SEARCH FOR THE HIGGS BOSON AND RARE STANDARD MODEL PROCESSES IN THE $\not\!\!E_T$+B-JETS SIGNATURE AT THE COLLIDER DETECTOR AT FERMILAB

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Karolos Potamianos

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2011

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

Research is a collaborative endeavor and this work is no exception. Many persons contributed directly or indirectly to the contents of this dissertation: supervisors, collaborators, friends, and family.

I am very grateful for the unlimited support I received from my research supervisor, Daniela Bortoletto. I could always turn to her, learning much from her feedback and advice. She always was available despite her busy schedule. She taught me how to do research, from the understanding of the problem (through many plots!) to its resolution, and that physics insight prevails over the bugs in the code. Her enthusiasm always made me set higher goals.

Fabrizio Margaroli took a capital role in the completion of this work, providing guidance and feedback on a daily basis. Without his continuous impulse and flow of ideas, I wouldn't have finished my first analysis so quickly. He was always encouraging, and confident of my ability to cope with the difficulties of the job. However, his major contribution is for enabling me to be in Chicago city more often. I really appreciate the hospitality (this also goes to his wife Federica Moscato). I am grateful to have met him.

I would like to thank the close collaborators involved in this analysis for their help and advice. Artur Apresyan helped me start working on CDF, without ever complaining of me continuously bugging him; the $\not\!\!E_T$+jets framework I inherited from him was of invaluable help. I also thank him for the countless discussions, both on physics and other topics. Qiuguang Liu took as special part in this endeavor, by his contribution to this analysis, and by his presence at the Purdue house. We shared many interesting discussions. I learned a lot from interacting with him and his wife Mindy. I am thankful for meeting him. Oscar Gonzalez much impacted this work by his physics insight. I thank him for ensuring I was being a careful analyst,

for the interesting and varied conversations we had together, and for playing tennis and other sports. Thomas Junk helped me understand the subtleties of statistical analysis more than any textbook. I thank him for the countless times he answered my questions on Bayesian inference in practice. I thank the members of the $\not\!\!E_T b\bar{b}$ and Higgs groups (Ben, Homer, Mike, Kyle, and many more) and the Higgs conveners (Eric and Craig) for their constrictive feedback and comments as I was progressing in my work. Numerous people in the operations and offline groups of CDF have also contributed to this work, by providing data to analyze. I thank them for their dedication in having the detector an other infrastructure up and running. I would also like to thank Emjai and Heather, the secretaries of the High Energy Physics group at Purdue, for their invaluable help in dealing with administrative tasks.

I met many interesting people during my graduate studies, and am very grateful for that. The numerous Italians (Diego, Matteo, Viviana, Angelo, Federico, Pierluigi, Matteo), Spaniards (Jesus, Barbara), Greeks (Costas, John, Victoria), and many more from all around the world have contributed to who I have become.

Since I left the Polytechnic School of the University of Brussels (ULB), I have had numerous occasions to appreciate the quality of the education I received there. I would like to thank all the Professors with whom I interacted, and particularly Daniel Baye, Alain Dubus, Marc Haelterman and Anne Delandtsheer. The same gratitude goes towards the Professors who taught me at Purdue.

I am indebted to my family, my father Gerassimos, my mother Rita, and my brother Alexandros, for the continuous support and for giving me the strength to cross the Atlantic to do research. I thank my parents for the education they have provided me, which has helped me to cope with many situations. Even though my research is not always clear to them, they always were confident in my capabilities, especially my brother.

Last but certainly not least, I thank Pauline, who stood next to me. She made important sacrifices for the completion of this research work. Without her presence and encouragement, my life would be very different. I thank her for who she is, and

for everything I share with her. I also thank her parents, Jean-Paul and Jenny, for their encouragements.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

SYMBOLS

CEM	Central electromagnetic calorimeter
CES	Central electromagnetic showermax
CHA	Central hadronic calorimeter
COT	Central outer tracker
EWK	Electroweak
L1	Level 1 trigger subsystem
L2	Level 2 trigger subsystem
L3	Level 3 trigger subsystem
$\not\!\!E_T$	missing transverse energy
JetProb	Jet probability b-tagging algorithm
MJ	Multi-jet
$\not\!\!p_T$	missing transverse momentum from tracks
PES	Plug electromagnetic showermax
PEM	Plug electromagnetic calorimeter
PHA	Plug hadronic calorimeter
SecVtx	Secondary vertex b-tagging algorithm
SVX	Silicon vertex detector
TRM	Tag-Rate-Matrix
TOF	Time-of-flight dectector
WHA	Wall hadronic calorimeter

ABSTRACT

Potamianos, Karolos. Ph.D., Purdue University, December 2011. Search for the Higgs Boson and Rare Standard Model Processes in the $\not\!\!E_T$+b-jets Signature at the Collider Detector at Fermilab. Major Professor: Daniela Bortoletto.

We study rare processes of the standard model of particle physics (SM) in events with missing transverse energy ($\not\!\!E_T$), no leptons, and two or three jets, of which at least one is identified as originating from a b-quark ($\not\!\!E_T$+b-jets signature). We present a search for the SM Higgs boson produced in association with a W or Z boson when the Higgs decays into $b\bar{b}$. We consider the scenario where $Z \to \nu\nu$, or $W \to l\nu$ and the lepton escapes detection. This dissertation analyzes 7.8 fb^{-1} of data collected by the CDF II experiment at Fermilab.

For the first time, we analyze events with *relaxed kinematic* requirements, yielding an increase of 30-40% in acceptance to the WH/ZH signal. We collect events from three different triggers and parametrize the efficiency of their logical combination (OR) using a novel artificial neural network (NN) technique. To increase the sensitivity to the signal, we implement a NN to remove the huge instrumental background. An additional NN is used to discriminate the Higgs signal from the remaining background.

We check our background modeling by comparing data against backgrounds in many control regions, and find good agreement. Observing no significant excess in the data, we place 95% confidence level (C.L.) upper limits on the Higgs boson production cross section. For a mass of 115 GeV/c^2 the expected (observed) limit is 2.9 (2.3) times the standard model prediction. Compared to the last iteration of this analysis, this result improves the significance by 10% throughout the $100 - 150$ GeV/c^2 mass range. This is one of the most sensitive at the Tevatron in this mass range.

We cross-check the tools developed in this dissertation by measuring the cross-section of top pair, electroweak single top and diboson ($WZ + ZZ$) production.

We have a habit in writing articles published in scientific journals to make the work as finished as possible, to cover up all the tracks, to not worry about the blind alleys or describe how you had the wrong idea first, and so on. So there isn't any place to publish, in a dignified manner, what you actually did in order to get to do the work.

Feynman, Richard Philips (1918 – 1988), Nobel Lecture, 1966.

1. INTRODUCTION

The standard model of particle physics (SM) is an attempt to describe nature. It is currently the most complete theory developed by mankind. Yet, it does not explain observed phenomena, such as neutrino oscillations [1, 2] and CP-violation [3]. Many theories beyond the standard model include these and other effects. Current and future particle accelerators and other experiments probe or will probe such theories. They will hopefully reveal new secrets about nature.

Of the particles predicted by the SM, only the Higgs boson has yet to be observed. Also, many SM parameters are far from being determined precisely. This can partially be addressed by using particle accelerators such as the Tevatron and the LHC, and looking at the rarest standard model processes, such as diboson, single top and Higgs production. There is plenty of work to be done regarding these physical processes, either to discover them or determine their properties precisely.

Usually, experimentalists start using the cleanest signatures in their searches. Typically this would imply using particles with easily measurable properties, such as photons, electrons, or muons, and avoiding as much as possible to rely on the measured properties of hadrons or on quantities such as the transverse energy imbalance in the detector ($\not\!\!E_T$). Analyses using the latter have an extra level of difficulty and are very challenging. However, it is often necessary to go the challenging way because of the physics involved. In addition, as time goes by, experimentalists gain understanding of the processes at place and confidence in their tools. With the associated gain in precision, these results become significant enough. This dissertation is about these challenging analyses. In particular, we analyze the $\not\!\!E_T$+b-jets signature.

1.1 Analysis flow

Everything begins with several simulated samples of known SM processes, which constitute the analysis background, and a signal simulated sample, which stems from the physics process being characterized. The goal is to compare these simulations to the data recorded from the detector.

For this purpose, kinematic selections are defined in order to isolate as much signal events as possible in a region with as small backgrounds as possible. Next to the signal region, in which we compare data to the simulation, we define several other regions, which serve for *background model verification*. This step is necessary to give confidence in the simulation in regions where no sensitivity to the signal is expected.

In this dissertation, we use multivariate analysis techniques to isolate the main background – QCD multijet – in order to further improve the signal purity. After rejecting QCD, we plot distributions of quantities that can distinguish as much as possible the signal from the backgrounds. These histograms are analyzed using statistical analysis tools in order to determine the level of compatibility of the simulation with the data.

1.2 Outline of this Dissertation

In chapter 2, we present a brief history of physics, followed by a more technical description of the standard model of particle physics in chapter 3. We also report on the status of the field. We then proceed with the description of the experimental apparatus: the Tevatron collider and the CDF detector (chapter 4).

The particles involved in the processes taking place when a proton p and an anti-proton $\bar{p}$ collide in the Tevatron are not easy to identify: they interact with the detector material in various ways, leaving small bits of information. Chapter 5 is about the combination of that information to reconstruct the identity and properties of the products of the collision. In chapter 6, we discuss techniques to identify b quarks; this identification stage significantly reduces the backgrounds.

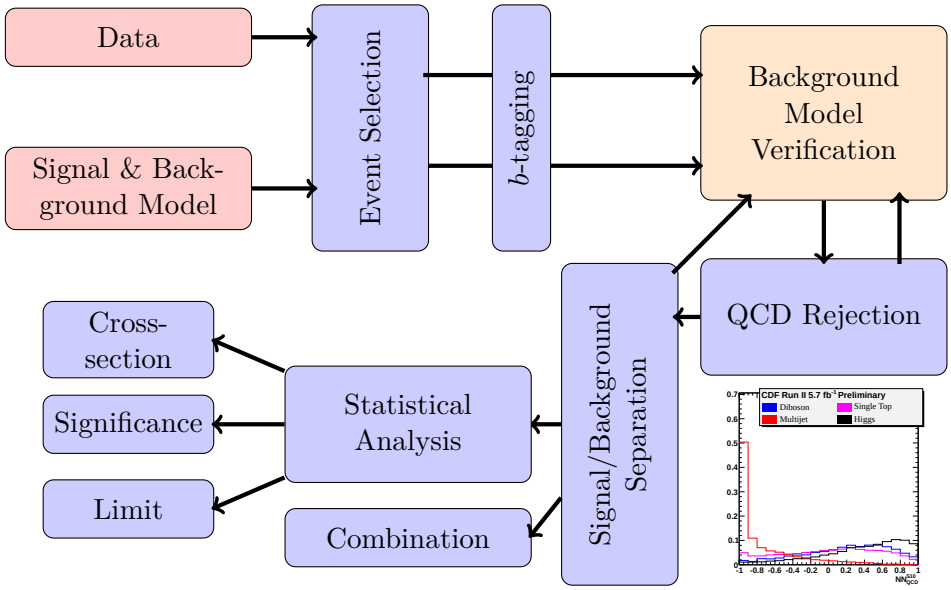


Figure 1.1.: $\not\!\!E_T$+b-jets analysis flow.

