	XFT_PT = 10.1 GeV XFT_LAYERS = 4 ET = 10 GeV
L2_TAU10_PT10_MET20	MISSING_ET = 20 GeV ABS_ETA_MAX = 1.1 ABS_ETA_MIN = 0 CLUS_NTOWERS = 5 ET = 10 GeV DELTA_PHI_6_MIN = 10 DELTA_PHI_6_MAX = 30 XFT_PT = 10.1 GeV ISO_FRACTION = .14
L3_TAU_MET	eta = 1 isolationTracks = 0 nTau = 1 pt = 15.0 trackMass = 2.0 MetCut = 20.0

Table 5.1: TAU_MET Trigger Requirements, version 16

grade (see [25]), when the Level 2 trigger started using full 10-bit calorimetric energy resolution. This allowed the development of a new clustering algorithm with better energy resolution and new isolation criteria. Up until then, the τ -candidate isolation requirement was imposed using a two-dimensional, track based veto, since no stereo information was available. Specifically, the algorithm does not consider a τ object isolated if there is a track with $p_T > 1$ GeV/ c , and azimuthal angle between 10 and 30 degrees from the τ -candidate direction. After the Level-2 calorimeter upgrade, the isolation requirement is imposed using the amount of energy deposited in the calorimeter towers surrounding the τ cluster. More details can be found in section 2.3.2.

The data were collected between February 2002 and February 2010, for a total integrated luminosity of 6.7 fb^{-1} . After requiring that all the necessary subdetectors were working correctly at the moment of data taking, the data analyzed correspond to a integrated luminosity of 5.7 fb^{-1} .

To compare the background processes estimated using Monte Carlo to the data, it is necessary to know the efficiency of the TAU-MET trigger, so that the acceptance measured in simulated events can be corrected for the events lost during data taking. In fact, because of the different resolution between online and offline variables, not all the Monte Carlo events where a τ -candidate is identified and the \cancel{E}_T is above 20 GeV satisfy the trigger requirements.

It is not possible, though, to measure the TAU-MET trigger efficiency in an alternative data sample: the TAU and MET part of the trigger are correlated, so the efficiency depends on the process.

As the CDF trigger is fully digital, the trigger information is included in the event simulation, and the online selection can be applied directly to Monte Carlo events. Possible differences between the selection efficiency in data and in Monte Carlo events are account for by what is called scale factor, defined as

$$S.F. = \frac{\epsilon^{Data}}{\epsilon^{M.C.}} \quad (5.1)$$

where ϵ^{Data} is the selection efficiency in data, and $\epsilon^{M.C.}$ in Monte Carlo events. In case of the online selection, it is reasonable to assume that, although the trigger efficiency depends on the process, the scale factor does not, and can be described by a common constant number. The trigger scale factor calculation will be discussed in 5.5.

The next section describes the trigger emulation in Monte Carlo events.

5.1.1 TAU_MET trigger emulation

As described in section 2.3, the trigger at CDF is a three-layer system: each of the 3 levels reduces the rate of the events that are fed to the next one and eventually saved to tape. The informations available in the two first layers, Level-1 and Level-2, are simulated using Monte Carlo events. I implemented the trigger logic in C++ code that was used to emulate the online trigger decision on simulated events. The algorithm was tested with data, by comparing the online trigger decision with the emulated one.

The Level-3 trigger takes advantage of the full event reconstruction, so it is not simulated by TrigSim. The Level-3 trigger is emulated applying the online algorithm to the offline reconstructed quantities, but this introduces a difference between the Level-3 efficiency in data and the emulation in simulated events. For example, at the trigger level, the transverse component of the energy deposited in the towers is calculated assuming that the primary interaction happens at the center of the detector, while offline its z -position is determined using the tracks present in the event. This, and the differences in online and offline calibrations will affect the \cancel{E}_T calculations, for example. These effects will be included in the estimate of an overall trigger scale factor (see section 5.5).

The emulation of each trigger version is applied to a fraction of Monte Carlo events that correspond to the fraction of integrated luminosity the trigger version was live. Simulated events used in the analysis are required to satisfy the emulated trigger logic.

5.2 Offline Event Selection

The data events used in the analysis are required to contain exactly one identified τ candidate with $E_T^{Vis} \geq 25$ GeV. Since we consider events where $H \rightarrow b\bar{b}$, the purity of the sample is improved by requiring the presence of at least one jet consistent with

TAU_MET trigger	PASS
τ -candidate	$E_T > 25 \text{ GeV}$
$ Z_0^{vtx} $	$\leq 60 \text{ cm}$
Identified Electron	VETO
Identified Muon, $\Delta R(\mu, \tau) \geq 0.4$	VETO
Loose Muon, $\Delta R(\mu, \tau) \geq 0.4$	VETO
Isolated Track, $\Delta R(\mu, \tau) \geq 0.4$	VETO
Number of tight jets	$=2$
$\Delta R(jet, \tau)$	≥ 0.4

Table 5.2: Event Level cuts

the originating from a b -quark. Jets, clustered with a cone of radius $\Delta R = 0.4$, are required to verify $E_T \geq 20 \text{ GeV}$ and $|\eta| \leq 0.2$. To optimize the analysis sensitivity, the data are split into two statistically independent subsamples: one containing events with exactly one b -tagged jet (Single Tag sample), and another one containing events with exactly two b -tagged jets (Double Tag sample). See section 3.6.1 for details about the definition of b -tagging, and the SecVtx algorithm.

The quality cuts listed in Table 5.2 are applied during the event selection process. A good quality reconstructed primary vertex is required to be in the region consistent with the beam-beam interaction, with a position along the beam direction that verifies $|Z_0^{vtx}| \leq 60 \text{ cm}$. Events that contain an identified electron are vetoed, to reduce the contribution from the $W \rightarrow e\nu + \text{jets}$ background. The contribution from $W \rightarrow \mu\nu + \text{jets}$ where the muon is identified as a τ is substantially smaller, since the τ is required to deposit more than 20 GeV in the calorimeter in order to be identified. Events that contain an extra μ , a loose e or an isolated track are vetoed as well, to suppress the Drell-Yan background. Details of the lepton identification algorithms are described in chapter 3.

To avoid double counting, each jet and the identified τ -candidate must be separated in $\eta - \phi$ space by $\Delta R > 0.4$. The number of jets presents in an event is used to determine independent samples: the “2-jets” sample is used to search for the Higgs

signal, while the other ones “1-jet”, “3-jets”, “4-jets” sample) are used as control regions. The samples with a τ candidate, before any b -tagging requirement is applied, is called Pretag, and is used to measure the identification scale factor, the number of W +jets events in the sample, and as a control region.

5.3 Sample Composition and Background Estimate

Chapter 4 lists all the process that are thought to contribute to the $\tau + \cancel{E}_T + jets$ sample. The Z +jets, diboson, $t\bar{t}$ pairs and single top production rates can be predicted using the Monte Carlo simulation. Corrections accounting for differences between data and the simulation are applied to the simulated events. They correct for object identification efficiencies, trigger efficiencies and primary $p\bar{p}$ interaction position (primary vertex), and are derived from control data samples. After applying all corrections, the yields for the backgrounds are calculated as the product of the acceptance (from the simulation) times the luminosity and measured or theoretical cross sections, as in the following equation:

$$N_{p\bar{p} \rightarrow X} = \epsilon \times (\sigma_{p\bar{p} \rightarrow X} \times \int L dt) \quad (5.2)$$

The main background to the search is multijet production. It is modeled by data, and its rate, N^{QCD} , is estimated fitting the \cancel{E}_T^{perp} spectrum. \cancel{E}_T^{perp} is the component of the \cancel{E}_T that is perpendicular to the closest object (either the τ -candidate or a jet), and is very effective in separating events containing a real W decay from ones where the W signature is due to object mis-reconstruction.

The inclusive W +jets cross section is well known [50], but ALPGEN is a Leading Order (LO) event generator, and does not describe correctly the W acceptance as a function of the number of jets presents in the event. Because of this, the normalization of the W +jets background in the Pretag sample is determined by the same \cancel{E}_T^{perp}

fit to data used to determine the size of the non- W component. It is performed separately for events with exactly one, two, three or four jets. If the relative contribution of the W produced in association with heavy and light flavor is known, it is possible to predict the number of W +jets events that satisfy the b -tagging requirements. This procedure determines the background composition of the Pretag and b -tagged samples. Each step will be described in more details in the next sections.

5.3.1 Monte Carlo Based Background Estimate

The first step is to determine the contribution of processes such as WW, WZ, ZZ, Z +jets and top production, relying on the known values of their cross sections. The cross section values are listed in table 4.3. The number of events of a given process that are expected in the sample is calculated using the formula:

$$N_{p\bar{p} \rightarrow X} = \epsilon \times (\sigma_{p\bar{p} \rightarrow X} \times \int L dt). \quad (5.3)$$

where $\sigma_{p\bar{p} \rightarrow X}$ is the production cross section for a given process, $\int L dt$ the integrated luminosity of the data sample considered, and ϵ the selection efficiency. The selection efficiency, ϵ , is determined in Monte Carlo, and corresponds to the fraction of simulated events that are reconstructed in the detector and satisfy the event selection. Corrections accounting for differences between data and the simulation are applied to the simulated events. When calculating the event yield in the b -tagged sample, ϵ needs to be corrected as well for the b -tagging efficiency. Since Monte Carlo simulation is not able to describe the tagging correctly, the b -tagging scale factor (see 3.6.1) and mistag matrix (see 3.6.2) are incorporated in the event selection efficiency. Instead of counting the number of Monte Carlo events where jets are b -tagged by the SecVtx algorithm, events are weighted by the probability that the jets are b -tagged. For example, in case we were considering events with at least one jet to be b -tagged, the probability will be defined as:

$$P_{event}^{b-tag} = 1 - \prod_i^{jets} (1 - p_{tag}^i). \quad (5.4)$$

For jets originating from a heavy flavor quark (b or c) p_{tag}^i is the b -tagging scale factor if the jet is tagged, and zero if it is not tagged. If the jet is originated by a light flavor, p_{tag}^i is the mistag probability. The selection efficiency is then calculated as:

$$\epsilon = \epsilon^{event} \cdot \epsilon^{tag} = \epsilon^{event} \cdot \frac{\sum_i^{events} P_i^{b-tag}}{N_{pretag}} \quad (5.5)$$

The calculation for the tagging efficiency for Double Tag events is performed with the same methodology only the combinatorics is slightly more complicated. The uncertainty in the normalizations from the tag scale factor and the mistag matrix are calculated by fluctuating these values by ± 1 within the nominal values and then recalculating the normalizations.

5.3.2 Non- W background

non- W Normalization in the Pretag Sample

As it was mentioned earlier, this approach to the background estimate was developed at CDF to determine the sample composition of a $l + \cancel{E}_T + jets$ sample, where $l = e, \mu$ [53]. The contributions of multijet background is not negligible in such a sample, but it is not the dominant one. In this approach, called Method II, the size of the multijet contribution in the Pretag sample is determined performing a binned likelihood fit to the \cancel{E}_T spectrum in data to the sum of \cancel{E}_T background shapes. The fit has one fixed component and two templates whose normalizations can float. The fixed component is coming from the MC-based processes, whose normalization are constrained to the values calculated according to the equation 5.3 within the uncertainties. The two floating templates are a Monte Carlo $W + jets$ template and a non- W template. The non- W template comes from a multijet enriched data sample.

When trying to apply the same approach to this analysis, we realized that, in multijet events, there is a correlation between the value of the \cancel{E}_T and the track multiplicity of the τ -candidate, especially when the τ -candidate points toward the central or

the ϕ cracks. The seed track of the τ candidate is forced to be fiducial to the CES by the identification algorithm, but not the other tracks in the signal cone. Depending on the track multiplicity of the τ -candidate, also the fraction of energy associated with the seed track varies: on average, it will be greater for a 1-prong τ -candidate, and smaller for a 3-prong one. So, in case the τ -candidate is pointing toward a crack, the fraction of energy that can be lost in the crack is bigger for 3-prong than for 1-prong one. Because the trigger requires $\cancel{E}_T > 20$ GeV there is a bias towards this event topology. The non- W model is not able to describe the behavior of the multijets events in the signal sample in the \cancel{E}_T distribution, given the different track multiplicity. This is seen when we try to fit the non- W and W +jets template to data in \cancel{E}_T distribution, in the Pretag sample with 2 jets (see fig. 5.1). However, the other variables, especially the jet-related ones, are not affected by this correlation, and the non- W model describes them well. In this analysis, to

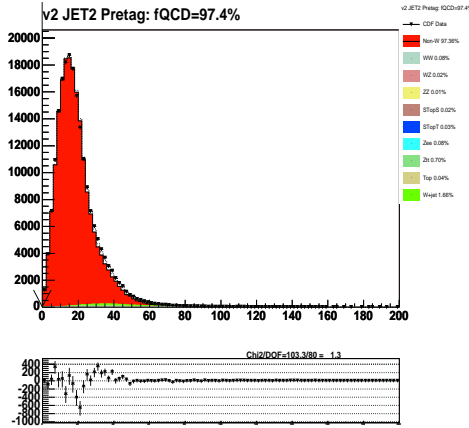


Figure 5.1: Comparison between the data \cancel{E}_T distribution and expected background contribution, after the fit. On the bottom, we show the difference in number of events between the sum of the backgrounds and the data. Residuals shows discrepancy in the peak region.

determine the normalization of the multijet sample, we fit the component of the \cancel{E}_T perpendicular to the closest object (either one the jets, or the τ -candidate), \cancel{E}_T^{perp} .