The information provided by the detector is too complex to be looked at without proper processing. The probabilistic nature of quantum physics and the similarity between the processes involved make it impossible to claim that a given collision (event) recorded by the detector is of a particular process. However, on average, we are able to compare our theories with experiment. Indeed, we can gain insight into nature by comparing the distribution of observables (i.e. measurable properties of the collisions) in simulations with that in the data. Chapter 7 describes how these simulations are performed. Chapter 8 describes the $\not\!\!E_T$+b-jets dataset; we present the numerous physics processes yielding a large $\not\!\!E_T$ and two high-energy jets[1], as well as how they are simulated.

We address the extra level of difficulty associated with the $\not\!\!E_T$+b-jets dataset using multivariate analysis techniques to combine the information from multiple observables. In chapter 9, we introduce artificial neural networks (NN) and discuss their purpose, implementation, and limitations. The events of interest to us are col-

[1]Jets are a spay of particles originating from the decay of a parent particle taking part in the collision.

lected using dedicated triggers, which decide whether an event should be recorded or not. Chapter 10 is about the parameterization of the trigger efficiency using a NN.

In order to be able to identify the small signal we are after, we must isolate it from the numerous backgrounds. Chapter 11 describes the event selection process. We describe the kinematic selection criteria used to increase the signal to background ratio (S/B), and a NN exploiting the correlations among observables to reject the background even more.

Before discussing the results of our analysis, we present in chapter 12 the tools used for the statistical interpretation of the distributions mentioned above, and discuss in chapter 13 the sources of uncertainty affecting our measurement.

We finally proceed with a search for the SM Higgs boson produced in association with a vector boson (chapter 14). We discuss the results, as well as those form several validations of the analysis technique (chapter 15). In chapter 16, we summarize our work, and briefly discuss future prospects.

1.3 Publications

The work presented in this dissertation was published in peer reviewed journals:

- T. Aaltonen *et al.* [The CDF Collaboration], "First Observation of Electroweak Single Top Quark Production", Phys. Rev. Lett. **103**, 092002 (2009), arXiv:0903.0885 [hep-ex]. [100+ citations]

- T. Aaltonen *et al.* [The CDF Collaboration], "A Search for the Higgs Boson Using Neural Networks in Events with Missing Energy and b-quark Jets in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV", Phys. Rev. Lett. **104**, 141801 (2010), arXiv:0911.3935 [hep-ex].

- T. Aaltonen *et al.* [The CDF Collaboration], "Search for single top quark production in pbar p collisions at $\sqrt{s} = 1.96$ TeV in the missing transverse

energy plus jets topology", Phys. Rev. D **81**, 072003 (2010), arXiv:1001.4577 [hep-ex].

Another paper, about the results in chapter 14, is under preparation. We have also plans to publish the details of the trigger parameterization presented in chapter 10.

2. A BRIEF HISTORY OF PHYSICS

Since the dawn of time, mankind has tried to understand its condition. Where do we come from? What is the world made of?

Early human attempts to answer this question involve a belief or a reasoning. Events occurring in nature were first explained by supernatural causes, which involved religion or superstition. Around the sixth century B.C., Greek philosophers tried to explain the world around them solely through reasoning. At that time, the focus was on ontology, cosmology, and mathematics.

2.1 The Philosophers of Nature

Thales of Miletus (Θαλῆς ὁ Μιλήσιος, c. 624 - c. 546 B.C.) is reportedly the first of the philosophers of nature. The Milesian school of thought he founded presented a view of nature in terms of methodologically observable entities. Therefore, Thales and his followers were considered the first scientific philosophers. They were after the quintessential substance of which the world was made of (ἀρχή). Thales thought it to be water, Anaximander (Ἀναξίμανδρος, c. 610 - c. 546 B.C.) thought it to be an undefined element (ἄπειρον), while Anaximenes (Ἀναξιμένης, 585 - 528 B.C.) considered that only air (in the sense of an *aether*, or αἰθήρ), could play this role. All these doctrines implied things were corpuscular. Later on, Heraclitus of Ephesus (Ἡράκλειτος ὁ Ἐφέσιος, c. 535 - c. 475 B.C.) taught that everything is in flux[1], adding the idea of motion.

Parmenides of Elea (Παρμενίδης ὁ Ἐλεάτης, c. 5th century B.C.) rejected the existence of motion, change and void. He claimed that change and motion were mere

[1] ποταμοῖσι τοῖσιν αὐτοῖσιν ἐμβαίνουσιν, ἕτερα καὶ ἕτερα ὕδατα ἐπιρρεῖ: Ever-newer waters flow on those who step into the same rivers.

illusions, and that all existence is a single, all-encompassing, and unchanged mass (this conception is termed monism). He rejected sensory experience as the path to understanding the universe, and instead used purely abstract reasoning. He believed that there is no such thing as void, since if the void is, then it is not nothing, and therefore it is not the void. This meant for him that motion is impossible, because there is no void to move into.

The philosophers after Parmenides concluded that there must exist multiple elements, which are not reducible to one another, and abandoned monism in favor of pluralism. Empedocles (Ἐμπεδοκλῆς c. 490 – c. 430 B.C.) argued that it is impossible to come into existence from a non-existence, nor to die completely or annihilate, and that what is commonly referred to as coming into existence and dying is in fact the mixture and separation of what was mixed, i.e. fire, air, water and earth. These four elements[2] were set in motion by love (φιλία) and strife (νεῖκος).

Leucippus (Λεύκιππος, first half of the 5th century B.C.) and his pupil Democritus (Δημόκριτος, c. 460 - c. 370 B.C.) are reportedly the first to develop the theory of atomism, in which everything is composed entirely of imperishable, indivisible elements called atoms (ἄτομα), and vacuum, since like Parmenides, they believed there can be no movement in the absence of vacuum. The atomists only admitted mechanistic answers, and thus looked exclusively for mechanistic questions, leaving out teleological questions. Aristotle (Ἀριστοτέλης, 384 - 322 BCE) even suggested that observation of physical phenomena ultimately leads to the discovery of the laws governing them. Epicurus (Ἐπίκουρος, 341 - 270 B.C.), however, was less sure that atomism could adequately explain specific natural phenomena, such as thunderstorms or earthquakes, as he lacked the necessary knowledge about the laws that rule the atoms. He is a key figure in the development of scientific method because of his insistence that only that which was tested through direct observation and logical deduction should be believed.

[2]The term element (στοιχεῖον), however, is reportedly first used by Plato (Πλάτων, 424 - 348 B.C.).

Lacking the experimental tools that were developed only one millennium later, these philosophers tried to go as far as they could to explain the world around them, solely relying on their reasoning or sensory experience. Yet, they achieved to develop most of the concepts behind modern science (and thus physics). The point-like particles of today's models are nothing more than atoms, and Empedocles' love and strife grasp exactly what modern physics refers to as interactions. What these early philosophers truly lacked, however, is the ability to express their theories in quantitative terms. Except for (macroscopic) measurements in geography and astronomy[3], they lacked the experimental tools and the mathematical formalism to tackle problems quantitatively.

2.2 Classical Physics and Chemistry

The successors of the Greek philosophers were more occupied with teleological questions rather than mechanistic ones. Religious faith and the fear for repercussions hindered progress towards the true laws of nature. It took over one millennium, until the Renaissance, before scientists such as Galileo Galilei (1564-1642), Johannes Kepler (1571-1630), and Sir Isaac Newton (1642-1727) went against the Church, and used experimental and quantitative methods to discover the laws of nature, giving birth to classical physics and paving the way to modern science. Every one of them, physicist, astronomer, and mathematician, proposed mathematical laws to explain the motion of mechanical bodies. These laws are still taught today in physics courses, and are valid approximations to modern theories. Like the Greek philosophers, they believed that mathematics is the language in which to describe the laws of physics:

> La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi agli occhi (io dico l'universo), ma non si può intendere, se prima non s'impara a intender la lingua, e conoscer i caratteri ne' quali è scritto. Egli à scritto in lingua matematica, e i caratteri son triangoli,

[3]Eratosthenes (Ἐρατοσθένης, c. 276 - c. 195 B.C.) is reportedly the first to have computed the Earth's circumference with a good accuracy.

> cerchi ed altre figure geometriche, senza i quali messi è impossibile intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro labirinto.
>
> [The universe] cannot be read until we have learnt the language and become familiar with the characters in which it is written. It is written in mathematical language, and the letters are triangles, circles and other geometrical figures, without which means it is humanly impossible to comprehend a single word.
>
> Opere Il Saggiatore, p. 171 (1623)

Around the same time in Europe, Robert Boyle (1627-1691) advocated that matter consisted of atoms and cluster of atoms in motion, and that observable phenomena were the result of collisions of these atoms in motion, thus reinstating the ideas of Ancient philosophy. A little later, Antoine-Laurent de Lavoisier (1743-1794) discovered that water was made of oxygen ($_8$O) and hydrogen ($_1$H), while air was made of oxygen and nitrogen ($_7$N), which was a great step into explaining everyday notions in term of atoms. Next to the discovery and terming of these elements, he devised, together with others, a systematic chemical nomenclature. More importantly, he advanced chemistry to the level of a science.

It is John Dalton (1766-1844), however, who first devised the first *atomic theory*. The key concepts of which were that: (a) atoms of the same (a different) element have the same (different) properties; (b) atoms could combine in whole-number ratios (of atomic weights) to form chemical compounds, and (c) atoms were combined, separated or re-arranged in *chemical reactions*. He published the first table of relative atomic weights, which is the key idea which differentiate his theory from that of the Ancients.

Dmitri Ivanovich Mendeleev (Дми́трий Ива́нович Менделе́ев, 1834-1907) noticed that the elements exhibit a periodicity of properties when arranged according to their atomic weight. He is the author of the first periodic table of elements, using which he

was able to predict the existence of germanium ($_{32}$Ge), gallium ($_{31}$Ga) and scandium ($_{21}$Sc), which were found several years later. It was the first time that a theory *predicted* the existence of something not yet observed, and later confirmed to exist.

The corpuscular and undulatory nature of light

The nature of light has puzzled physicists for centuries. Indeed, two different viewpoints emerged to explain light, and yet none of them could prevail: some observable phenomena involving light were clearly in favor of one approach, while others favored the other. The milestones of these developments are as follows:

- Christiaan Huygens (1629-1695) and Isaac Newton had developed two competitive theories of light, in which light was thought to consist respectively of waves, or of particles.

- Thomas Young (1773-1829) designed the double-slit experiment[4], and established the wave theory of light and the concept of interference.

- James Clerk Maxwell (1831-1879) theoretically explained light as the propagation of electromagnetic waves, which Heinrich Rudolf Hertz (1857-1894) verified experimentally.

- After these developments, the undulatory nature of light seemed to have prevailed. Yet, in 1887 Hertz observed the *photoelectric effect*[5], during which matter emits electrons when absorbing energy from electromagnetic radiation. In 1902, Philipp Eduard Anton von Lenard (1862 -1947) observed that the energy of the emitted (photo-)electrons increased with the frequency of the light, and not its intensity, as the theory of Maxwell was thought to predict.

[4]In this experiment, a coherent light source such as a laser beam illuminates a thin plate pierced by two parallel slits, and the light passing through the slits is observed on a screen behind the plate. The undulatory nature of light causes the light waves passing through the two slits to interfere, producing bright and dark bands on the screen.
[5]This effect was initially known as the Hertz effet.

The 19th century ended without an answer to the fundamental question of the nature of light. Many other fields of physics developed in the 18th and 19th centuries, such as thermodynamics, fluid mechanics, statistical mechanics, as well as the formalism behind the mathematical expression of the universe, developed mainly by Joseph-Louis Lagrange (1736-1813) and Sir William Rowan Hamilton (1805-1865). But these fields are not directly relevant to what concerns us in this dissertation.

2.3 Modern Physics

In 1905, Albert Einstein (1879-1955) solved the apparent paradox implied by the photoelectric effect by suggesting that light was composed of discrete *quanta*[6] rather than continuous waves. Based upon the theory of black-body radiation by Max Karl Ernst Ludwig Planck (1858-1947), he conjectured that the energy in each quantum of light was proportional to the frequency multiplied by a constant, later to be known as Planck's constant h. In this theory, a photon with a frequency above a certain threshold has the energy required to eject an electron in the material, as observed in the photo-electric effect. The correctness of this postulate was confirmed in 1914 by Robert Andrews Millikan (1868-1953). Besides earning Einstein and Millikan the Physics Nobel Prize in 1921 and 1923 respectively, the photoelectric effect propelled the concept of the dualistic nature of light, i.e. that light possesses simultaneously the characteristics of both waves and particles, each manifesting itself depending on the circumstances. This complementarity is called the wave-particle duality.

This paved the way to the quantum revolution in physics, and the idea that every particle possesses this dual nature. Indeed, Sir Joseph John Thomson (1856-1940) showed that electrons are particles, and his son, Sir George Paget Thomson (1892-1975), that they also have the properties of waves; they both earned the Physics Nobel Prize, in 1906 and 1937 respectively, for their observations.

[6]The term *photon* was introduced by Gilbert Newton Lewis (1875-1946) only in 1926.

The Principle of Relativity

Galilean relativity was known for a long time. It states that the fundamental laws of physics are the same in all inertial frames, which requires physical laws to be the same whether a body is at rest or moving at constant velocity. This is however not the case for Maxwell's equations if the speed of light is to be the same for every reference frame, a fact experimentally confirmed in 1887 by Albert Abraham Michelson (1852-1931) and Edward Williams Morley (1838-1923). In 1899, Hendrik Antoon Lorentz (1853-1928) discovered that Maxwell's equations are invariant only by a certain change of time and length units. This lead to the development of *special relativity*[7], according to which Galilean relativity is a good approximation only at relative velocities much smaller than the speed of light $c = 299792458$ m/s.

Special relativity only addressed the invariance of Maxwell's equations in the special case of inertial reference frames. This theory prompted Hermann Minkowski (1864-1909) to develop the concept of space-time, in which special relativity, and subsequent theories based upon it, can be nicely represented. In 1916, Einstein developed *general relativity*, which applies the principle of relativity to arbitrary coordinate transformations using the contributions to differential geometry of Georg Friedrich Bernhard Riemann (1826-1866). This theory also includes the effect of gravity.

General relativity is also much related to symmetries of the laws of electromagnetism. For example, Theodor Franz Eduard Kaluza (1885-1954) showed that Maxwell's equations can be derived by extending general relativity into five dimensions. This strategy of using higher dimensions to unify different forces remains an active area of research in particle physics. These extra dimensions may be physically real but curled up and very small, an idea essential to string theory.

[7]In 1905, Einstein re-derived the Lorentz transformation, termed so by Jules Henri Poincaré (1854-1912), from his postulates of special relativity.

Quantum Mechanics

The discovery that the wave-particle duality applies to every particle and not just to light lead to another revolution in physics: *quantum mechanics*. The concept of quanta was already introduced into physics by Max Planck, but the mathematical formulation was derived by Erwin Rudolf Josef Alexander Schrödinger (1887-1961), who introduced the concept of *wave equation*, which describes the evolution of the *wave function*. The wave function encapsulates the probability that the system is to be found in a given state at a given time.

According to the Copenhagen interpretation, devised by Niels Henrik David Bohr (1885-1962) among others, quantum mechanics does not describe an objective reality but deals only with probabilities of measuring a system's properties. Energy quanta fit neither the classical idea of particles nor the classical idea of waves. Also, the act of measurement causes the set of probabilities to immediately and randomly assume only one of the possible values. This leads to the uncertainty principle postulated in 1927 by Werner Karl Heisenberg (1901-1976), which states a fundamental limit on the accuracy with which certain pairs of physical properties of a particle, such as position and momentum, can be simultaneously known to arbitrary accuracy. The uncertainty principle has profound consequences in modern particle physics, since it allows the creation of virtual particles which mediate the interactions in nature.

Another important quantum mechanical principle is that of the *exclusion principle* formulated in 1925 by Wolfgang Ernst Pauli (1900-1958), which states that no two identical fermions (i.e. particles with half-integer spin[8]) may occupy the same quantum state simultaneously. Bosons (i.e. integer spin particles) are not subject to this exclusion principle, and any number of identical bosons can occupy the same quantum state[9].

[8]The spin is the internal angular momentum, a quantum mechanical property which has no equivalent in classical mechanics.
[9]The statistical distribution of indistinguishable fermions (bosons) obeys the Fermi-Dirac (Bose-Einstein) statistics.

Early quantum mechanics only addressed the quantization of classical mechanics. In 1928, Paul Adrien Maurice Dirac (1902-1984) was the first to fully account for relativity in the context of quantum mechanics. The Dirac equation, which describes the behavior of fermions, explains spin as a consequence of the union of quantum mechanics and relativity, and predicts the existence of *antimatter*. The *positron* (antielectron) was experimentally observed in 1932 by Carl David Anderson (1905-1991).

Quantum Field Theory

Dirac founded quantum field theory (QFT) by reinterpreting equation as a many-body equation. QFT nicely accounts for the existence of antimatter and matter-antimatter annihilation. However, early theories had serious theoretical difficulties, as they gave infinite predictions for basic physical quantities such as the self-energy of the electron (which was already a problem in classical electromagnetic field theory).

This problem was solved in the case of quantum electrodynamics (QED) around 1950 by Richard Phillips Feynman (1918-1988), Julian Seymour Schwinger (1918-1994) and Shin'ichirō Tomonaga (**朝永 振一郎**, 1906-1979) through the procedure known as renormalization, which redefines the concepts of charge and mass in interacting physical systems as renormalized quantities. They were awarded the Nobel prize in 1965 in recognition of their work.

Nuclear Physics and Radioactivity

Another area of physics that developed in the 20th century and that is relevant to our purpose is nuclear physics, which starts with the discovery of radioactivity by Antoine Henri Becquerel (1852–1908), Marie Salomea Skłodowska Curie (1867-1934), and Pierre Curie (1859-1906). Radioactivity is the process by which the nucleus of an unstable atom loses energy by spontaneously emitting ionizing particles.

The discovery of radioactivity and that of the electron by Thomson was an indication that the atom had an internal structure.

In 1911, Ernest Rutherford (1871-1937) and his team observed that when helium nuclei were fired at a thin film of gold foil, a few particles were scattered at large angles, and even backwards. This lead to the Rutherford model of the atom, in which the atom has a very small and dense positive nucleus (at that time supposed to be made only of protons) surrounded by electrons. In 1932, Sir James Chadwick (1891-1974) discovers the neutron.

Two aspects related to nuclear physics are of importance to particle physics. First, the study of β decay, during which an electron or a positron is emitted from an atom, lead Enrico Fermi (1901-1954) to describe the weak force in 1930[10] and Pauli to postulate the existence of the neutrino to preserve the conservation of energy, conservation of momentum, and conservation of spin in β decay. Second, the binding mechanism of the atomic nucleus, made of protons and neutrons, was not well understood. Around 1970, a strong attractive force was postulated to explain this binding. Quantum chromodynamics (QCD) describes the strong interaction in the standard model.

2.4 Particle Physics

Although notable, the advent of particle physics was not the only breakthrough of the 20th century. However, it is the one that concerns us in this dissertation.

Since the experiment carried in Rutherford's laboratory, there has been much progress in the design and use of particle accelerators and detectors. Accelerator-based particle physics has allowed gaining a lot of insight on the constituents of matter. The idea is to accelerate subatomic particles (protons, antiprotons, electrons, and positrons, and perhaps one day even muons) to very high energies, i.e. very close to the speed of light in vacuum, and then make them collide either with another beam (of the same or different particles) or with a fixed target. In these setups, the large

[10]This force was later unified with electromagnetism.

amounts of energy are transformed into elementary particles, which are then observed in particle detectors as they decay.

Another area of particle physics, called *astroparticle physics*, studies collisions of very energetic particles (much more energetic than what can be produced in today's accelerators) with atoms from the atmosphere of the Earth. These collisions produce a shower of particles, which are detected. Despite the much larger energies (which allow to probe the universe even closer to its origins), this field suffers from (very) low event rates, and, although in a lesser extent, of the non-reproducibility of the experimental conditions.

In this dissertation, we discuss results obtained from the head-on collision of protons and antiprotons, at a center-of-mass energy of 1.96 TeV.

2.5 Summary

In this chapter, we have presented a very brief history of what lead to modern physics, particularly to the description of the world in terms of *elementary particles* and *interactions*. We presented the main ideas leading to the current understanding of this aspect of nature, introducing their inceptors. But because scientific advancement is a collaborative effort (especially across generations), there are many persons that contributed to this knowledge.

We now proceed to a more formal (and technical) presentation of the theoretical framework underlying high-energy particle physics: the *standard model*.

3. THE STANDARD MODEL OR PARTICLE PHYSICS

3.1 Introduction

The standard model of particle physics is the theory which describes electromagnetic, weak and strong interactions based on a local symmetry of the form:

$$SU(2)_{I_W} \times U(1)_{Y_W} \times SU(3)_C \tag{3.1}$$

where the local (gauge) groups act on the weak isospin and hyper-charge, I_W and Y_W(electroweak interaction), and the color (strong interaction).