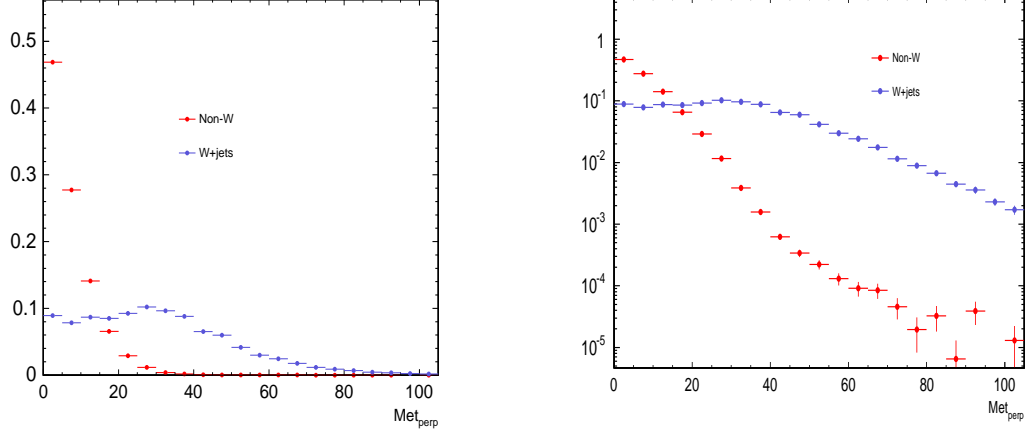


Figure 5.2: E_T^{perp} distribution shape comparison for Non-W events and W plus jets. The two histograms are normalized to unit area.

Figure 5.2 shows the distribution in E_T^{perp} for multijet events and for W+jets. As the non-W distribution falls quickly, the tail becomes sensitive to the small contribution of W+jets, making the results of the fit meaningful. An example of the fit result is in fig 5.3.

We fit to the E_T^{perp} distribution in data the templates from the different backgrounds. The rate of the multi-jet and the W+jets backgrounds are two free parameters of the fit, while the rate of the other backgrounds are constrained to their predicted yields within the uncertainty. The minimization of the likelihood fit determines the best estimate for the normalization of the multijet background, as well as for the number of W+jets events in the sample.

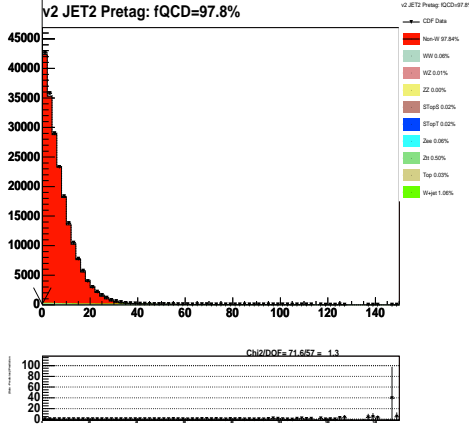


Figure 5.3: Comparison between data H_T^{perp} distribution and expected background contribution after the fit

Multi-jet Background Normalization in the Single Tag Sample

The rate of the non- W background in the Single Tag sample is determined via a two template fit to the data. The H_T^{perp} template for the multi-jet background is determined using events from the same multijet enriched sample described in section 4.2.6, but considering only events with exactly one b -tagged jet. The other template is given by the sum of all other background templates, normalized to the expected yields in the Single Tag sample. The normalization of the multijet template is allowed to float in the fit, while the other one is constrained, within the uncertainty, to the expected value.

The uncertainty associated with the multijet normalization is the one returned by the fit.

Multi-jet Background Normalization in the Double Tag Sample

The non- W background normalization in the Double Tag sample is determined via a two template fit to the data, similar to the one for the Single Tag sample. One

template is given by the sum of the \mathcal{H}_T^{perp} distributions for all the backgrounds (except the multijet one), normalized to the expected yield in the Double Tag sample. To describe the \mathcal{H}_T^{perp} distribution for the multi-jet background we consider all the events in the multijet enriched sample that contain either two b -tagged jets or one b -tagged and one taggable jet. A taggable jet is defined by the presence of at least two good tracks in its cone, necessary to reconstruct a secondary vertex.

If only the non- W events with both the jets b -tagged were used, the statistic of the template would be very low, and it would be very difficult to perform the fit. Since selecting events with a b -tagged jet enhances the probability for the other one to be a heavy flavor jet, events with only one tight tag jet could be added to the template. As the tracking efficiency and, consequently, the b -tagging efficiency, depends on the jet η , the second jet taggability requirement results in a better background modeling.

5.3.3 W +jets

As it was mentioned earlier, the normalization of the W +jet background in the Pretag sample is determined by the two component fit to the \mathcal{H}_T^{perp} spectrum in data. To predict its contribution to the tagged sample, W +jets is broken down into events where W is produced in association with heavy flavor quarks (W +h.f.) and in association with light flavor quarks (W +l.f., or mistag).

W +heavy flavor

The fraction of W +jets events that contains heavy flavor quarks, $f_{H.F.}$ is estimated from the ALPGEN + PYTHIA Monte Carlo, but a correction factor is needed. ALPGEN, in fact, is a LO event generator, while the production cross section for Wcc +jets and Wbb +jets processes increases when NLO and higher order calculations are performed. The K-factor $K = 1.4 \pm 0.4$, is measured at CDF in the W +1jet bin and applied to the rest of the sample, and account for the difference in heavy flavor fractions observed between data and Monte Carlo prediction. The number of

W +h.f events in the b -tagged sample can be calculated as:

$$N_{W+HF}^{tag} = N_{W+jets}^{pretga} \cdot f_{HF} \cdot K \cdot \epsilon_{tag} \quad (5.6)$$

where ϵ_{tag} is the tagging probability, described in section 5.3.1.

W+ light flavor

Once we know the W +jets heavy flavor content, the remainder of the W +jets sample is assumed to be the W +light flavor jet events. This process contributes to the tagged samples only when one or more jets originated from light flavor quarks are mistakenly b -tagged. The mistag rate is estimated by applying the mistag probability to the W +light flavor sample: multiplying each event by its mistagging probability allows an estimate of the number that makes it into the final sample. The uncertainty on this estimate comes from the mistag matrix, which has an uncertainty in each bin because of the finite statistics of the sample it was derived from.

5.4 Signal Expectation

As explained in section 4.1, the signal samples were generated using PYTHIA, for Higgs mass $100 \leq m_H \leq 150$ GeV/ c^2 in 5 GeV/ c^2 steps. The number of expected signal events is calculated in a similar way to the Monte Carlo based background, via:

$$N_{WH} = \epsilon \times (\sigma_{p\bar{p} \rightarrow WH} \cdot BR(H \rightarrow b\bar{b}) \cdot \int L dt). \quad (5.7)$$

where the value for $\sigma_{p\bar{p} \rightarrow WH}$ and $BR(H \rightarrow b\bar{b})$ are listed in table 4.1 and 4.2.

The fraction of Higgs events containing an hadronically decaying τ that survive the Pretag event selection is 4.3 %. These events contain an identified τ , no other extra lepton in the event, satisfy the trigger and have exactly two jets. Most of the inefficiency is due to the energy and isolation cuts. The thresholds for these cuts are set so to be at least as tight as the corresponding cuts applied online, at the trigger level.

5.5 Scale Factor Measurement

The selection efficiency used to determine the yield for Monte Carlo based backgrounds (see equations 5.3) and the Higgs boson signal (equation 5.7) needs to be corrected for the difference in τ lepton identification and TAU_MET trigger efficiencies between data and simulations. There is no need to measure these two scale factors separately, as they are always multiplied by each other, so this section describes how we determine their product.

In case of e and μ , the identification and trigger efficiency scale factors can be measured using the $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events in the data and Monte Carlo with the tag and probe method. This is not possible in case of τ . A sample consistent with a $Z \rightarrow \tau\tau \rightarrow \tau l \nu\nu$ decays contains a non-negligible contribution from the other background processes, and it does not have enough statistics. The τ lepton decays into e or μ only $\sim 35\%$ of the time, and these leptons have a softer energy spectrum, resulting in a lower efficiency for high p_T lepton trigger. Furthermore, the TAU_MET trigger requires both, a τ -candidate and \cancel{E}_T at the same time. The latter requirement is less likely to be satisfied by Z events, where the \cancel{E}_T is due to the neutrinos from the decay of the τ .

A better way to measure the identification and trigger scale factors is with $W \rightarrow \tau\nu$ events. If we were measuring the W cross section, we would rely on:

$$\sigma = \frac{N^{Data} - N^{background}}{\epsilon \times SF \times \int L dt} \quad (5.8)$$

where $N^{Data} - N^{background}$ represents the number of events from W +jets as measured in data, ϵ is the Monte Carlo acceptance, SF is the product of the scale factors that are needed to correct the raw Monte Carlo acceptance to represent the data, and $\int L dt$ is the integrated luminosity corresponding to the data used. Since the W production cross section at the Tevatron is known, if we could determine the number of W events present in a sample we could resolve the equation 5.8 for SF : the product of the trigger and identification scale factors can be estimated. In principle

this could be done for the inclusive W (no cut on number of jets), or exclusively in any jet bin. Since there could be a correlation between the trigger efficiency and the number of jets in the event (for example, due to the two-dimensional isolation trigger requirement), the best sample to perform this measurement is the $W \rightarrow \tau\nu$ sample with exactly two jets. This is also the sample that we use to search for the Higgs, but its contribution in the sample, before applying b-tagging, is negligible. To avoid the difficulty in determining a value for the exclusive $W+2$ jets cross section, the scale factor is determined normalizing the yield of $W \rightarrow \tau\nu + 2$ jets events collected with TAU_MET trigger to the yield of the $W \rightarrow e(\mu)\nu + 2$ jets events collected with the inclusive high-pt electron and muon triggers. The product of the identification and trigger scale factors is 0.72 ± 0.12 . For reference, the inclusive W boson production cross section derived from the $W \rightarrow e(\mu)\nu + 2$ jets samples is 2609 ± 399 pb ([54]).

5.6 Event Yield

We have described how to estimate the contributions of the individual backgrounds in the data sample under consideration, as well as the expected number of signal events. In this section we show the \mathcal{H}_T^{perp} fit, used to determine the non- W and W +jets normalizations in the Pretag (fig. 5.4), the Single Tag (fig. 5.5) and the Double Tag (fig 5.6) samples. The summary of the background composition can be found in Tables 5.3, 5.4 and 5.5. Figure 5.4 shows results of the \mathcal{H}_T^{perp} fit in the Pretag sample. The histograms represent the backgrounds normalized to their expected yields. All the histograms are added together, except the non- W one. The sum of all the backgrounds is represented by the dashed histogram. The points correspond to the data. Figures 5.5 and 5.6 show the result of the \mathcal{H}_T^{perp} fits in the Single and Double Tag samples correspondingly. In these samples, the fits are performed using only two templates: the non- W one, and the sum of all other backgrounds normalized

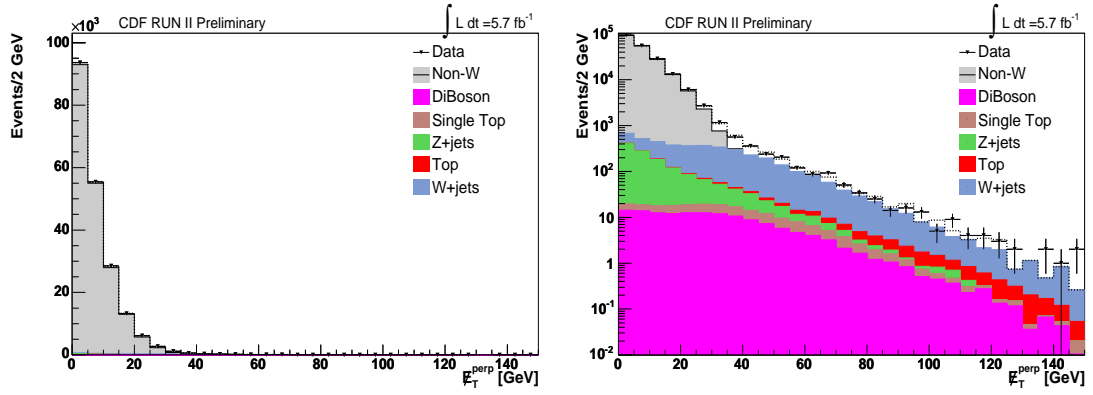


Figure 5.4: Fit of the E_T^{perp} distribution in the Pretag sample

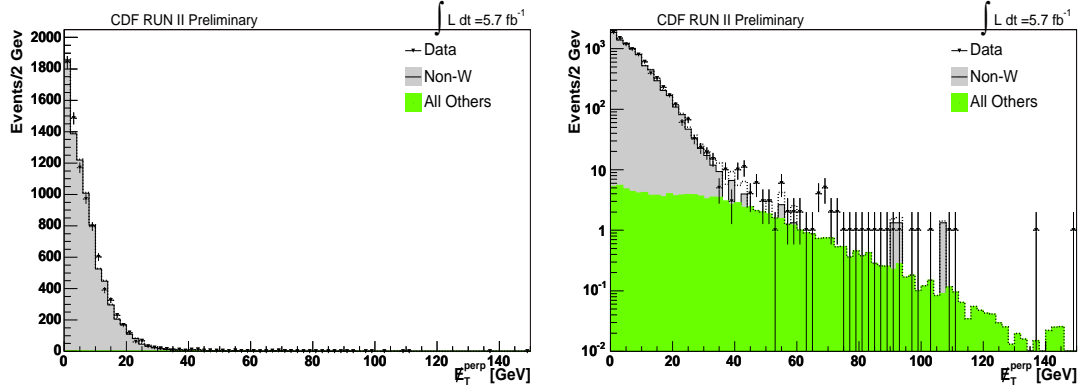


Figure 5.5: Fit of the E_T^{perp} distribution in the Single Tag sample

according to their predicted yields.

Figures 5.7, and 5.8 compare the expected and observed distributions in different kinematic variables for events with $N_{jets} = 2$.

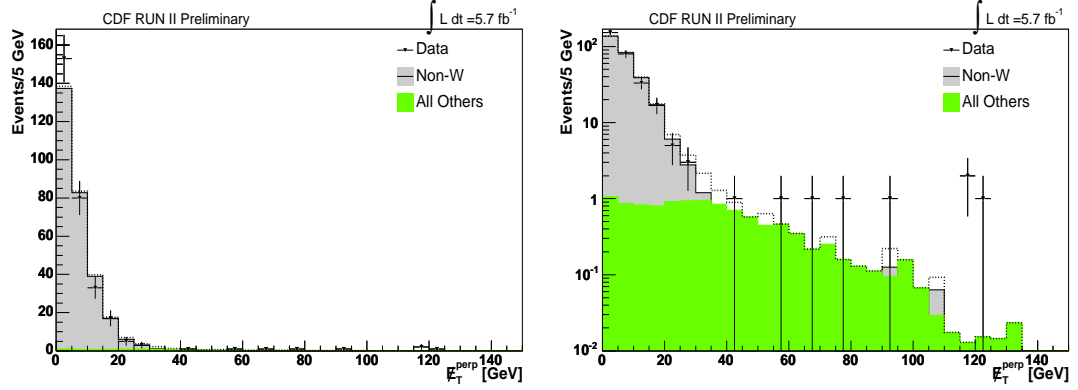


Figure 5.6: Fit of the H_T^{perp} distribution in the Double Tag sample

NJets	1jet	2jets	3jets	4jets	5jets
Pretag Events	1046855	202674	34298	5322	869
Non-W	1027529.3 ± 1166.0	198305.8 ± 480.5	33448.1 ± 190.1	5098.2 ± 74.7	792.1 ± 30.9
WW	102.8 ± 19.1	115.2 ± 21.5	31.3 ± 5.9	5.8 ± 1.1	1.2 ± 0.3
WZ	20.3 ± 3.8	22.9 ± 4.4	8.2 ± 1.6	1.7 ± 0.3	0.4 ± 0.1
ZZ	3.1 ± 0.6	2.6 ± 0.5	1.5 ± 0.3	0.4 ± 0.1	0.1 ± 0.0
Single Top (s-chan)	13.6 ± 2.6	33.6 ± 6.3	10.5 ± 2.0	2.4 ± 0.5	0.5 ± 0.1
Single Top (t-chan)	32.1 ± 6.3	42.1 ± 8.3	13.2 ± 2.6	2.6 ± 0.5	0.4 ± 0.1
$Z \rightarrow ee$	313.4 ± 60.4	118.3 ± 23.5	32.4 ± 7.2	3.6 ± 1.4	0.9 ± 0.7
$Z \rightarrow \tau\tau$	4025.8 ± 760.9	1022.1 ± 193.4	179.2 ± 34.2	20.1 ± 4.1	3.2 ± 0.8
$t\bar{t}$	10.1 ± 1.9	55.6 ± 10.3	111.2 ± 20.6	106.2 ± 19.6	44.2 ± 8.2
Wjj	14682.4 ± 315.6	2955.1 ± 121.7	459.4 ± 56.6	72.2 ± 32.9	22.0 ± 16.3
$Wb\bar{b}$	168.4 ± 50.7	117.8 ± 35.7	30.2 ± 9.8	7.5 ± 4.1	2.6 ± 2.1
$Wc\bar{c}$	357.5 ± 107.5	221.0 ± 66.9	60.8 ± 19.7	12.9 ± 7.1	4.9 ± 3.9
Wc	570.7 ± 171.7	165.1 ± 50.0	13.5 ± 4.4	0.7 ± 0.4	0.1 ± 0.1
W+l.f.	13585.7 ± 378.8	2451.2 ± 152.8	355.0 ± 61.6	51.2 ± 34.6	14.4 ± 17.4

Table 5.3: Background composition of the Pretag sample. Diboson, Z +jets and top production yield are predicted based on their cross section, while the W +jets and non- W contribution are determined by the fit to data.

NJets	1jet	2jets	3jets	4jets	5jets
Pretag Events	1046855	202674	34298	5322	869
WW	2.0 ± 0.4	5.2 ± 1.0	2.0 ± 0.4	0.6 ± 0.1	0.1 ± 0.0
WZ	0.9 ± 0.2	1.8 ± 0.4	0.8 ± 0.2	0.2 ± 0.0	0.0 ± 0.0
ZZ	0.1 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Single Top (s-chan)	5.1 ± 1.0	15.2 ± 2.9	4.7 ± 0.9	1.1 ± 0.2	0.2 ± 0.0
Single Top (t-chan)	11.4 ± 2.3	17.0 ± 3.4	5.6 ± 1.1	1.2 ± 0.2	0.2 ± 0.0
$Z \rightarrow ee$	1.7 ± 0.9	2.3 ± 1.1	0.8 ± 0.6	0.6 ± 0.5	0.0 ± 0.1
$Z \rightarrow \tau\tau$	32.0 ± 6.3	25.7 ± 5.2	8.9 ± 2.0	1.0 ± 0.4	0.3 ± 0.2
$t\bar{t}$	3.4 ± 0.6	22.6 ± 4.2	47.8 ± 8.9	46.8 ± 8.7	19.3 ± 3.6
Mistags	101.3 ± 3.7	34.7 ± 2.3	8.7 ± 1.5	1.9 ± 1.3	0.8 ± 0.9
$Wb\bar{b}$	51.9 ± 15.6	45.6 ± 13.8	11.5 ± 3.7	3.0 ± 1.7	1.0 ± 0.8
$Wc\bar{c}$	25.0 ± 7.6	21.2 ± 6.5	7.1 ± 2.3	1.9 ± 1.0	0.7 ± 0.6
Wc	39.8 ± 12.1	15.8 ± 4.8	1.6 ± 0.5	0.1 ± 0.1	0.0 ± 0.0
Non-W	25322.4 ± 159.5	8272.9 ± 91.5	1998.8 ± 45.6	448.4 ± 22.5	72.5 ± 9.8
Total Background	25596.9 ± 163.8	8480.4 ± 96.3	2098.2 ± 47.8	506.7 ± 24.6	95.3 ± 10.6
WH ($m_H = 115\text{GeV}$)	0.24 ± 0.04	0.72 ± 0.13	0.19 ± 0.04	0.03 ± 0.01	0.01 ± 0.00
Observed	25460	8412	2084	506	101

Table 5.4: Background composition of the Single Tag sample. The number of events for all the processes, except the non- W , are predicted based on the Pretag normalizations and the b -tagging efficiency. The non- W normalization is determined by the fit to data.

nJets	2jets	3jets	4jets	5jets
Pretag Events	202674	34298	5322	869
WW	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
WZ	0.3 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
ZZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Single Top (s-channel)	4.2 ± 0.8	1.5 ± 0.3	0.3 ± 0.1	0.1 ± 0.0
Single Top (t-channel)	1.4 ± 0.3	1.3 ± 0.3	0.3 ± 0.1	0.1 ± 0.0
$Z \rightarrow ee$	0.0 ± 0.1	0.0 ± 0.1	0.0 ± 0.1	0.0 ± 0.0
$Z \rightarrow \tau\tau$	0.4 ± 0.2	1.0 ± 0.4	0.2 ± 0.2	0.0 ± 0.0
$t\bar{t}$	4.6 ± 0.9	13.3 ± 2.5	17.2 ± 3.2	8.2 ± 1.5
Mistags	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$Wb\bar{b}$	7.1 ± 2.2	2.0 ± 0.7	0.5 ± 0.3	0.2 ± 0.2
$Wc\bar{c}$	0.4 ± 0.1	0.2 ± 0.1	0.1 ± 0.0	0.1 ± 0.0
Wc	0.3 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Non-W	286.9 ± 17.1	108.7 ± 10.9	24.9 ± 6.2	4.9 ± 3.3
Total Background	305.7 ± 17.4	128.4 ± 11.4	43.7 ± 7.1	13.6 ± 3.7
WH ($m_H = 115\text{ GeV}$)	0.23 ± 0.04	0.09 ± 0.02	0.02 ± 0.00	0.00 ± 0.00
Observed	299	128	43	16

Table 5.5: Background summary table for Double Tag events. The number of events for all the processes, except the non- W , are predicted based on the Pretag normalizations and the b -tagging efficiency. The non- W normalization is determined by the fit to data.

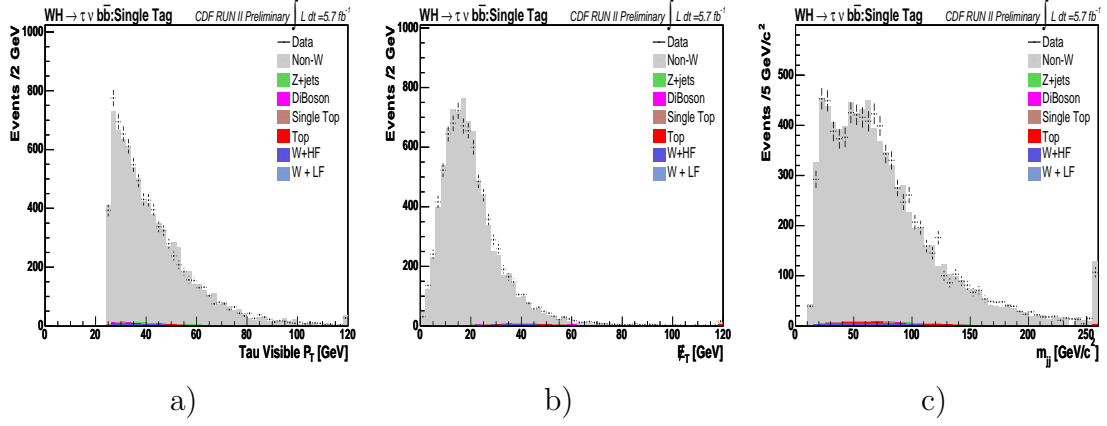


Figure 5.7: Single Tag sample. Comparison of observed and expected distribution of a) visible p_T of the τ -candidate b) E_T c) dijet invariant mass. Validation plots show agreement.