Imposing the gauge symmetry leads to the requirement that the W and Z gauge bosons as well as the fermions be massless, since a corresponding mass term in the Lagrangian would violate the gauge symmetry. The way of providing a mass to the gauge bosons while preserving a local gauge symmetry leads to the concept of spontaneous symmetry breaking. This does not rely on mass terms in the Lagrangian but rather on the existence of a scalar field with a specific form of interaction, which gives mass to all elementary particles.[1] Here, the mass of the particles is not a result of emission or absorption of quanta of the scalar field but rather the result of their interaction with the field, which extends over all space.

The Brout-Englert-Higgs (BEH) mechanism [4–6] applies a spontaneous symmetry breaking to the local symmetry. It is considered as the origin of the mass of the gauge bosons and the fermions[2] in the standard model. The existence of the yet undiscovered Higgs scalar would be the physical manifestation of the mechanism.

[1]The mass of composite particles such as protons and neutrons is related to the energy of the composite system and has nothing to do with the mechanism described here.
[2]It is an assumption of the standard model, for besides relating the γ, W, and Z, the Higgs boson has no other observable effect.

Table 3.1: The particles of the standard model. The gravitational interaction is not described by the standard model. When charge is implied, the upper sign refers to the particle, and the other to the antiparticle, which has the same properties except for the opposite charge. Source: [7]

	Particle	Mass (GeV/c^2)	Charge	Interaction
Leptons	electron e	$511 \cdot 10^{-6}$	± 1	EM, Weak
	neutrino ν_e	$< 2.3 \cdot 10^{-6}$	0	Weak
	muon μ	$105.6 \cdot 10^{-3}$	± 1	EM,Weak
	neutrino ν_μ	$< 0.17 \cdot 10^{-3}$	0	Weak
	muon τ	1.776	± 1	EM,Weak
	neutrino ν_τ	$< 15.5 \cdot 10^{-3}$	0	Weak
Quarks	up u	$[1.5 - 3.3] \cdot 10^{-3}$	$\pm 2/3$	Strong, EM, Weak
	down d	$[3.5 - 6.0] \cdot 10^{-3}$	$\mp 1/3$	
	charm c	$1.27^{+0.07}_{-0.11}$	$\pm 2/3$	
	strange s	$[7 - 130] \cdot 10^{-3}$	$\mp 1/3$	
	top t	173.2 ± 0.9	$\pm 2/3$	
	bottom b	$4.20^{+0.17}_{-0.07}$	$\mp 1/3$	
Gauge Bosons	g (gluon)	0	0	Strong
	γ (photon)	0	0	EM
	$W^\pm$ boson	80.399 ± 0.023	± 1	Weak
	Z^0 boson	91.188 ± 0.002	0	

In this chapter, we review formally the concepts behind the standard model. We present a general discussion on quantum gauge theories and proceed with the description of electrodynamics, the theory of weak interactions, electroweak unification, and the strong interaction. The final sections of this chapter review the constraints on the SM Higgs boson, and present prospects of physics beyond the SM.

Table 3.1 lists the particles in the standard model together with some of their properties.

3.2 Quantum Field Theory

Quantum field theory (QFT) provides a theoretical framework for constructing quantum mechanical models of systems parametrized by an infinite number of dynamical degrees of freedom, i.e. *fields*[3]. It is the natural and quantitative language for particle physics. In the SM, the elementary particles and their interactions are formulated as a relativistic quantum field theory [8]. In QFT, particles are regarded as excited states of a field, or *field quanta*. Such a formalism is necessary to allow the description of physics processes during which the particle count may change (which requires the ability to *create* and *destroy* particles).

In perturbative[4] quantum field theory, the forces between particles are mediated by other particles. The electromagnetic force between two electrons is caused by an exchange of *photons*. Intermediate *vector bosons* mediate the weak force and *gluons* mediate the strong force. There is currently no complete quantum theory of gravity, but many proposed theories postulate the existence of a *graviton* that mediates it. These force-carrying particles are virtual particles, i.e. particles that exists only for a limited time and space. The energy and momentum of a virtual particle are uncertain according to the uncertainty principle

$$\Delta E \cdot \Delta t \geq \hbar. \tag{3.2}$$

By definition, they cannot be detected while carrying the force, because such detection will imply that the force is not being carried.

QFT is also used in other areas of physics, e.g., for the description of critical phenomena and quantum phase transitions, such as in the BCS theory of superconductivity [9–11].

[3]A field is a physical quantity associated with each point of spacetime.

[4]Perturbation theories start from simple system for which a mathematical solution is known, and add an additional *perturbing* Hamiltonian representing a weak disturbance to the system. If the disturbance is small, the physical observables of the perturbed system are expressed as 'corrections' to those of the simple system.

3.2.1 The Euler-Lagrange Principle of Least Action

The *action* $\mathcal{S}$ is an attribute of the dynamics of a physical system. Mathematically, it is a functional which operates on the *path* or history of the system to yield a real number (of dimensions energy $\times$ time), it is expressed as

$$\mathcal{S} = \int \mathcal{L}(\phi, \partial_\mu \phi) d^4x \tag{3.3}$$

where $\mathcal{L}$ is the Lagrangian density and is a function of one or more fields $\phi(x)$ and their derivatives $\partial_\mu \phi(x)$.

The *principle of least action* led to the development of the Lagrangian and Hamiltonian formulations of classical mechanics. It states that the path followed by a physical system is that for which the action is stationary (minimal), i.e. for which $\delta\mathcal{S} = 0$. The equations of motions are derived from this principle. Indeed,

$$\begin{aligned} \delta S &= \int d^4x \left\{ \frac{\partial \mathcal{L}}{\partial \phi} \delta\phi + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \delta(\partial_\mu \phi) \right\} \\ &= \int d^4x \left\{ \frac{\partial \mathcal{L}}{\partial \phi} \delta\phi - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) \delta\phi + \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \delta\phi \right) \right\} = 0. \end{aligned} \tag{3.4}$$

The third term of (3.4) can be converted to a surface integral over the boundary of the space-time region. This term vanishes if we assume that $\delta\phi$ is zero at the temporal beginning and end of the region and on the spatial boundary. Since the integral (3.4) must vanish for an arbitrary $\delta\phi$, we obtain the *Euler-Lagrange equation of motion* for a field

$$\frac{\partial \mathcal{L}}{\partial \phi} \delta\phi - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) = 0. \tag{3.5}$$

There is one such equation for each field in the Lagrangian.

The stationary action formulation of classical mechanics extends to quantum mechanics in the Feynman path integral formulation, where a physical system follows simultaneously all possible paths with amplitudes determined by the action [12, 13].

The Lagrangian formulation of field theory is particularly suited to relativistic dynamics because all expressions are explicitly Lorentz invariant.

3.2.2 Gauge Theories

A gauge theory is a field theory in which the Lagrangian is invariant under a continuous group of local transformations. The *gauge* refers to the degrees of freedom in the Lagrangian. The *gauge transformations* form a Lie group, which is the gauge (symmetry) group of the theory. The Lie algebra of group generators is associate with any such group. To each generator corresponds a vector field: the *gauge field*. These fields are included in the Lagrangian to ensure its *gauge invariance* (invariance under group transformations). For quantized theories such as those presented below, the quanta of the gauge fields are termed *gauge bosons*. Groups with a commutative (non-commutative) symmetry are referred to as Abelian (non-Abelian).

Gauge theories possess the important property of being *renormalizable* [8]. In QFT, renormalization is used to address the divergence of many integrals in perturbative calculations, which give infinite answers. These divergences arise from the fact that, in the Lagrangian, quantities such as the charge and mass of the particles do not correspond to the physical observables measured in the laboratory (in un-renormalized theories). The renormalization process consists in re-writing the Lagrangian in terms of observable quantities (renormalized quantities), introducing counter terms that cancel out the divergences. Renormalization restores the predictive power of quantum field theories.

Global Gauge Symmetry

The first theory with a global gauge symmetry was proposed by P.A.M. Dirac in an attempt to describe the equations of motion of a free electron (or generally any particle with electric charge and spin). The Lagrangian of this theory is[5]

$$\mathcal{L}(x) = \bar{\psi}(x)\left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi(x), \tag{3.6}$$

where ψ represents a Dirac field of mass m, γ^{μ} are Dirac's matrices (i.e. four-dimensional matrices obeying Clifford's algebra [8]). The Dirac Lagrangian is invariant under the global $U(1)$ transformation

$$\psi(x) \rightarrow e^{iQ\alpha}\psi(x), \tag{3.7}$$

where Q is the electric charge of the particle and α is a constant. In (3.6), $\bar{\psi} = \psi^{\dagger}\gamma_0^*$ is the adjoint spinor of ψ.

According do Noether's theorem [8,14], every symmetry conserves a certain quantity. In the case of $U(1)$, the current

$$j^{\mu}(x) = -Q\bar{\psi}(x)\gamma^{\mu}\psi(x) \tag{3.8}$$

obeys

$$\partial_{\mu}j^{\mu} = 0. \tag{3.9}$$

Equation (3.8) states that the charge is conserved.

[5]We use the Einstein notation which implies summation over repeated indices. Greek indices go from 0 to 4.

Local Gauge Symmetry

The description of the interaction between particles (through fields) require a space-dependent symmetry, or *local* symmetry. In the case of $U(1)$,

$$\psi(x) \to e^{iQ\alpha(x)}\psi(x) \tag{3.10}$$

where $\alpha(x)$ is a position-dependent field. The operator that keeps this Lagrangian invariant is the *covariant derivative*

$$D_\mu = \partial_\mu + iQA_\mu \tag{3.11}$$

where A_μ is a four-vector field that transforms according to

$$A_\mu \to A_\mu - \partial_\mu \alpha(x). \tag{3.12}$$

From the above, we have

$$D_\mu \psi(x) \to e^{iQ\alpha(x)} D_\mu \psi(x). \tag{3.13}$$

Introducing local gauge symmetry to the Dirac equation (3.6) yields the Lagrangian of *quantum electrodynamics* (QED)

$$\begin{aligned} \mathcal{L}_{\text{QED}} &= \bar{\psi}(x)\left(i\gamma^\mu D_\mu - m\right)\psi(x) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &= \bar{\psi}(x)\left(i\gamma^\mu \partial_\mu - m\right)\psi(x) - Q\bar{\psi}(x)\gamma^\mu\psi(x) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \end{aligned} \tag{3.14}$$

where $F_\mu\nu = \partial_\mu A_\nu - \partial_\nu A_\mu$. This theory describes the electromagnetic interaction, represented by the term $Q\bar{\psi}(x)\gamma^\mu\psi(x)$, where the field ψ interacts through the ex-

change of a massless[6] vector field A_μ. Indeed, applying the principle of least action to obtain the equations of motion (3.5) for (3.14) yields

$$\left(i\gamma^\mu D_\mu - m\right)\psi(x) = Q\gamma^\mu A_\mu \psi(x). \tag{3.15}$$

Thus, the existence and properties of the photon naturally arise from the requirement of local gauge invariance under $U(1)$ transformations. The fact that the mediator of the electromagnetic interaction is massless implies that the latter has an infinite range of action.

3.3 The Theory of Weak Interactions

The theory of weak interactions, developed to explain the properties of β decay by Fermi in 1934 [15], is analogous to QED. In the latter, processes such as electron-proton scattering have an amplitude given by a product of currents:

$$\mathcal{M} = \left(e\bar{\psi}_p\gamma^\mu\psi_p\right)\cdot\frac{-1}{q^2}\cdot\left(-e\bar{\psi}_e\gamma^\mu\psi_e\right) \tag{3.16}$$

where ψ_p refers to the proton, of charge $Q = +e$, and ψ_e to the electron, of charge $Q = -e$. By analogy, Fermi proposed a theory with the following amplitudes for the charged currents of β decays

$$\mathcal{M} = G_F\left(e\bar{\psi}_n\gamma^\mu\psi_p\right)\left(-e\bar{\psi}_{\nu_e}\gamma^\mu\psi_e\right) \tag{3.17}$$

which corresponds to $n \to p + \nu_e + e$ and where $G_F = 1.16637(1)\times 10^{-5}\text{GeV}^{-2}$ is Fermi's constant. This amplitude however, grows infinitely as the energy of the process increases.

The observation of parity violation [3] in the beta decay of $^{27}_{60}$Co, and that it was so maximally, lead to the modification of the theory, and the introduction of the idea that

[6]The inclusion of a mass term $\frac{1}{2}m^2A_\mu A^\mu$ in the Lagrangian (3.14) would violate Lorentz invariance.

the weak interaction affects only left-handed particles, and requires the replacement of γ^μ by $\gamma^\mu(1-\gamma^5)$ in (3.16) [16, 17]. As of the time of writing, all the reported experimental evidence suggests that neutrinos interact with leptons only through the weak force, and therefore are left-handed. This is referred as the $V-A$ (vector-axial) structure of weak interactions. The introduction of this structure in a gauge theory is based on the $SU(2)$ weak isospin group. In 1973, experiments at CERN observed weak neutral currents, through the detection of neutrino scattering [18–20].

3.4 The Electroweak Interaction

The unification of the electromagnetic (QED) and the weak interactions was performed by Glashow in 1961 [21] using the $SU(2)\otimes U(1)$ gauge group. He associated $U(1)$ to the weak hypercharge Y_W and related it to the weak isospin I_W and the electric charge Q through the relation

$$Q = I_{W,3} + \frac{Y_W}{2}. \tag{3.18}$$

This theory thus requires a gauge boson triplet, $(W^1_\mu, W^3_\mu, W^3_\mu)$, which corresponds to the $SU(2)$ group, and a neutral field B_μ, which corresponds to $U(1)$. As for QED, this requires modifying the covariant derivative to

$$D_\mu \equiv \partial_\mu + ig\frac{\tau_a}{2}W^a_\mu + i\frac{g'}{2}Y_W B_\mu. \tag{3.19}$$

where summation over a roman (greek) letter is carried over the space (space-time) coordinates, g and g' are coupling constants, and τ^a are the Pauli matrices.

The Lagrangian for the electroweak interaction is

$$\begin{aligned}\mathcal{L}_{\text{EWK}} = \; & - \; \frac{1}{4}W^{\mu\nu}_a W^a_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \\ & + \; \bar{E}_{Li}\left(i\gamma^\mu D_\mu\right)_{ij} E_{Lj} + \bar{e_{Ri}}\left(i\gamma^\mu D_\mu\right)_{ij} e_{Rj} \\ & + \; \bar{Q}_{Li}\left(i\gamma^\mu D_\mu\right)_{ij} Q_{Lj} + \bar{u_{Ri}}\left(i\gamma^\mu D_\mu\right)_{ij} u_{Rj} + \bar{d}_{Ri}\left(i\gamma^\mu D_\mu\right)_{ij} d_{Rj}\end{aligned} \tag{3.20}$$

where the E_L and Q_L represent the lepton and quark left-handed doublets

$$\begin{aligned} E_L &= \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \text{ or } \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \\ Q_L &= \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \text{ or } \begin{pmatrix} t \\ b \end{pmatrix}_L, \end{aligned}$$

and e_R, u_R and d_R the lepton and quark right-handed singlets, respectively for any lepton family, any up-type quark family and down-type quark family. W_a^μ and B^μ represent the vector gauge fields, which, after mixing between the neutral fields, lead to the physical fields W_μ^+, W_μ^-, Z_μ and A_μ, which respectively mediate the charged (CC) and neutral (NC) current processes, and the electromagnetic interaction. The corresponding particles are respectively the $W^\pm$ and Z bosons, and the photon γ. The mixing occurs as follows:

$$W_\mu^\pm \equiv \sqrt{\frac{1}{2}} \left(W_\mu^1 \mp W_\mu^2 \right) \tag{3.21}$$

$$Z_\mu \equiv -B_\mu \sin(\theta_W) + W_\mu^3 \cos(\theta_W) \tag{3.22}$$

$$A_\mu \equiv B_\mu \cos(\theta_W) + W_\mu^3 \sin(\theta_W) \tag{3.23}$$

$$\tag{3.24}$$

where θ_W is Weinberg's angle and is determined from $g'/g = \tan\theta_W$. In this formalism, the coupling of the electromagnetic current with the photon field is given by the electric charge

$$Q = g' \cos\theta_W = g \cos\theta_W. \tag{3.25}$$

While this theory explains all known electroweak phenomena (except neutrino oscillations [1, 2], and thus their masses), the masses of the vector bosons where introduced by hand, and break the gauge invariance of the theory. To address this issue, several authors introduced the idea of *spontaneous symmetry breaking* (EWSB).

3.4.1 The BEH Mechanism

The Brout-Englert-Higgs (BEH) mechanism[7] [4–6] is the most popular among the EWSB mechanisms; it is implemented in the above picture by introducing a complex scalar $SU(2)_{I_W}$ doublet

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$

and considering the following interaction Lagrangian

$$\mathcal{L}_{\text{BEH}} = (D_\mu \Phi)\,(D^\mu \Phi)^\dagger + V(\Phi) \tag{3.26}$$

where

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \qquad\qquad (\mu^2 > 0^8 \text{ and } \lambda > 0).$$

The potential is bounded from below and has a minima for

$$|\,\langle \Phi \rangle\,| = \frac{1}{\sqrt{2}} v = \sqrt{\frac{\mu^2}{2\lambda}},$$

where v is the vacuum expectation value. By choosing one of the many minima as a true minimum of the energy, the symmetry of the physical system is *spontaneously broken.*

Expanding this field about a particular minimum,

$$\Phi = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$$

where $h(x)$, called the Higgs field, is the only remaining scalar field. This yields

- Mass terms for the gauge bosons;

[7] This mechanism has multiple names. It refers to the contents of three important letters that were published in Phys. Rev. Lett. in 1964 by Higgs, by Brout and Englert, and by Guralnik, Hagan, and Kibble.

[8] A negative value for μ^2 would yield a mass term for Φ, which is not gauge invariant and therefore forbidden.

- A mass term for the neutral scalar boson h (a real field);
- Interaction terms between the scalar boson and the gauge bosons, proportional to the mass of the gauge boson;
- Self interaction terms of higher order for h.

Of course, an additional scalar field leads to the possibility of it being coupled to the fermion fields of the theory. These are considered under $\mathcal{L}_{\text{Yukawa}}$, the Lagrangian describing the Yukawa interaction that generates fermion masses after the Higgs boson acquires a vacuum expectation value. An example of such a term for a $SU(2)$ doublet f_L and a $SU(2)$ singlet f_R is given by

$$g_f \left[\bar{f}_L \Phi f_R + h.c.\right] \rightarrow \frac{g_f v}{\sqrt{2}} \bar{f} f = m_f \bar{f} f \tag{3.27}$$

where g_f is the Yukawa coupling for the fermion f.

However, the SM does not fix the fermion masses from the Higgs potential, nor can these be derived from the BEH mechanism. Additionally, the theory does not predict the mass of the Higgs boson. Yet, this theory is compelling, for it solely relies on symmetry principles.

Experimental Evidence In 1973, the Gargamelle collaboration at CERN performed the first measurement of neutrino-induced weak neutral currents [20]. The measurement of the ratio of NC to CC determined that

$$0.1 < \sin^2 \theta_W < 0.6 \tag{3.28}$$

at 90% C.L. Low-energy phenomena provide the following relation [22]

$$g = 2\sqrt{2 M_W^2 G_F}, \tag{3.29}$$

Table 3.2: The masses (in GeV/c^2) of the W and Z bosons as measured at the UA1 and UA2 experiments, together with the current world average. Source: [7]

	UA1	UA2	Current World Average
M_W	$83.5 \pm 1.1 \pm 2.7$	$80.2 \pm 0.6 \pm 0.5$	80.399 ± 0.023
M_Z	$93.0 \pm 1.4 \pm 3.0$	$91.5 \pm 1.2 \pm 1.7$	91.188 ± 0.002

which together with $Q = g \sin\theta_W$ provides constraints on the mass of the W and Z bosons

$$M_W^2 = \frac{Q^2}{4\sin^2\theta_W} v^2 \sim \left([48 - 118]\ \text{GeV}\right)^2 \tag{3.30}$$

$$M_Z^2 = \frac{M_W}{\cos^2\theta_W} \sim \left([76 - 124]\ \text{GeV}\right)^2 \tag{3.31}$$

In 1983, both the Z [23,24] and the W [25,26] bosons where discovered by the UA1 and UA2 experiments at the CERN SPS $p\bar{p}$ collider, and their masses determined to a few GeV/c^2 precision, in good agreement with the predictions from the SM. Table 3.2 shows the results of both collaborations, together with the current world average [7].

3.5 The Strong Interaction

Quantum Chromodynamics (QCD) is the gauge field theory that describes the *strong interactions* of colored quarks and gluons; it is the $SU(3)$ component of the SM. The dynamics of the quarks and the gluons are controlled by the QCD Lagrangian

$$\begin{aligned} \mathcal{L}_{\text{QCD}} &= \bar{\psi}_i \left(i\gamma^\mu (D_\mu)_{ij} - m\,\delta_{ij}\right)\psi_j - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \\ &= \bar{\psi}_i (i\gamma^\mu \partial_\mu - m)\psi_i - g G^a_\mu \bar{\psi}_i \gamma^\mu T^a_{ij} \psi_j - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \,, \end{aligned} \tag{3.32}$$

where $\psi_i(x)$ and $G^a_\mu(x)$ are respectively the quark and gluon fields in the fundamental representation of the $SU(3)$ gauge group, γ^μ are the Dirac matrices connecting the spinor representation to the vector representation of the Lorentz group, and T^a_{ij} are

the Gell-Mann matrices a representation for the generators of the SU(3) gauge group. $G^a_{\mu\nu}$ represents the gauge invariant gluonic field strength tensor, analogous to the electromagnetic field strength tensor, $F^{\mu\nu}$. It is given by

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g f^{abc} G^b_\mu G^c_\nu \,,$$

where f_{abc} are the structure constants of $SU(3)$ [8].