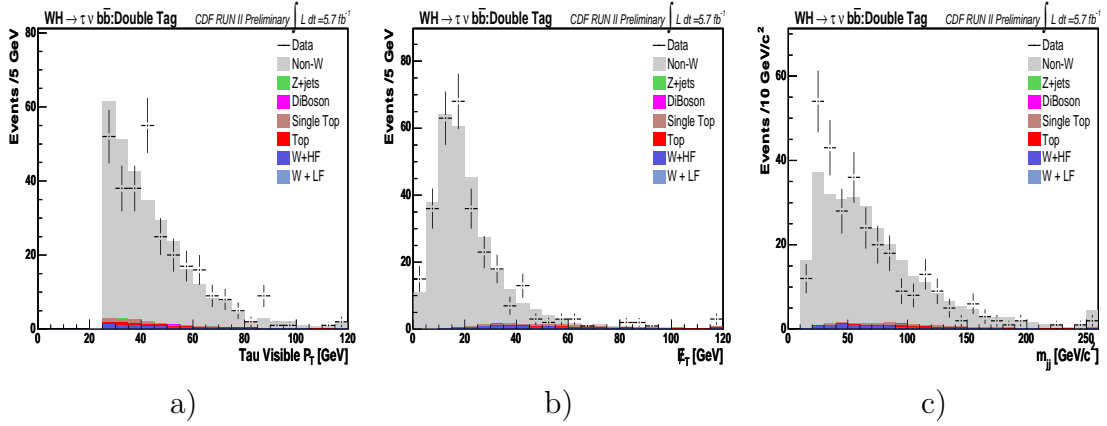


Figure 5.8: Double Tag sample. Comparison of observed and expected distribution of a) visible p_T of the τ -candidate b) E_T c) dijet invariant mass. Validation plots show agreement.

5.7 Event Selection Optimization

As shown in section 5.6, the Single Tag and Double Tag samples are dominated by the non- W background, and the signal is negligible. It is clear that this work does not have sensitivity to the production cross section of the Standard Model Higgs boson, and only a 95% upper limit can be extracted from the data.

It is necessary to improve the signal-to-background ratio to set the most stringent limit, and we optimize the cut value on \cancel{E}_T^{perp} to achieve the best sensitivity. The figure of merit used in the optimization procedure is the expected 95% C.L. upper limit on the Higgs production cross section.

The calculation of the Bayesian C.L. upper limit for a Higgs boson with mass $m_H = 115 \text{ GeV}/c^2$ is based on the binned likelihood fit to the invariant mass of the two high p_T jets present in the events. Systematic uncertainties on the rate of signal and background production are included in the limit calculation. A posterior density is obtained by multiplying the likelihood by Gaussian prior densities for the background normalizations and systematic uncertainties (see section 5.8 for details on the systematics uncertainties) leaving $\sigma \cdot \text{BR}(H \rightarrow b\bar{b})$ with a uniform prior density. A 95% C.L. limit is then determined such that 95% of the posterior density for $\sigma \cdot \text{BR}(H \rightarrow b\bar{b})$ falls below the limit. We use the MCLIMIT package [55] to determine the 95% limit: more details on the limit setting procedure are described in section 6.1.

The value of the \cancel{E}_T^{perp} cut is varied in steps of 2 GeV, between 18 and 40 GeV for the Single Tag sample, and between 18 and 28 GeV for the Double Tag sample. For each background process, the number of expected events is corrected for the efficiency of the \cancel{E}_T^{perp} cut. Because of the limited statistics of the non- W sample, the efficiency of the \cancel{E}_T^{perp} cut introduces a non-negligible background normalization uncertainty, included in the summary tables 5.8 and 5.9 and in the limit calculation.

The median expected 95% limits on the ratio of the Higgs boson rate and its expectation from the Standard Model for the two channels (Single and Double Tag

channel) are shown in tables 5.6 and 5.7 and as a function of the \cancel{E}_T^{perp} . The same information is shown as well in figure 5.9. In the Single Tag sample, the best limit corresponds to $\cancel{E}_T^{perp} > 34$ GeV. For the Double Tag sample, the differences in the limit between requiring $\cancel{E}_T^{perp} > 20$ GeV and $\cancel{E}_T^{perp} > 23$ GeV, that corresponds to the minimum, is not significant (the median expected limits with the number of pseudo-experiment considered here is stable at the 5% level). For simplicity we choose $\cancel{E}_T^{perp} > 20$ GeV to define the signal region and that will be used to set the final limit.

The summary of the background composition after including the \cancel{E}_T^{perp} cut is shown in table 5.8 for the Single Tag sample and in table 5.9 for the Double Tag sample. Although the number of expected events seems to overestimate the number of observed events in both the Single and Double Tag samples, this does not seem to be a systematic effect, but rather a statistical fluctuation. The expected and observed distributions in various kinematic variables for events with $N_{jets} = 2$ are shown in Figures 5.10, 5.11.

\cancel{E}_T^{perp} cut	Median expected limit
18	140.163
20	96.6202
22	87.616
24	90.5929
26	71.2261
28	71.6811
30	67.8448
32	60.6075
34	70.0519
36	74.5548
38	69.944
40	72.1288

Table 5.6: Median expected limit for the Single Tag sample, as a function of the \cancel{E}_T^{perp} cut.

\cancel{E}_T^{perp} cut	Median expected limit
18	69.2
20	67.8
22	67.6
24	67.1
26	66.6
28	69.4

Table 5.7: Median expected limit for the Double Tag sample, as a function of the \cancel{E}_T^{perp} cut.

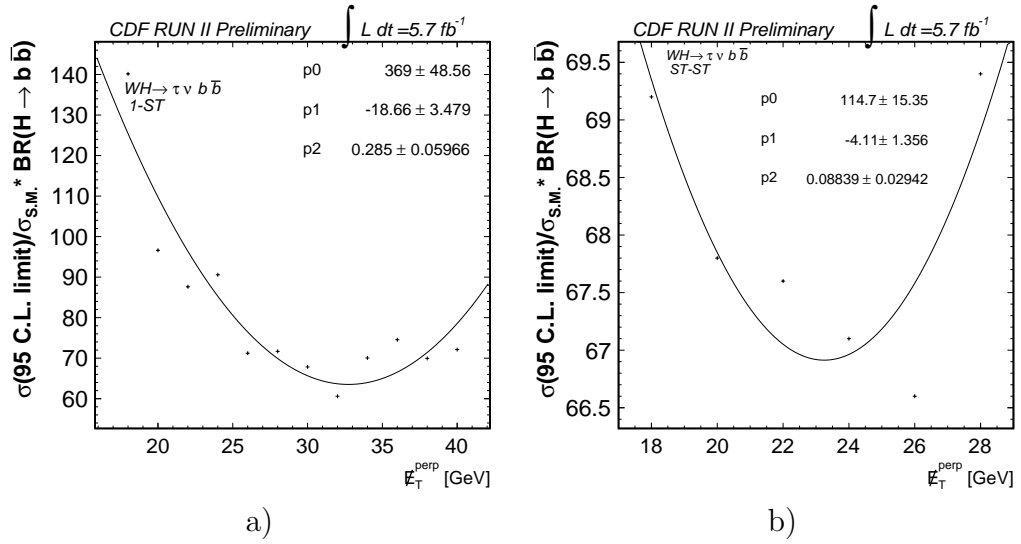


Figure 5.9: Median expected limit for 1) Single Tag and b) Double Tag sample, as a function of the \cancel{E}_T^{perp} cut.

NJets	1jet	2jets	3jets	4jets	5jets
WW	1.01 ± 0.20	1.96 ± 0.37	0.64 ± 0.13	0.14 ± 0.03	0.02 ± 0.01
WZ	0.37 ± 0.07	0.69 ± 0.13	0.18 ± 0.04	0.02 ± 0.01	0.01 ± 0.00
ZZ	0.04 ± 0.01	0.04 ± 0.01	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Single top (t-chan)	2.75 ± 0.52	6.71 ± 1.26	1.84 ± 0.35	0.35 ± 0.07	0.07 ± 0.02
Single top (s-chan)	6.05 ± 1.20	7.93 ± 1.57	2.15 ± 0.43	0.43 ± 0.09	0.03 ± 0.01
$Z \rightarrow ee$	0.10 ± 0.06	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
$Z \rightarrow \tau\tau$	5.92 ± 1.17	2.64 ± 0.53	0.57 ± 0.13	0.01 ± 0.01	0.00 ± 0.00
$t\bar{t}$	2.29 ± 0.43	12.70 ± 2.35	20.06 ± 3.71	14.67 ± 2.72	3.94 ± 0.73
Mistag	38.34 ± 1.41	13.67 ± 0.91	3.08 ± 0.54	0.67 ± 0.45	0.22 ± 0.27
$Wb\bar{b}$	17.81 ± 5.37	17.07 ± 5.17	3.95 ± 1.28	0.94 ± 0.51	0.27 ± 0.21
$Wc\bar{c}$	8.53 ± 2.60	8.24 ± 2.52	2.48 ± 0.81	0.56 ± 0.31	0.13 ± 0.10
Wc	10.37 ± 3.16	4.32 ± 1.32	0.38 ± 0.12	0.02 ± 0.01	0.00 ± 0.00
Non-W	50.68 ± 8.11	34.55 ± 6.77	12.70 ± 4.01	5.66 ± 2.83	2.27 ± 1.61
Total Background	144.27 ± 10.81	110.51 ± 9.57	48.04 ± 5.73	23.48 ± 3.99	6.96 ± 1.80
WH (115 GeV)	0.11 ± 0.02	0.30 ± 0.05	0.08 ± 0.01	0.01 ± 0.00	0.00 ± 0.00
Observed	155	100	41	18	3

Table 5.8: Single Tag sample backgrounds prediction, data, and expected signal for events with $\cancel{E}_T^{perp} > 34$ GeV.

NJets	2jets	3jets	4jets	5jets
WW	0.03 ± 0.02	0.04 ± 0.02	0.04 ± 0.02	0.00 ± 0.00
WZ	0.21 ± 0.05	0.04 ± 0.01	0.02 ± 0.01	0.00 ± 0.00
ZZ	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Single top (t-chan)	2.84 ± 0.54	1.00 ± 0.19	0.18 ± 0.04	0.05 ± 0.01
Single top (s-chan)	0.88 ± 0.18	0.76 ± 0.15	0.18 ± 0.04	0.02 ± 0.01
$Z \rightarrow ee$	0.00 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
$Z\tau\tau$	0.24 ± 0.14	0.23 ± 0.09	0.00 ± 0.00	0.00 ± 0.00
$t\bar{t}$	3.57 ± 0.67	8.45 ± 1.57	9.07 ± 1.68	2.84 ± 0.53
Mistag	0.11 ± 0.01	0.05 ± 0.01	0.02 ± 0.01	0.01 ± 0.01
$Wb\bar{b}$	4.42 ± 1.36	1.25 ± 0.41	0.27 ± 0.17	0.10 ± 0.08
$Wc\bar{c}$	0.24 ± 0.07	0.10 ± 0.03	0.03 ± 0.02	0.03 ± 0.03
Wc	0.20 ± 0.06	0.02 ± 0.01	0.00 ± 0.00	0.00 ± 0.00
Non-W	11.11 ± 1.21	4.21 ± 0.68	1.52 ± 0.50	0.20 ± 0.18
Total Background	23.88 ± 2.03	16.14 ± 1.78	11.33 ± 1.76	3.25 ± 1.0
WH (115 GeV)	0.14 ± 0.03	0.04 ± 0.01	0.01 ± 0.00	0.00 ± 0.00
Observed	16	17	11	5

Table 5.9: Double Tag sample backgrounds prediction, data, and expected signal for events with $\cancel{E}_T^{perp} > 20$ GeV.