Neither quarks nor gluons are observed as free particles. Hadrons are color-singlet (i.e. color-neutral) combinations of quarks, anti-quarks, and gluons. The fundamental parameters of QCD are the coupling g_s (or $\alpha_S = g_s^2/4\pi$) and the quark masses m_q.

In the framework of perturbative QCD (pQCD), predictions for observables are expressed in terms of the renormalized coupling $\alpha_S(\mu_R^2)$, a function of an (unphysical) renormalization scale μ_R. For a process with momentum transfer Q, $\alpha_S(\mu_R^2 \simeq Q^2)$ is indicative of the effective strength of the strong interaction. The coupling satisfies the renormalization group equation (RGE)

$$\mu_R^2 \frac{d\alpha_S}{d\mu_R^2} = \beta(\alpha_S) = -(b_0\alpha_S^2 + b_1\alpha_S^3 + b_2\alpha_S^4 + \ldots) \tag{3.33}$$

where b_i is the $(i-1)$-loop beta-function coefficient. $\beta(\alpha_S)$ encodes the *running* of the coupling. The minus sign in (3.33) is the origin of asymptotic freedom, i.e. the fact that the strong coupling becomes weak for processes involving large momentum transfers (hard processes), $\alpha_S \sim 0.1$ for momentum transfers in the $100 - 1000$ GeV range, as shown in figure 3.1. The dependence of the coupling α_S on the energy scale is a property of renormalized theories.

3.5.1 The Top Quark

The top quark is the third generation up-type quark. It was discovered in 1995 by the CDF and DØ experiments at Fermilab [28, 29], in events where it is produced together with an anti-top quark. It has a charge of $+\frac{2}{3}e$. With a mass of

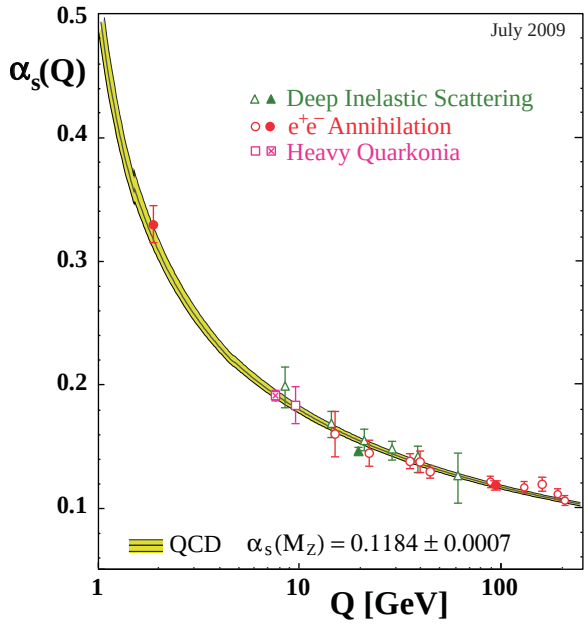


Figure 3.1.: Summary of measurements of α_S as a function of the energy scale Q. The curves are QCD predictions for the combined world average value of $\alpha_S(M_Z)$. Full symbols, open circles, and open triangles are results based respectively on NNNLO, NNLO, and NLO QCD. The cross-filled square is based on lattice QCD. The filled triangle at $Q = 20\ GeV$ is from DIS structure functions. Source: [27]

173.2 ± 0.9 GeV/c^2 [7, 30], it is the most massive elementary particle known to date. It interacts primarily by the strong interaction but can also decay through the weak force. It has a very short lifetime ($\sim 10^{-25}$ s) and hence decays before hadronizing. In 2009, the CDF and DØ experiments observed the production of a single top quark through the weak interaction [31, 32].

The Top Quark Mass and the BEH Mechanism

In the standard model, fermions acquire mass through the BEH mechanism. The top quark is no exception. The Higgs boson has a Yukawa coupling to both the right- and left-handed top quarks. After the electroweak symmetry breaking, the left- and right-handed components mix, leading to a mass term for the top quark:

$$\mathcal{L} = y_t h q u^c \rightarrow \frac{y_t v}{\sqrt{2}}(1 + h^0/v) u u^c$$

where $y_t = \sqrt{2} m_t / v \simeq 1$ is the top quark Yukawa coupling and $v = 246$ GeV is the value of the Higgs vacuum expectation value.

3.5.2 Quark Mixing and the CKM Matrix

In the standard model, the Cabbibo-Kobayashi-Maskawa (CKM) matrix specifies the relationship between the quantum states of quarks when they propagate freely and when they take part in the weak interaction [33]. It is an extension to three families of quarks fo the concept of the Cabibbo angle, introduced to preserve the universality of the weak interaction [34]. It is also an extension of the Glashow-Illiopoulos-Maiani (GIM) mechanism [35], which led to the prediction of the c-quark.

The relationship between free and weak interaction eigenstates can be conveniently represented using a matrix notation:

$$\begin{bmatrix} |d'\rangle \\ |s'\rangle \\ |b'\rangle \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} |d\rangle \\ |s\rangle \\ |b\rangle \end{bmatrix}$$

where the primed notation refers to the weak interaction eigenstates and the unprimed notation to the free eigenstates. The elements of the matrix are related to the relative probability that a up-type quark decays into a down-type quark.

A $N \times N$ matrix has $N(N-1)/2$ real parameters, called Euler angles, and also $(N-1)(N-2)/2$ non-trivial phase angles. The CKM matrix therefore has three Euler angles and one phase angle, δ. The phase will enter the wavefunction as $\exp[i(\omega t+\delta)]$, which is clearly not invariant under time reversal, $t \to -t$ [36].

The CKM matrix must be unitary, i.e. $V^\dagger V = 1$, which implies the weak universality: the sum of all couplings of any of the up-type quarks to all the down-type quarks is the same for all generations. It is a consequence of the fact that all $SU(2)$ doublets couple with the same strength to the vector bosons of the weak interaction. Several parameterizations of the CKM matrix have been proposed. One of them is the so-called standard parameterization [37]:

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13} \end{bmatrix}$$

where the three θ_{ij} are the Euler angles and δ_{13} is the CP-violating phase, and c_{ij} and s_{ij} represent the cosines and sines of the Euler angles, respectively. In this parameterization, θ_{12} is the Cabibbo angle. The currently best known values for the standard parameters are $\theta_{12} = 13.04^\circ \pm 0.05^\circ$, $\theta_{13} = 0.201^\circ \pm 0.011^\circ$, $\theta_{23} = 2.38^\circ \pm 0.06^\circ$

and $\delta_{13} = 1.20^\circ \pm 0.08^\circ$, which correspond to the following modular values of the elements in the matrix [38]:

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97419 \pm 0.00022 & 0.2257 \pm 0.0010 & 0.00359 \pm 0.00016 \\ 0.2256 \pm 0.0010 & 0.97334 \pm 0.00023 & 0.0415^{+0.0010}_{-0.0011} \\ 0.00874^{+0.00026}_{-0.00037} & 0.0407 \pm 0.0010 & 0.999133^{+0.000044}_{-0.000043} \end{bmatrix}.$$

The CKM matrix elements are empirical parameters and must be measured experimentally. They are among many constants of the standard model. At present, there is no understanding of their origin [36].

3.5.3 Parton Distribution Functions

The parton model was proposed by Richard Feynman in 1969 as a way to analyze high-energy hadron collisions [39]. In modern terminology, a parton refers either to a quark or a gluon.

A *parton distribution function* (PDF) is defined as the probability density for finding a particle with a certain longitudinal momentum fraction x at momentum transfer Q^2. Because of the inherent non-perturbative effect in a QCD binding state, PDFs cannot be obtained by pQCD. The PDFs can however be determined from data for deep inelastic lepton-nucleon scattering and for related hard-scattering processes initiated by nucleons. Experimentally determined PDFs sets are available from various groups: CTEQ [40–42], MSTW [43–46], and MRST [46, 47]. Figure 3.2 shows the PDFs for the MSTW2008NLO set.

Deep Inelastic Scattering

High-energy lepton-nucleon scattering, or *deep inelastic scattering* (DIS), plays a key role in determining the partonic structure of the proton [7]. To illustrate the key features of QCD cross sections in processes with initial-state hadrons, let us consider

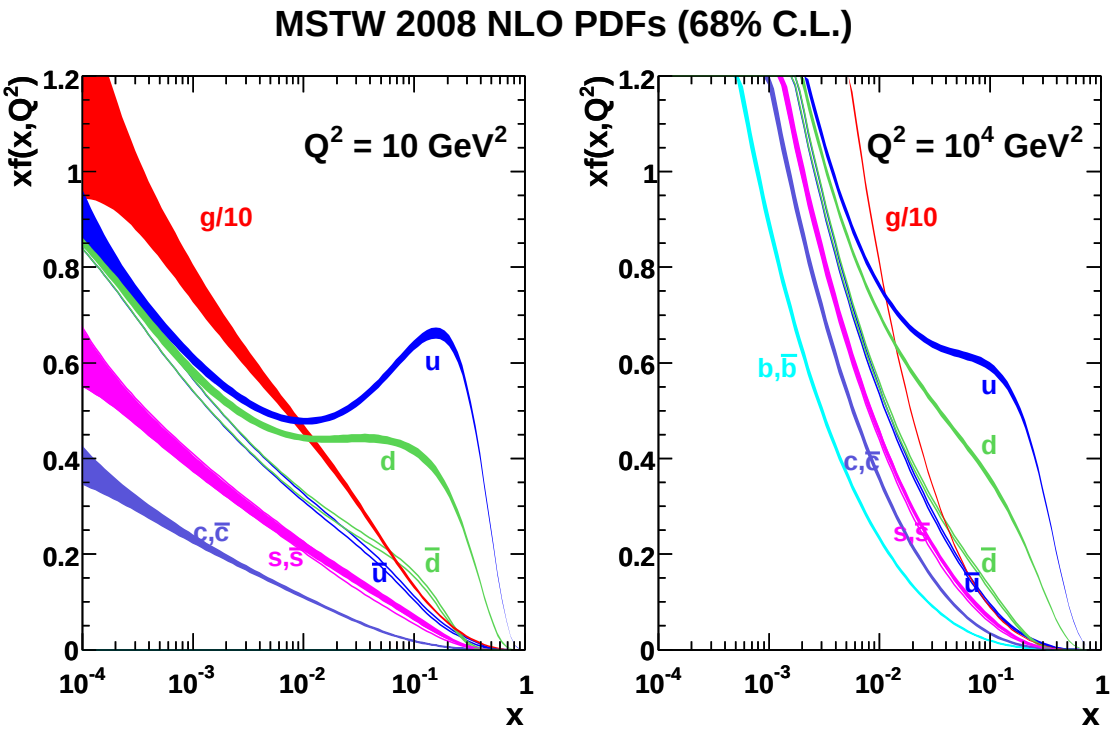


Figure 3.2.: MSTW 2008NLO PDFs at $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2. Source: [45]

$ep \to e + X$, where an electron e with four-momentum k emits a highly off-shell photon (momentum q) that interacts with the proton (momentum p). For photon virtualities $Q^2 \equiv -q^2$ far above the squared proton mass and far below the Z mass, the differential cross section is

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi g}{2xQ^4}\left[(1+(1-y)^2)F_2(x,Q^2) - y^2 F_L(x,Q^2)\right] \tag{3.34}$$

where $x = Q^2/(2p\cdot q)$, $y = (q\cdot p)/(k\cdot p)$, g is the electromagnetic coupling, and $F_2(x,Q^2)$ and $F_L(x,Q^2)$ are proton structure functions which encode the interaction between the photon and the proton. These are not calculable in pQCD, nor is any other cross section that involves initial-state hadrons. To zeroth order in α_S, the structure functions are given directly in terms of non-perturbative PDFs

$$F_2(x,Q^2) = x\sum_q e_q^2 f_{q/p}(x), \qquad F_L(x,Q^2) = 0 \tag{3.35}$$

where $f_{q/p}(x)$ is the PDF for quarks of type q inside the proton, i.e. the number density of quarks of type q inside a fast-moving proton that carry a fraction x of its longitudinal momentum. Since PDFs are non-perturbative, and difficult to calculate in lattice QCD [48], they must be extracted from data.