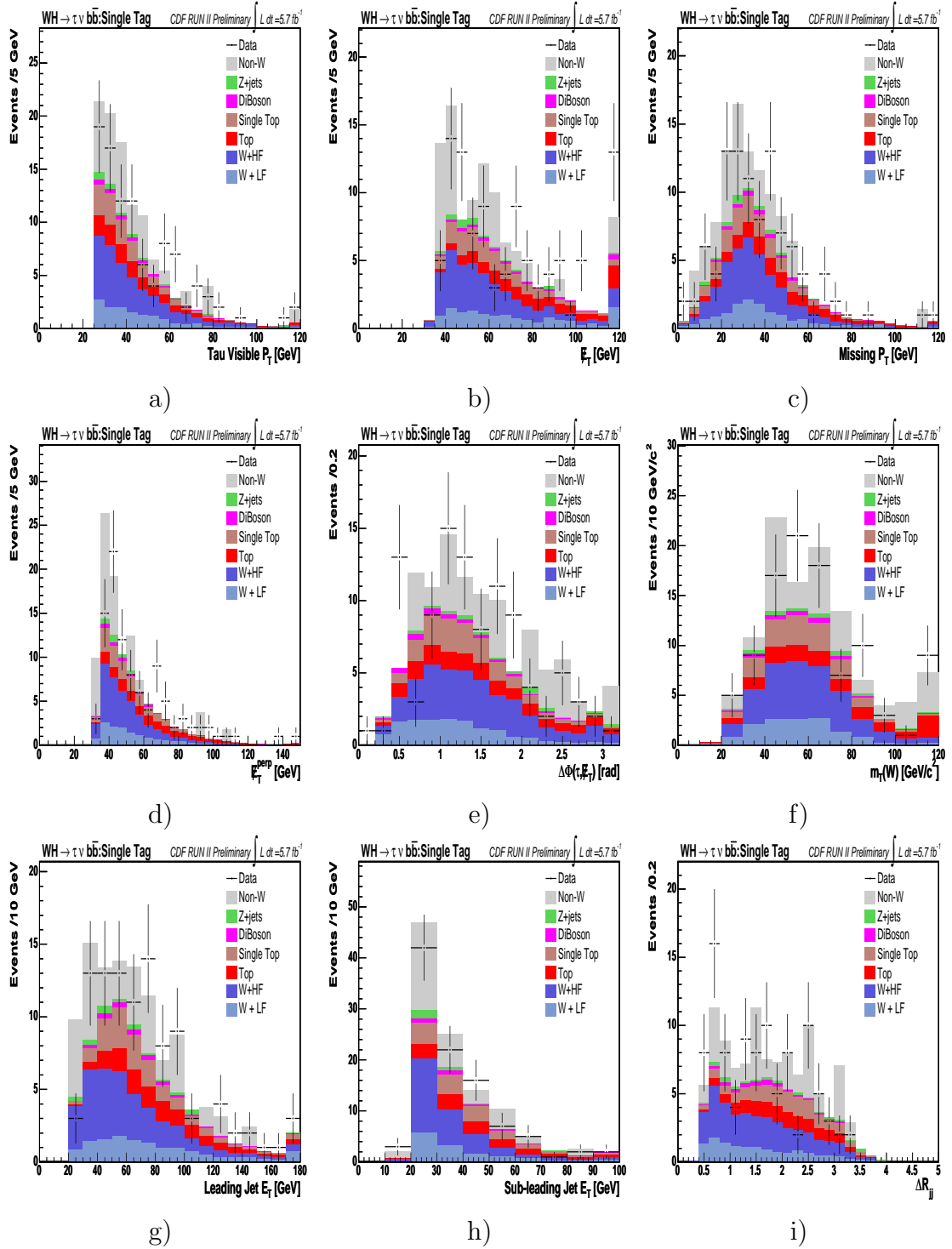


Figure 5.10: Single Tag sample, $E_T^{perp} > 34$ GeV. Comparison of observed and expected distribution of a) visible p_T of the τ -candidate, b) E_T , c) missing transverse momentum, d) $\Delta\phi(\tau - E_T^{perp})$, e) W transverse mass, f) Leading jet E_T , g) Sub-leading jet E_T , h) $\Delta R(j_1 j_2)$, i) $\Delta R(j_1 j_2)$. Validation plots show agreement within statistical uncertainties.

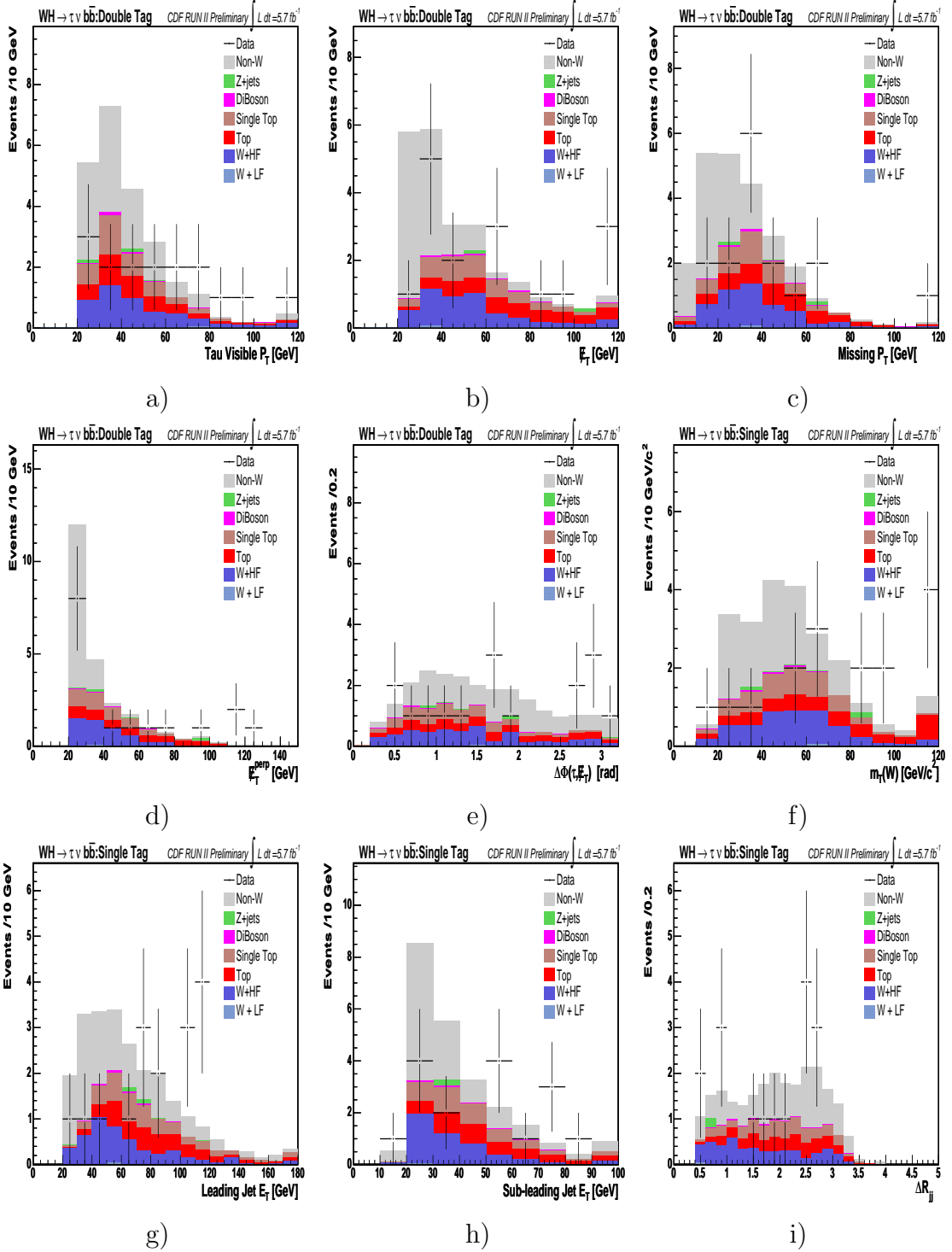


Figure 5.11: Double Tag sample, $E_T^{perp} > 20$ GeV. Comparison of observed and expected distribution of a) visible p_T of the τ -candidate, b) E_T , c) missing transverse momentum, d) $\Delta\phi(\tau - E_T^{perp})$, f) W transverse mass, g) Leading jet E_T , h) Sub-leading jet E_T , i) $\Delta R(j_1 j_2)$. Validation plots show agreement within statistical uncertainties.

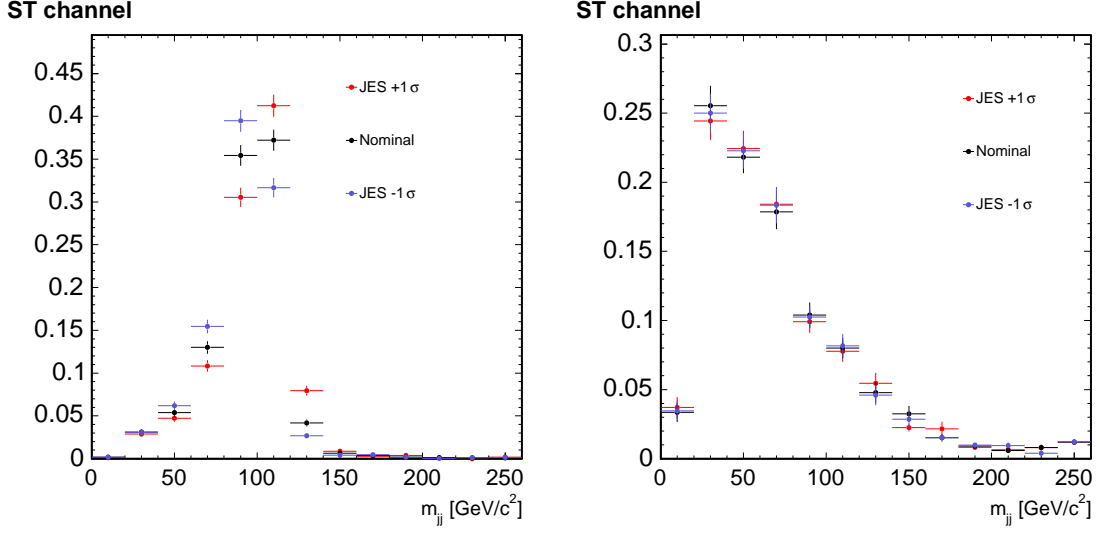


Figure 5.12: $\pm\sigma$ JES variation effect on di-jet mass distribution shape in events with one b-tagged jet. On the left for the signal ($m_H = 115\text{GeV}/c^2$) and on the right for $W \rightarrow b\bar{b}$ background.

5.8 Systematic Uncertainties

Two kinds of systematic uncertainties are considered in this analysis: uncertainties on the overall normalization of the background estimates, and uncertainties affecting the distribution shapes. The effect of the Jet Energy Scale (JES) uncertainty on the different processes is evaluated applying the jet energy correction corresponding to $\pm 1\sigma$ variations from the nominal correction. Since variations of JES change the energy spectrum of the jets and thus the shape of the dijet mass spectrum, it is considered as a shape systematic. Figures 5.12 and 5.13 show the effect on the dijet mass distribution of varying the JES by $\pm 1\sigma$ for the signal, and the $W \rightarrow b\bar{b}$ process. Since the varied and nominal background templates are consistent within statistical uncertainty for all the background processes, we consider the JES shape variation only for the signal.

We consider the following sources of the systematic uncertainties, affecting the

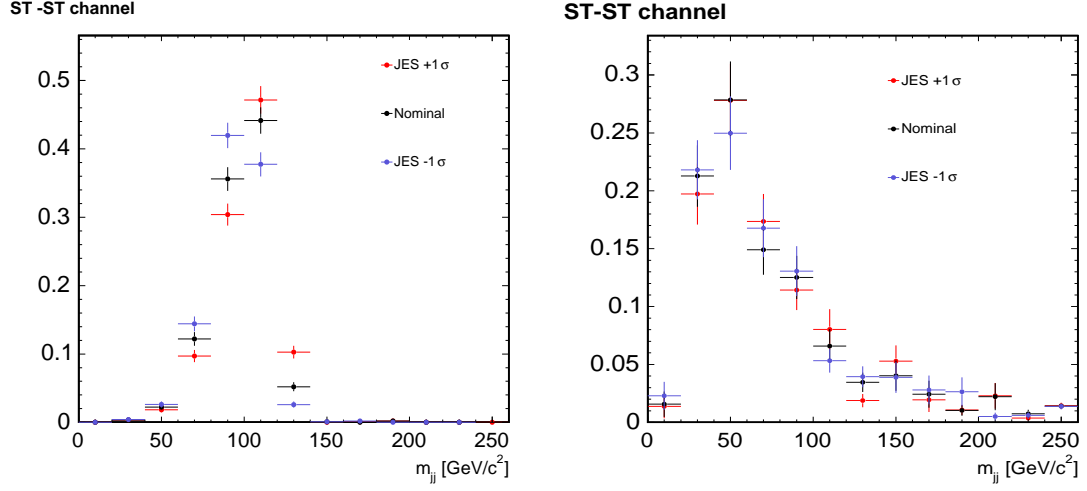


Figure 5.13: $\pm\sigma$ JES variation effect on dijet mass distribution shape in events with two b-tagged jets. On the left for the signal ($m_H = 115 \text{ GeV}/c^2$) and on the right for $W \rightarrow b\bar{b}$ background.

overall signal and background normalizations:

- **Jet Energy Scale:** applying the shift of $\pm 1\sigma$ to the jet energy scale could cause event migration between samples characterized with different numbers of jets, affecting the normalization of the Monte Carlo based backgrounds.
- **Cross section:** for each physics process modeled by Monte Carlo (except the W+jets one), the uncertainty on the cross section based either on a previous CDF measure or theoretical calculation is considered.
- **Tau Identification and trigger efficiencies:** the uncertainty on the scale factor measured in section 5.5
- **Luminosity:** we assign 6% uncertainty on the luminosity measurement
- **Z vertex cut:** the different efficiency in data and Monte Carlo simulation of this cut introduces a scale factor, known with an accuracy of 0.5%

- **W+HF normalization:** the fraction of heavy flavor events in the ALPGEN Monte Carlo is corrected to account for NLO effects, and the correction has an uncertainty of 30%. The b-tagging scale factor has an uncertainty of 4.3%, that, depending on the number of the tag in the sample, has a different contribution (the nominal 4.3% for Single Tag, and 8.6% for a Double Tag, since the uncertainties are 100% correlated between jets)
- **W+LF normalization:** the W +light flavor normalization is given by the difference between the amount of W +jets and W +HF in the Pretag sample, so the uncertainty on heavy flavor K-factor affects its normalization. The uncertainty on the mistag rate is calculated re-performing the calculation of the rate with the mistag matrix values for each jet fluctuated within $\pm\sigma$ of the uncertainty.
- **NonW normalization:** the fit returns the uncertainty on the parameter associated with the non- W normalization. The acceptance on \cancel{E}_T^{perp} cut has also a statistical uncertainty that affects the non- W normalization, and is accounted for in the uncertainty.
- **Initial- and Final-State Radiation:** uncertainty due to the change in strength of Initial State Radiation (ISR) and Final State Radiation (FSR) in parton showering. Dedicated Monte Carlo samples for WH signal events are used to estimate its effect.
- **PDF** uncertainty due to the choice of parton distribution function. The signal events are re-weighted for different PDF schemes associated with variation of the twenty independent eigenvectors of the CTEQ [56] PDFs. The acceptance varies between +0.6 and -0.4% .

Tables 5.10 and 5.11 summarize the impact of the considered systematic uncertainties.

CDF Run II Preliminary, $WH \rightarrow \tau\nu b\bar{b}$, Single Tag Sample, 5.7fb^{-1}							
Systematic Uncertainty (%)	Signal	Non-W	W+HF	W+LF	Z+jets	DiBoson	$t\bar{t}$ and Single t
b-tagging	4.3	-	4.3	-	4.3	4.3	4.3
Tau Trigger and ID	17.	-	17.	17.	17.	17.	17.
JES	2. - shape						
Heavy Flavor Fraction	30.						
Mistag Probability	7.						
Non-W normalization	20.						
ISR-FSR-PDF	2.						

Table 5.10: List of Systematic uncertainties affecting the signal and backgrounds in the Single Tag sample.

CDF Run II Preliminary, $WH \rightarrow \tau\nu b\bar{b}$, Double Tag Sample, 5.7fb^{-1}							
Systematic Uncertainty (%)	Signal	Non-W	W+HF	W+LF	Z+jets	DiBoson	$t\bar{t}$ and Single t
b-tagging	8.6	-	8.6	-	8.6	8.6	8.6
Tau Trigger and ID	17.	-	17.	17.	17.	17.	17.
JES	2. - shape						
Heavy Flavor Fraction	30.						
Mistag Probability	10.						
Non-W normalization	11.						
ISR-FSR-PDF	2.						

Table 5.11: List of Systematic uncertainties affecting the signal and backgrounds in the Double Tag sample.

Chapter 6

Results

Since no evidence for a Higgs boson signal excess is observed in the dijet invariant mass distribution, we set a 95% confidence level limit on its production cross section. We use MCLIMIT package [55] for the limit calculation, but we outline the main principle of the Bayesian limit calculation in this chapter, along with the search results.

6.1 Limit Calculation

The problem at hand can be formulated as how to extract a limit on the size of signal s as deduced from observing n events from a Poisson distribution with mean $s\epsilon + b$, in presence on uncertainty on the acceptance to the number of signal events ϵ and the backgrounds b . Given that the size of the expected Higgs signal is very small, a counting experiment would not be sensitive to it. To improve the signal-to-background ratio and the expected sensitivity of the search, the events in the final sample are classified in bins of a variable able to maximally distinguish signal from backgrounds: some bins will show a better signal-to-background ratio, and will be more sensitive to the presence of a signal. Since each bin is considered an independent experiment, the probability of observing the data, given the expected

background can be expressed as the product of the probability in each bin. So, the probability of obtaining the observed result is:

$$\prod_{k=1}^N \frac{e^{-(s\epsilon_k + b_k)} (s\epsilon_k + b_k)^{n_k}}{n_k!} \quad (6.1)$$

where N is number of bins, n_k the number of observed events in the k -th bin, and $k = 1, 2, \dots, N$. The number of signal events expected to be produced is represented by s , ϵ_k is the acceptance (that is, the fraction of signal events that survives the event selection) for the k -th bin, and b_k is the expected number of background events for the k -th bin. In a realistic case, both ϵ_k and b_k have uncertainties associated with them. Following the Bayesian approach, all the knowledge about the uncertainties is included into a probability distribution function called the joint nuisance prior. The joint nuisance prior can be written as:

$$\pi(\epsilon_1, b_1, \epsilon_2, b_2, \dots, \epsilon_N, b_N) d\epsilon_1 db_1 d\epsilon_2 db_2 \cdots d\epsilon_N db_N . \quad (6.2)$$

According to the Bayesian theorem, to determine the posterior density function for s , given n_k observed, the probability of equation 6.1 must be multiplied by the prior $\pi(s)$, that summarizes the knowledge about the signal size. Considering as well all the “nuisance parameter”, that is, all the uncertainties, the posterior is proportional to:

$$\pi(s) \int \int \int \pi(\epsilon_1, b_1, \epsilon_2, b_2, \dots, \epsilon_N, b_N) \left[\prod_{k=1}^N \frac{e^{-(s\epsilon_k + b_k)} (s\epsilon_k + b_k)^{n_k}}{n_k!} \right] d\epsilon_1 db_1 d\epsilon_2 db_2 \cdots d\epsilon_N db_N \quad (6.3)$$

where we have $2N$ integrals (two for each of the N bins, one for the uncertainty on signal yield and the other for the uncertainty on background yield) to be performed averaging over the information about uncertainties. This averaging over the information about uncertainties is called marginalizing, and this p.d.f. is called marginalized posterior.

The 2N integrals can be turned into finite sum over an ensemble of M random configuration for the nuisance parameters. The configuration can be generated from the known joint nuisance priors. The marginalized posterior for s can be written as

$$p(s|n) = \frac{1}{\mathcal{N} \cdot M} \sum_{i=1}^M \left[\prod_{k=1}^N \frac{e^{-(s\epsilon_{ki}+b_{ki})} (s\epsilon_{ki} + b_{ki})^{n_k}}{n_k!} \right] \quad (6.4)$$

where the normalization constant \mathcal{N} is given by

$$\mathcal{N} = \int_0^\infty \frac{1}{M} \sum_{i=1}^M \left[\prod_{k=1}^N \frac{e^{-(s\epsilon_{ki}+b_{ki})} (s\epsilon_{ki} + b_{ki})^{n_k}}{n_k!} \right] ds \quad (6.5)$$

The 95% confidence level (C.L.) upper limit on the signal s_u is defined with $\int_0^{s_u} p(s|n) ds = 0.95$.

6.2 Dijet Invariant Mass

The single most discriminating variable in the search for $H \rightarrow b\bar{b}$ is the invariant mass of the two jets in the event, as it represents the mass of the Higgs boson. A Higgs boson signal would, in fact, appear in the invariant mass distribution as a narrow peak over a falling background. We will use this variable to calculate the 95% C.L. upper limit on the Higgs production rate.

In the Single Tag sample, the number of events in the template used to model the multijet background is limited (26 events), so it is necessary to account for the bin-by-bin statistical uncertainty in the limit calculation. The effect of this uncertainty could be reduced increasing the number of events in the multijet template. For this, we studied the correlation between \cancel{E}_T^{perp} and dijet invariant mass. Figure 6.1 shows distributions in dijet mass for multijet events that satisfy different cuts on \cancel{E}_T^{perp} . Figure 6.1a shows in red the dijet mass distribution for multijet events that satisfy $\cancel{E}_T^{perp} > 34$ GeV, normalized to unit area: this is the default sample used to describe the background in this channel. In blue, it shows the distribution for multijet events that satisfy $\cancel{E}_T^{perp} > 14$ GeV. Given the small number of entries in the

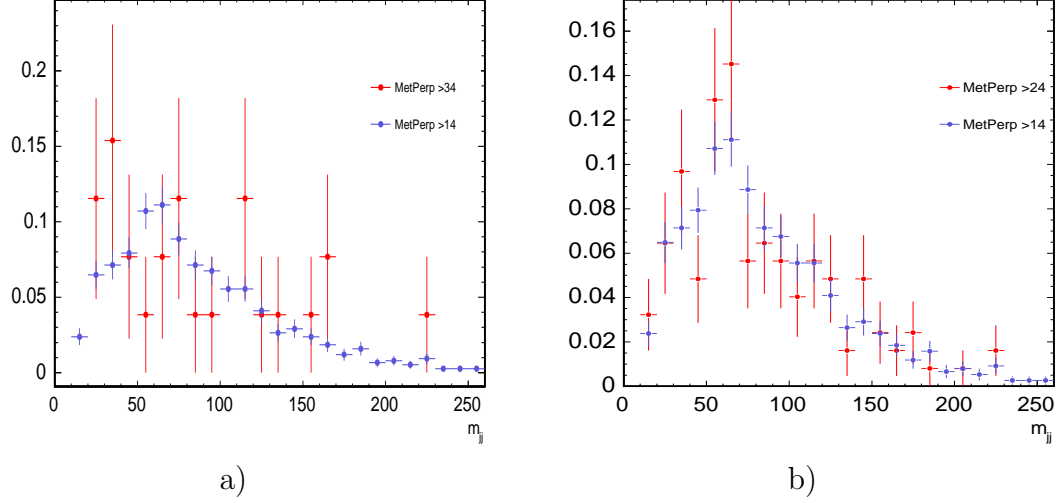


Figure 6.1: a) Shape comparison of dijet mass distribution for multijet events satisfying $\cancel{E}_T^{perp} > 34$ GeV, in red, and satisfying $\cancel{E}_T^{perp} > 14$ GeV, in blue. b) Shape comparison of dijet mass distribution for multijet events satisfying $\cancel{E}_T^{perp} > 24$ GeV, in red, and satisfying $\cancel{E}_T^{perp} > 14$ GeV, in blue.

default template, and the size of the bin-by-bin statistical uncertainty on it, the two templates are consistent. We did an additional check (see figure 6.1b) comparing the dijet mass distribution for events that satisfy $\cancel{E}_T^{perp} > 24$ (in red) and $\cancel{E}_T^{perp} > 14$ (in blue), and the templates are also consistent, given the statistical uncertainty. So, we conclude that we can use events with $\cancel{E}_T^{perp} > 14$ GeV to model the multijet background dijet mass distribution for events satisfying $\cancel{E}_T^{perp} > 34$ GeV.

The background and signal di-jet mass distributions for Single and Double Tag signal regions are shown in figures 6.2.