3.6 Constraints on the Higgs boson

Although the SM successfully predicts almost every observable physical phenomenon[9], the Higgs boson has yet to be observed (as of 2011). It is one of the main challenges of experimental (collider) particle physics. Although further studies will be required to measure its properties, the discovery of the Higgs boson would provide strong evidence that the SM is correct. Departures from the SM will provide hints to new phenomena. However, would there be no Higgs boson (or Higgs-like

[9]Except for several phenomena that are beyond it's scope, since the SM is expected to be a low-energy approximation of a more universal theory.

boson for this matter), we would have to re-think our approach to the equations of nature.

Because the particles are assumed to get their mass by coupling with the Higgs boson, the latter is most easily produced in association with heavy particles. The same argument implies that the largest branching fractions are for the decays to heavy particles (provided that such decay is kinematically allowed). The Higgs boson also couples to $\gamma\gamma$ through a loop of charged particles (such as the W), and to gluons via a quark loop. These considerations are of utmost importance when defining the experimental approach to observing the Higgs.

3.6.1 Theoretical Bounds on the Higgs Boson

The requirement of a stable Higgs potential allows to set lower bounds on m_H; by requiring that the SM should be valid up to the Planck scale ($\Lambda \sim 1.22 \cdot 10^{19}$ GeV), one obtains [49]

$$m_H > 133 + 1.92(M_t - 175) - 4.2 \left(\frac{\alpha_s - 0.12}{0.006} \right). \tag{3.36}$$

Upper bounds can also be obtained by requiring that WW scattering does not violate unitarity, which yields $m_H \lesssim 1$ TeV. Additional constrains are shown in figure 3.3. We redirect the reader to the literature for more details [50, 51]

3.6.2 Experimental Bounds on the Higgs Boson

The experimental constrains on the mass of the Higgs boson come from two sources: fits to precision measurements (indirect constraints) and direct searches.

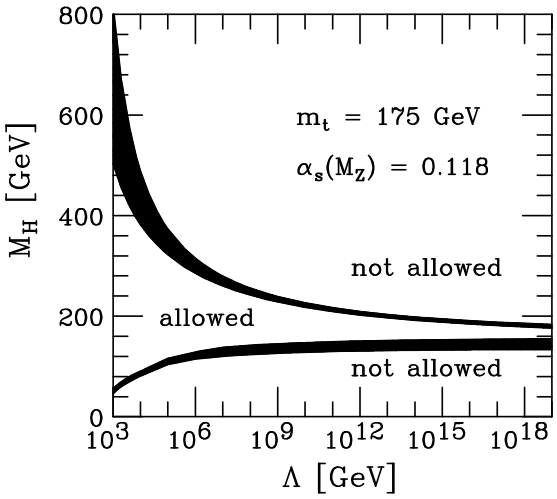


Figure 3.3.: Summary of the uncertainties connected to the bounds on M_H. The upper solid area indicates the sum of theoretical uncertainties on the upper bound for M_H (assuming $M_t = 175$ GeV/c^2). The upper edge corresponds to Higgs masses for which the SM Higgs sector ceases to be meaningful at scale Λ, and the lower edge indicates a value of M_H for which perturbation theory is certainly expected to be reliable at scale Λ. The lower solid area represents the theoretical uncertainties on the lower bounds for M_H derived from stability requirements using $M_t = 175$ GeV/c^2 and $\alpha_S = 0.118$. Source: [51]

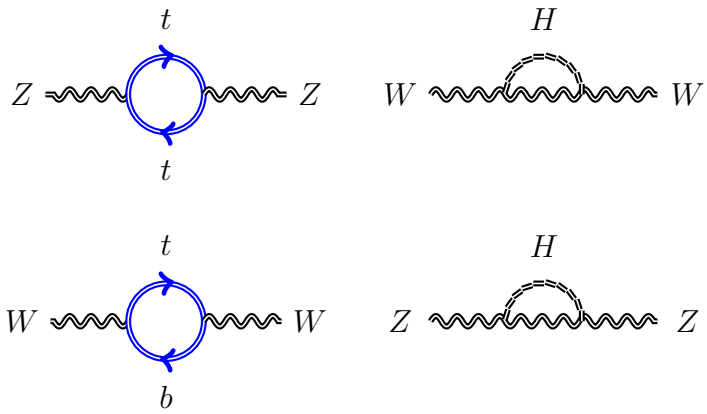


Figure 3.4.: Radiative loop contributions to the mass of SM particles. Precision measurements of the properties of the gauge bosons and the top quark constrain the mass of the Higgs boson. Source: [8]

Indirect Constrains from Precision Measurements

Although the SM does not predict the mass of the Higgs boson, it relates it to electroweak parameters that can be measured experimentally. For instance, radiative corrections (loops) relate m_H to the mass of the heaviest known particles (W, Z, and the top (t) quark). Figure 3.4 shows several of these corrections. The contribution from all such corrections yields [52]

$$M_W^2 = \left(M_Z^2 \cos^2 \theta_W\right) \cdot (1 + \Delta\rho) \tag{3.37}$$

where (to first order)

$$\Delta\rho \equiv \frac{3G_F}{8\sqrt{2}\pi^2} M_t^2 + \frac{\sqrt{2}G_F}{16\pi^2} M_W^2 \left(\frac{11}{3} \ln \frac{m_H^2}{M_W^2} + \ldots\right). \tag{3.38}$$

It is convenient to represent the constrains on m_H as a function of M_W and M_t, which are the most massive particles[10]. This is shown in figure 3.5, which groups together constraints from the LEP-I and SLD experiments (dashed line) and from LEP-II and Tevatron (solid line). The green band shows the allowed range for M_t

[10]The mass of the Z is directly related to that of the W, as shown by (3.31).

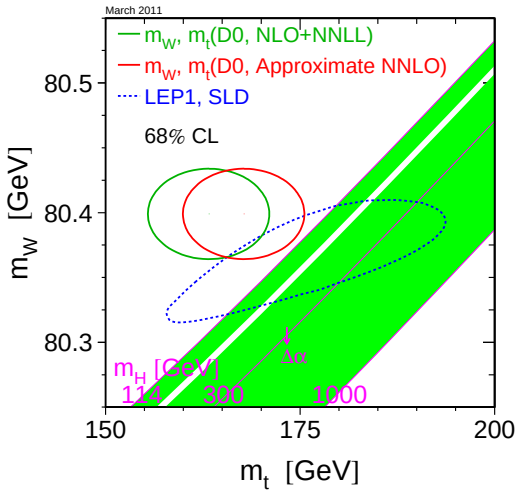


Figure 3.5.: The constraints on m_H as a function of m_W and m_t. The results are from the LEP-I and SLD experiments (dashed line) and from LEP-II and Tevatron (solid line). The green band shows the allowed range for m_t and m_W as a function of m_H. $\Delta\alpha$ is the variation when $\alpha(M_Z)$ is changed by one standard deviation. Source: [53]

and M_W as a function of m_H. $\Delta\alpha$ is the variation when $\alpha(M_Z)$ is changed by one standard deviation, and provides an additional uncertainty. Another representation is shown in figure 3.6, where the SM is fitted to data, and the quality of the fit χ^2 is shown as a function of m_H. This fit is performed by the LEP Electroweak Working Group [53] using $M_t = 173.2 \pm 0.90$ GeV/c^2 [30] and $M_W = 80.399 \pm 0.023$ GeV/c^2. and yields

$$m_H = 92^{+34}_{-26} \text{ GeV}/c^2 \tag{3.39}$$

with a 95% C.L. upper limit of $m_H = 161$ GeV/c^2. Including the results of direct searches at LEP (next section), this limit increases to $m_H = 185$ GeV/c^2.

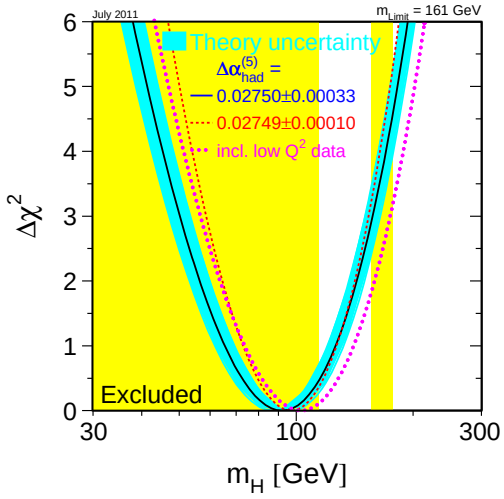


Figure 3.6.: Quality of the fit to electroweak precision data versus the Higgs boson mass. The mass range excluded by direct searches at LEP and the Tevatron is also shown. The solid blue line is the nominal fit; the theoretical uncertainties on the fit are represented by the light blue band. The dashed and dotted curves represent the results of the fit with different input parameters. Source: [53]

Direct Searches at LEP

From 1989 to 2000, the Large Electron-Positron Collider (LEP) smashed electrons and positrons at a center-of-mass energy of $\sqrt{s}$ between 189 and 209 GeV[11]. The OPAL, ALEPH, L3 and DELPHI experiments each collected (and analyzed) a total integrated luminosity of 2.4 fb^{-1}. Given the initial state at LEP, these experiments looked for $e^+e^- \to ZH$, with $H \to b\bar{b}$ and $Z \to \ell\ell$ or $Z \to \nu\nu$, or with $H \to \tau\tau$ and $Z \to q\bar{q}$. Their results were combined by the LEP Electroweak Working Group [53], the outcome of which is shown in figure 3.7. The Higgs boson was excluded for a mass less than 114 GeV/c^2 at 95% C.L.

Direct Searches at the Tevatron

The search for the SM Higgs boson has been a priority of the Tevatron physics program, both at CDF and DØ, the two experiments collecting data at the Tevatron (cf. chapter 4).