6.3 Expected and Observed Limit

We calculate the expected limit using pseudo-experiments. For each pseudo-experiment, the dijet mass distribution is randomly populated, according to the background only

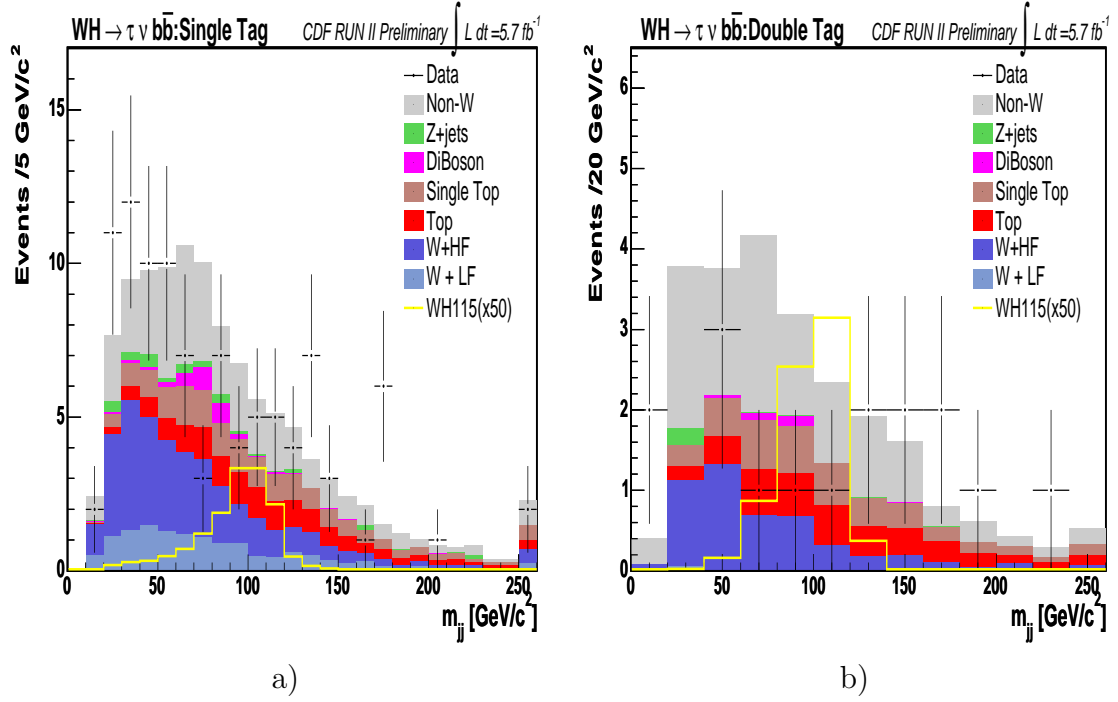


Figure 6.2: Dijet system mass distribution, for events with a) one b-tagged jet, b) 2 b-tagged jets. The background templates are normalized to the expected yields listed in table 5.8. The yellow curve correspond to the Higgs signal contribution ($m_H = 115 \text{ GeV}/c^2$), if the production cross section was 50 time the one predicted by the Standard Model.

expectation and the upper limit is computed. The median value of these upper limits defines the expected limit. The expected and observed 95% C.L. limit for the combined sample (Single Tag and Double Tag samples) are shown in table 6.1 and in figure 6.3. The expected and observed 95% C.L. limit for the individual channels are also shown (table 6.2 and figure 6.4 for the Single Tag sample, table 6.3 and figure 6.5 for the Double Tag sample).

CDF Run II Preliminary, $WH \rightarrow \tau\nu b\bar{b}$, 5.7fb^{-1}						
$\sigma(95 \text{ C.L. limit})/\sigma_{S.M.} * BR(H \rightarrow b\bar{b})$						
m_H	Observed	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
100	13.4	12.9	20.4	30.0	43.3	60.4
105	16.7	15.2	23.8	35.2	50.4	70.0
110	24.0	17.1	26.9	40.1	57.4	82.1
115	28.7	19.7	31.7	46.6	67.0	94.0
120	33.9	24.7	38.6	57.1	82.0	115.1
125	53.4	29.7	46.5	68.9	98.6	136.7
130	82.6	37.7	59.7	87.7	126.3	175.0
135	112.7	50.3	79.8	118.1	168.3	234.7
140	186.5	76.3	118.1	174.7	248.4	345.3
145	294.6	107.0	169.8	248.7	355.8	505.2
150	488.0	172.8	265.7	395.3	569.8	787.6

Table 6.1: Observed and expected limits on the ratio of the measured cross section to the one predicted by the Standard Model, for the Single and Double Tag samples combined.

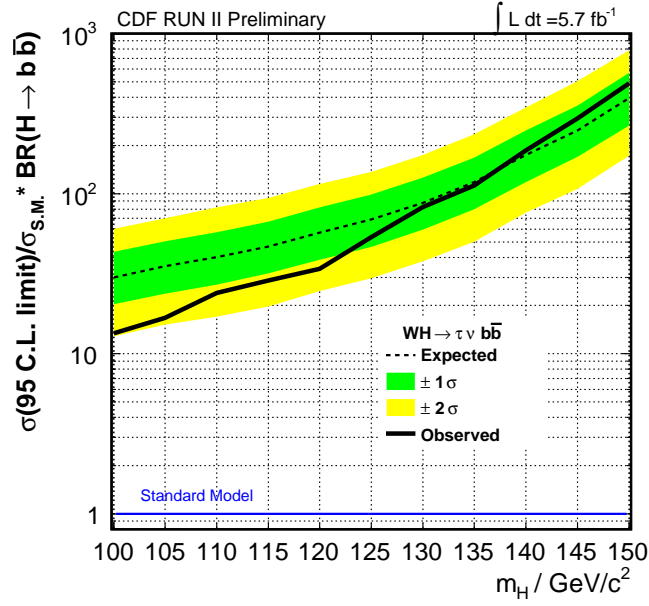


Figure 6.3: Expected and observed 95% C.L. limits, for the Single and Double Tag sample combined.

CDF Run II Preliminary, 5.7fb^{-1}
 $WH \rightarrow \tau\nu b\bar{b}$, Single Tag

m_H	Observed	-2σ	-1σ	Median Value	$+1\sigma$	$+2\sigma$
100	25.1	19.2	30.1	44.6	64.4	91.0
105	32.6	21.6	34.8	51.3	73.6	102.2
110	38.6	25.0	39.9	58.4	83.4	114.4
115	48.8	29.4	46.5	69.3	98.6	138.3
120	63.6	35.4	56.2	83.9	120.4	169.7
125	93.4	44.0	69.0	102.7	145.8	205.9
130	133.3	55.4	87.3	129.8	185.5	256.9
135	193.8	73.5	115.8	172.5	245.6	342.9
140	303.3	107.1	169.0	248.7	354.6	497.7
145	440.3	153.4	240.4	356.5	509.2	719.6
150	685.5	240.9	380.9	564.5	799.5	1124.7

Table 6.2: Observed and expected limits on the ratio of the measured cross section to the one predicted by the Standard Model, in the Single Tag sample.

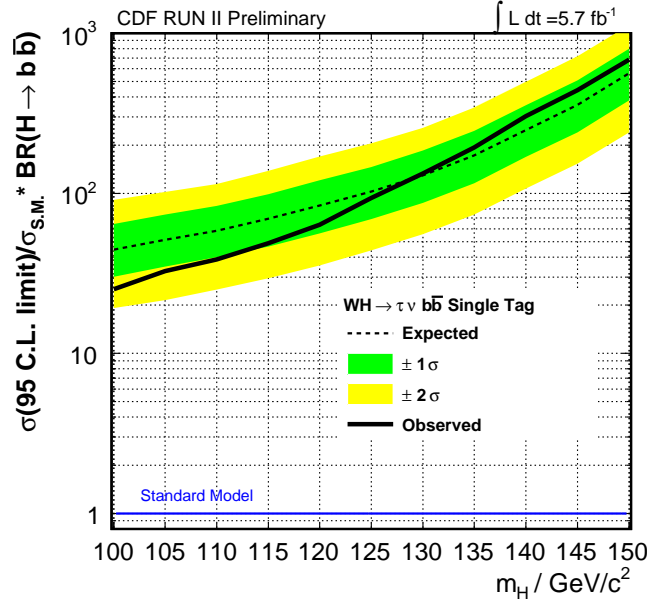


Figure 6.4: Expected and observed 95% C.L. limits, Single Tag channel.

CDF Run II Preliminary, 5.7fb^{-1}
 $WH \rightarrow \tau\nu b\bar{b}$, Double Tag

m_H	Observed	-2σ	-1σ	Median Value	$+1\sigma$	$+2\sigma$
100	22.0	18.6	28.7	42.1	59.6	85.4
105	23.5	22.1	34.0	50.2	70.9	98.8
110	31.6	25.0	38.9	57.4	82.0	114.4
115	38.0	29.7	45.6	67.3	95.1	132.8
120	50.4	36.9	56.2	82.4	118.5	169.5
125	62.9	43.0	66.7	98.1	138.3	190.2
130	82.1	55.6	83.5	122.3	174.0	248.8
135	127.2	76.0	115.5	171.1	241.3	341.4
140	218.7	113.2	174.2	257.3	364.7	513.1
145	314.6	164.8	254.9	370.4	524.6	737.2
150	589.0	262.0	401.2	588.2	844.1	1184.4

Table 6.3: Observed and expected limits on the ratio of the measured cross section to the one predicted by the Standard Model, in the Double Tag sample.

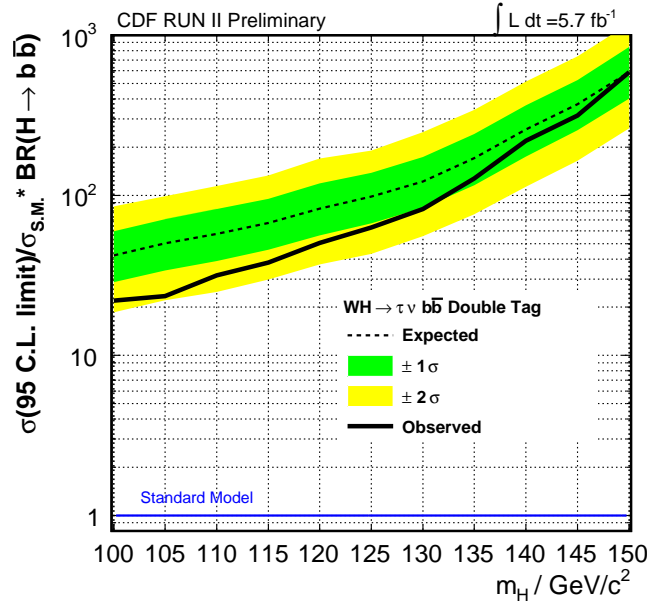


Figure 6.5: Expected and observed 95% C.L. limits, for the Double Tag channel.

Chapter 7

Conclusions

We presented the results of a search for the Standard Model Higgs boson decaying to $b\bar{b}$, produced in association with a W boson decaying into an hadronically decaying τ lepton and neutrinos. We find that for the dataset corresponding to an integrated luminosity of 5.7 fb^{-1} , the data agree with the SM background predictions within the systematic uncertainties, so we set a 95% confidence level upper limits on the H cross section times the branching ratio of the Higgs decay to a $b\bar{b}$ pair ($\sigma(p\bar{p} \rightarrow W^\pm H) \times \text{BR}(H \rightarrow b\bar{b})$). The resulting observed (expected) upper limit for a Higgs mass of $115 \text{ GeV}/c^2$ is 28.7 (46.6) times the Standard Model expectation. We set limits across a range of Higgs masses from $100 \text{ GeV}/c^2$ to $150 \text{ GeV}/c^2$, with observed(expected) limits ranging from $13.4 (30.0) \times \text{S.M. expectation}$ to $488.0 (395.3) \times \text{the S.M. expectation}$.

This is the first time that a dedicated search for $WH \rightarrow \tau\nu b\bar{b}$ is performed at CDF, and the composition of the data acquired with the TAU_MET trigger is studied extensively. It was clear from the start that the analysis of this channel would not set an exclusion limit on the Higgs boson by itself. Actually, at the Tavatron there is no single channel that can reach the Standard Model sensitivity for a low mass Higgs boson, not even using the full dataset. For this reason, all the Higgs boson searches

performed at CDF are combined together to set the most stringent limit. Combining all the most recent CDF results, that included analyses with $4.0 \leq \int L dt \leq 8.2 \text{ fb}^{-1}$, the 95% upper observed(expected) limit on the Higgs boson production rate is 1.55 (1.49) times the Standard Model expectation, for $m_H = 115 \text{ GeV}/c^2$ [57]. If we were to perform all the Higgs searches with the full dataset, this would still not be enough to reach the sensitivity to the SM production rate. Hence, it makes sense to combine the results from the Tevatron experiments. Assuming D0 reaches similar performances to CDF, then the combined limit should scale as $1/\sqrt{2}$: as of July 2011, the combined 95 % upper observed (expected) limit on the Higgs boson production rate is 1.16 (1.17) times the Standard Model expectation, for $m_H = 115 \text{ GeV}/c^2$ [15]. In the last few years CDF and D0 have steadily improved the sensitivity of the Higgs boson searches: a constant effort has been made to improve analysis of the most sensitive channels, and efforts have been started to include analyses of channels with lower expected sensitivity. The search for $WH \rightarrow \tau \nu b \bar{b}$ is part of the current effort to improve the sensitivity to the Higgs boson searches at the Tevatron. The channel analyzed in this thesis is not the most sensitive one and does not improve significantly the current limits on the Higgs boson production cross section. Working on it, though, proved to be a good opportunity to learn about many aspects of data analysis, from optimizing the trigger and developing a procedure to model its effect on simulated events, to techniques to build background models. Although the specific procedures I developed will not be applied to other analyses, the main ideas are still relevant for LHC analyses.

With the Tevatron end of operations on September 30th, the final data sample of $\int L dt \sim 10 \text{ fb}^{-1}$ is available to both CDF and D0 collaborations. In the mean time, the Large Hadron Collider at CERN is performing very well, and the detectors, ATLAS and CMS, are collecting $\sim 450 \text{ pb}^{-1}$ per week of pp collision at $\sqrt{s} = 7 \text{ TeV}$, and have excluded at 95% C.L. the existence of the Higgs boson for a large range of possible masses. Already, the results from both the Tevatron and LHC presented

at the summer conferences of 2011 corners the Higgs boson, if it exists, to a very narrow window, $114 \leq m_H \leq 145 \text{ GeV}/c^2$.

Figure 7.1 shows the projected median expected 95 % C.L. upper limit for CDF, as a function of the integrated luminosity analyzed by the experiment. The CDF results are scaled to twice the integrated luminosity, in an attempt to describe the effect of combining the result with D0, assuming the performances of the two experiments are similar. The solid lines represent the $1/\sqrt{\int L dt}$ projection of how the limit changes when new data are included in the analysis. The orange band describes the range of improvements that were considered possible in Summer 2007. These projections seem to indicate that it may be possible for the final Tevatron results to reach the S.M. sensitivity to a Higgs boson with $m_H = 115 \text{ GeV}/c^2$.

Figure 7.2a shows the projected median expected 95% C.L. upper limit for ATLAS

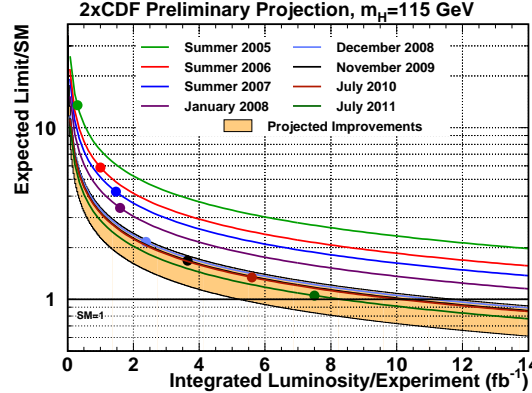


Figure 7.1: Projected median expected limits on Higgs boson production rate, scaling CDF performances by $1/\sqrt{2}$ to represent CDF and D0 combination.

as a function of the Higgs boson mass. The colored area shows the excluded region from LEP [12] and the from the Tevatron [59]. The different lines correspond to the size of the analyzed sample. Similar performances are expected from CMS as well. A single LHC experiment is expected to exclude the presence of a Higgs boson in the whole mass region with 5 fb^{-1} of data. As of October 2011, more than 4 fb^{-1} were

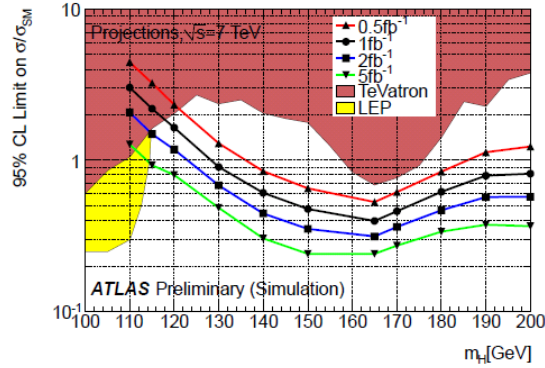


Figure 7.2: Projected median expected limits on Higgs boson production rate as a function of Higgs mass for ATLAS [58]

collected by both CMS and ATLAS and 5 fb^{-1} are expected for early spring 2011.

After being the most sought after particle of the past 40 years, the truth about the Higgs boson will be at hand in the next year or so. Should it exist, we would have understood the mechanism at the origin of particle masses. If no sign of its presence appears in LHC and Tevatron data, a different mechanism must be at the origin of the electroweak symmetry breaking, in which case there are reasons to believe that some new physics should be at play at the TeV scale, in reach of the LHC. And this may be even more exciting.

Bibliography

- [1] F. Halzen, A. D. Martin, Quarks & Leptons: An Introductory Course in Modern Particle Physics, (1984), John Wiley & Sons, Inc
- [2] S. Weinberg, "A Model Of Leptons," Phys. Rev. Lett. 19, 1264 (1967)
- [3] P. W. Higgs, Phys. Lett. 12, 132 (1964).
- [4] J. F. Gunion, H. E. Haber, G. Kane and S. Dawson, "The Higgs Hunter's Guide" (1990) Perseus Publishing, Cambridge, MA
- [5] TeV4LHC Higgs Working Group
url<http://maltoni.home.cern.ch/maltoni/TeV4LHC/>
- [6] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), CERN-2011-002 (CERN, Geneva, 2011), [arXiv:hep-ph/1101.0593]
- [7] Dawson, Lectures given at the 1994 TASI summer school, Boulder, Colorado, [arXiv:hep-ph/9411325v1].
- [8] T. Hambye and K. Riesselmann, Phys. Rev. D55, 7255 (1997), [hep-ph/9610272].
- [9] The CDF, D0 Collaborations, the TEVNPBWG Working Group [arXiv:hep-ex/1103.3233v3]

- [10] The LEP Electroweak Working Group <http://lepewwg.web.cern.ch/LEPEWWG/>
- [11] ALEPH Collab., Phys. Lett. B526, 191 (2002).
- [12] LEP Working Group for Higgs boson searches, R. Barate et al., Phys. Lett. B565, 61 (2003), [arXiv:hep-ex/0306033]
- [13] C. F. Kolda and H. Murayama, JHEP 07, 035 (2000), [arXiv:hep-ph/0003170]
- [14] S. Martin, A Supersymmetry Primer [arXiv:hep-ph/9709356v6]
- [15] Tevatron New Phenomena & Higgs Working Group [arXiv:hep-ex:1107.5518]
- [16] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka
url:<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>
- [17] The ATLAS collaboration, ATLAS-CONF-2011-135,
url:<http://cdsweb.cern.ch/record/1383838>
- [18] The CMS collaboration, CMS-PAS-HIG-11-022,
<http://cdsweb.cern.ch/record/1376643>
- [19] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, Phys. Lett. B513, 232 (2001);
D.B. Kaplan and H. Georgi, Phys. Lett. B136, 183 (1984) and Phys. Lett. B145,
216 (1984);, T. Banks, Nucl. Phys. B243, 123 (1984); D.B. Kaplan, H. Georgi,
and S. Dimopoulos, Phys. Lett. B136, 187 (1984); M.J. Dugan, H. Georgi, and
D.B. Kaplan, Nucl. Phys. B254, 299 (1985).
- [20] S. Weinberg, Phys. Rev. D19, 1277 (1979); L. Susskind, Phys. Rev. D20, 2619
(1979)
- [21] Fermilab Beams Division, Tevatron Run 2 Handbook (1998)

- [22] F. Abe *et al.*, Nucl. Instrum. Methods Phys. Res. A **271**, 387 (1988);
D. Amidei *et al.*, Nucl. Instrum. Methods Phys. Res. A **350**, 73 (1994);
F. Abe *et al.*, Phys. Rev. D **52**, 4784 (1995);
P. Azzi *et al.*, Nucl. Instrum. Methods Phys. Res. A **360**, 137 (1995);
The CDF II Detector Technical Design Report, Fermilab-Pub-96/390-E.
- [23] P. T. Lukens (CDF IIb) (2003), FERMILAB-TM-2198
- [24] R. Blair et al. (CDF II Collaboration), The CDF Run II Detector Technical Design Report, 1996, FERMILAB-Pub-96/390-E.
- [25] A. Canepa *et al.* L2CAL - The L2 Calorimeter Trigger Upgrade, Nuclear Science, IEEE (2009)
- [26] W.Badgett *et al.* The CDF Run II Event Structure, CDF/PUB/ELECTROWEAK/PUBLIC/4152
- [27] A. Heijboer http://www-cdf.fnal.gov/internal/physics/joint_physics/HiggsTTF/EffTool
- [28] A. Anastassov *et al.* Tau Reconstruction Efficiency and QCD Fake Rate in Run2 CDF internal note 6308
- [29] Robert G. Wagner, Electron Identification for Run II: Algorithms, CDF internal note 5456 (2003)
- [30] T. Spreitzer et al., Electron Identification in Offline Release 6.1.2, CDF internal note 7950 (2006)
- [31] T. Aaltonen et al (CDF Collaboration), Phys. Rev. Lett. 103, 101802 (2009). arXiv: 0906.5613
- [32] A.Apresyan et al, Search for the Standard Model Higgs boson in the \cancel{E}_T plus jets sample, CDF internal note 9585

- [33] Simona Rolli, Lubomir Lovas, Muon ID, Reconstruction and Trigger Efficiencies and Scale Factors for Period 9-12 data, CDF internal note 9085 (2007).
- [34] B. Flaughner, J. Mueller, A Guide to JETCLU: The CDF Jet Clustering Algorithm, CDF internal note 1814 (1992).
- [35] A. Batthi et al, Nucl. Instrum. Meth. A, 566:375, 2006
- [36] Acosta, D. et al. , CDF collaboration, Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\text{TeV}$ using lepton + jets events with secondary vertex b -tagging, Phys Rev D.71.052003 (2005)
- [37] T. Sjostrand, P. Eden, C. Friberg, L. Loennblad, G. Miu, S. Mrenna and E. Norrbin, Computer Phys. Commun. 135 (2001) 238 (LU TP 00-30) [hep-ph/0010017].
- [38] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A.D. Polosa, J. Energy Phys., 0307 (2003) 001 [hep-ph/0206293]
- [39] J.A. Wall et al. Madgraph/madevent v4: the new web generation. J. High Energy Phys., 0709:028, 2007 Fabio Maltoni, Tim Stelzer, J. High Energy Phys., 0302 (2003) 027 [hep-ph/0208156]
- [40] S. Jadach, J. H. Kuhn, and Z. Was, Comput. Phys. Commun. 64 (1990) 275; M. Jezabek, Z. Was, S. Jadach, and J. H. Kuhn, Comput. Phys. Commun. 70 (1992) 69; S. Jadach, Z. Was, R. Decker, and J. H. Kuhn, Comput. Phys. Commun. 76 (1993) 361380; P. Golonka, E. Richter-Was, and Z. Was, [hep-ph/0009302]
- [41] S. Agostinelli *et al.* (GEANT 4 Collaboration), NIM A 506 (2003) 250-303.
- [42] J. Baglio and A. Djouadi [arXiv:hep-ph/1003.4266v2]
- [43] A. Djouadi, J. Kalinowski and M. Spira, Comput. Phys. Commun. 108, 56 (1998) [arXiv:hep-ph/9704448]

- [44] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, Eur. Phys. J. C 63, 189 (2009) [arXiv:hep-ph/0901.0002].
- [45] M.L. Mangano et al., J. High Energy Phys., 07:001 (2003)
- [46] D. Sherman (Ph.D. Thesis), FERMILAB-THESIS-2007-82
- [47] U. Langenfeld, S. Moch and P. Uwer, Phys. Rev. D 80, 054009 (2009) [arXiv:hep-ph/0906.5273].
- [48] N. Kidonakis [arXiv:hep-ph/1005.3330].
- [49] N. Kidonakis, Phys. Rev. D 74, 114012 (2006) [arXiv:hep-ph/0609287].
- [50] A. Abulencia et al. (CDF Collaboration), Measurements of inclusive W and Z cross sections in p anti-p collisions at $s^{*}(1/2) = 1.96\text{-TeV}$, J. Phys. G 34, 2457 7 (2007) [arXiv:hep-ex/0508029].
- [51] J. Campbell, K. Ellis, <http://mcfm.fnal.gov/>
- [52] J. M. Campbell and R. K. Ellis, Phys. Rev. D 60, 113006 (1999) [arXiv:hep-ph/9905386].
- [53] A. Abulencia *et al.*, Phys. Rev. **D71**, 072005 (2005)
- [54] from private correspondance with Federico Sforza
- [55] MCLimit - Limit Calculator
<https://plone4.fnal.gov:4430/P0/phystat/packages/0711001>
- [56] H. L. Lai et al. (CTEQ Collaboration), Eur. Phys. J. C 39 12, 375 (2000).
- [57] http://www-cdf.fnal.gov/physics/new/hdg/Results_files/results/cdfcombeps2011/
- [58] he ATLAS collaboration, ATL-PHYS-PUB-2010-015 \ <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/PublicHGPlotsAtlPhysPub2010726>

[59] T. Aaltonen et al (CDF Collaboration), CDF Conference Note 10223