There are various SM processes through which a Higgs boson can be produced, as shown in figure 3.8a. The most abundant of these is gluon-gluon fusion ($gg \to H$), with a cross-section between 1.0 to 0.2 pb for $100 < m_H < 200$ GeV/c^2. Next comes associated production, $p\bar{p} \to VH$, where the Higgs is produced in association with a vector boson V (W or Z); this process has a cross-section about five times smaller than gluon fusion. Other modes, such as vector boson fusion, VBF, and the production in association with a top pair, $p\bar{p} \to t\bar{t}H$ or $gg \to t\bar{t}H$, contribute with much smaller cross-sections.

The SM predicts the Higgs boson decay modes as shown in figure 3.8c. Together with the direct constraints from LEP and those from indirect fits to precision measurements, which yield a 95% C.L. interval of $114 < m_H < 185$ GeV/c^2, the decay modes determine the search strategy at the Tevatron. Indeed, for $m_H < 135$ GeV (low mass), the dominant decay mode is $H \to b\bar{b}$, while $H \to WW$ dominates above

[11] These values were determined based on the then available information about the SM.

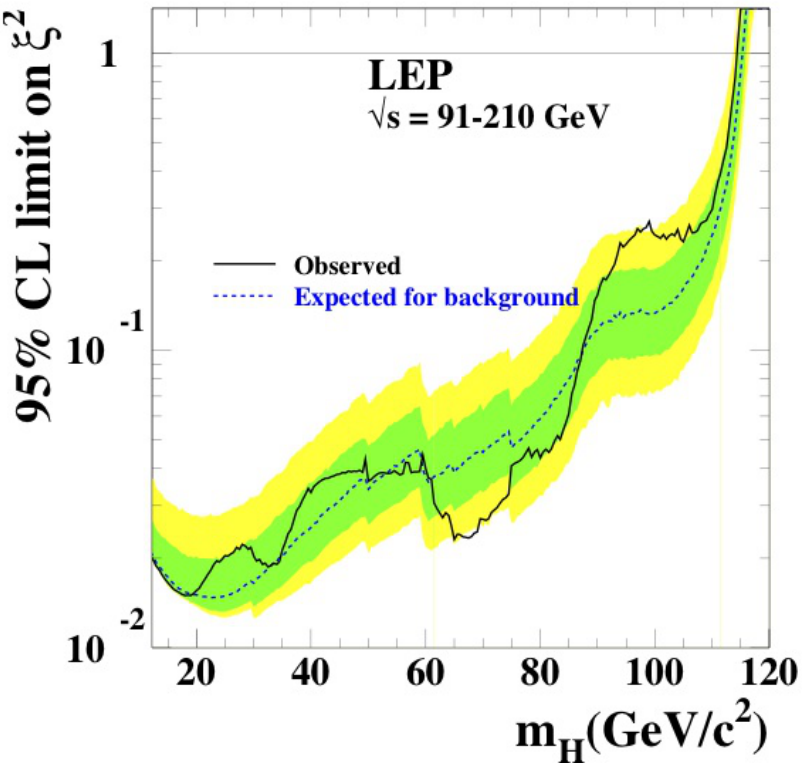


Figure 3.7.: The 95% confidence level upper bound on the ratio ξ^2 of the $H \to ZZ$ coupling to the corresponding SM prediction. The solid line indicates the observed limit, and the dashed line indicates the median limit expected in the absence of a Higgs boson signal. The dark and light shaded bands around the expected limit line correspond to the 68% and 95% probability bands, indicating the range of statistical fluctuations of the expected outcomes. The horizontal line corresponds to the Standard Model coupling. Standard Model Higgs boson decay branching fractions are assumed. Source: [7, 54]

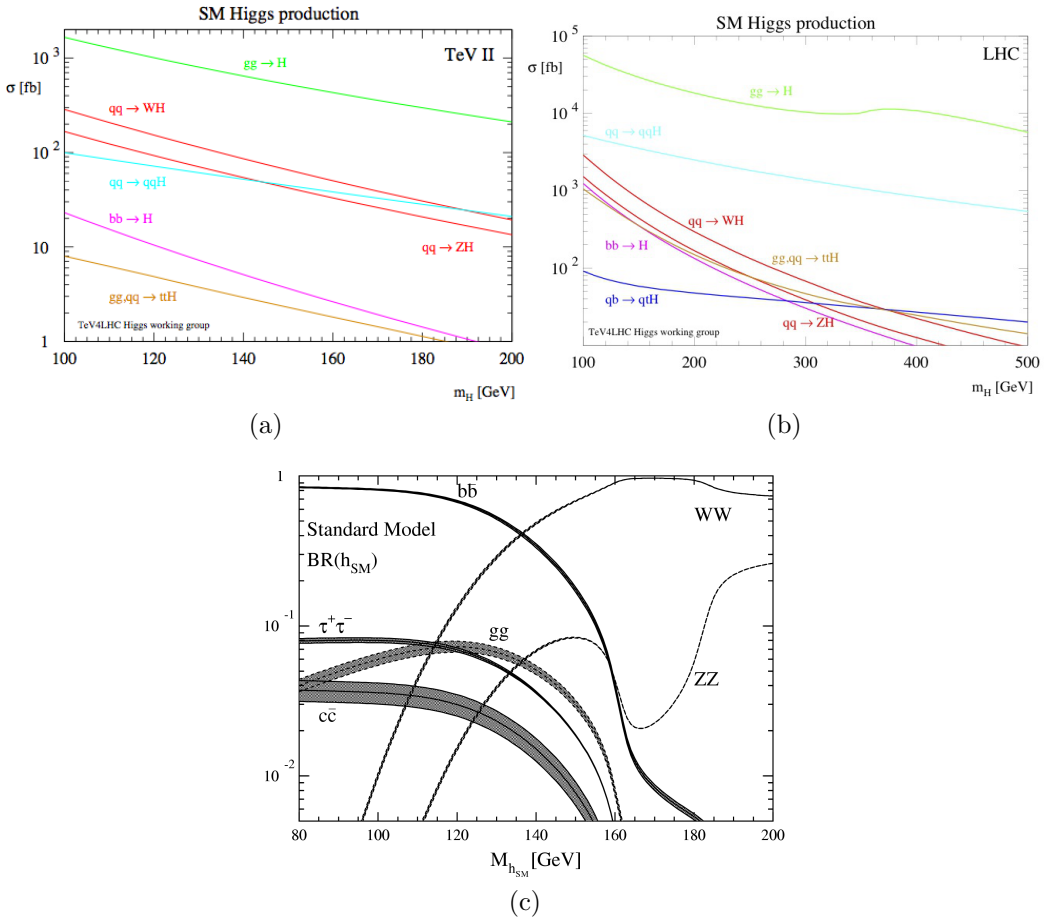


Figure 3.8.: Higgs production cross-sections at (a) the Tevatron ($\sqrt{s} = 1.96$ TeV), and (b) the LHC ($\sqrt{s} = 7$ TeV); (c) Branching ratios for the main decay modes of the SM Higgs boson. Source: [7]

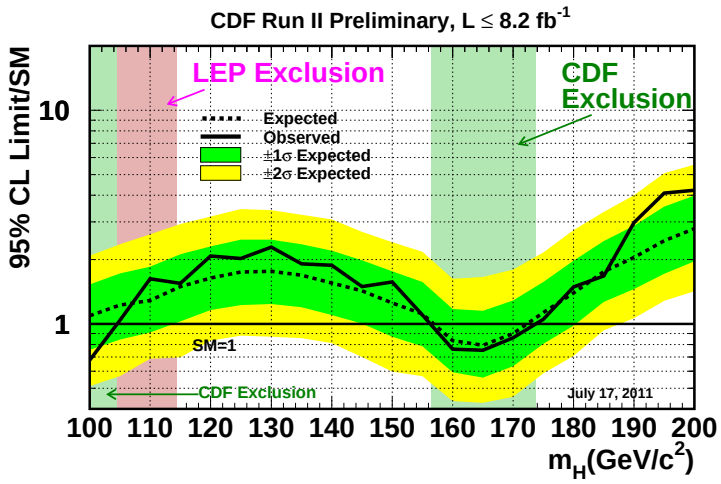
135 GeV/c^2. Because QCD production is very abundant, it is necessary to identify a striking signature of the signal[12]. For this purpose, analysts use the W or Z boson, either through their decay to leptons (more sensitive) or to quarks; these decays are triggered upon to collect events.

The low mass searches at the Tevatron focus on the associated production, VH, with a final state of a $b\bar{b}$ pair (two b-jets), and either zero ($\not\!\!E_T b\bar{b}$), one ($\ell\nu b\bar{b}$) or two ($\ell\ell b\bar{b}$) leptons. We mention that, because of the hadronic τ decays and reconstruction inefficiencies, the two first signatures are sensitive to both $WH \to \ell\nu b\bar{b}$ and $ZH \to \ell\ell b\bar{b}$, although to a different extent. The zero and one lepton channels have similar sensitivity; it is larger than that of the two lepton channel, which is purer but has much smaller acceptance. Gluon fusion can be probed at low mass through the decay to a pair of τ leptons, i.e. $gg \to H \to \tau\tau$, but this is much less sensitive than the other low mass searches.

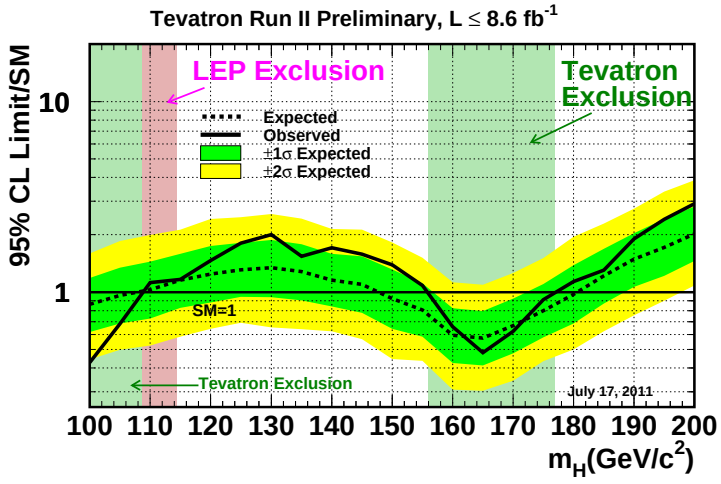
At high mass, the golden channel is $gg \to H \to WW$, where the reconstruction of both W's significantly reduces the QCD and electroweak backgrounds. This search is the single most sensitive at high mass.

The approach followed by CDF and DØ is to separate the different channels according to the final state, perform a dedicated analysis of each channel, and *combine* the results, in order to maximize sensitivity; this is especially true for the low mass searches. In a first step, each experiment combines its results, as shown for CDF in figure 3.9a for the Summer 2011 results, and then all channels from both experiments are combined, as shown in figure 3.9b, again for the Summer 2011 results. Figure 3.10 shows the achieved limits in the search for the Higgs at CDF by date, compared to projections made in 2007.

[12]For instance, $gg \to H \to b\bar{b}$ is impossible to distinguish from the large backgrounds, which are several orders of magnitude larger.

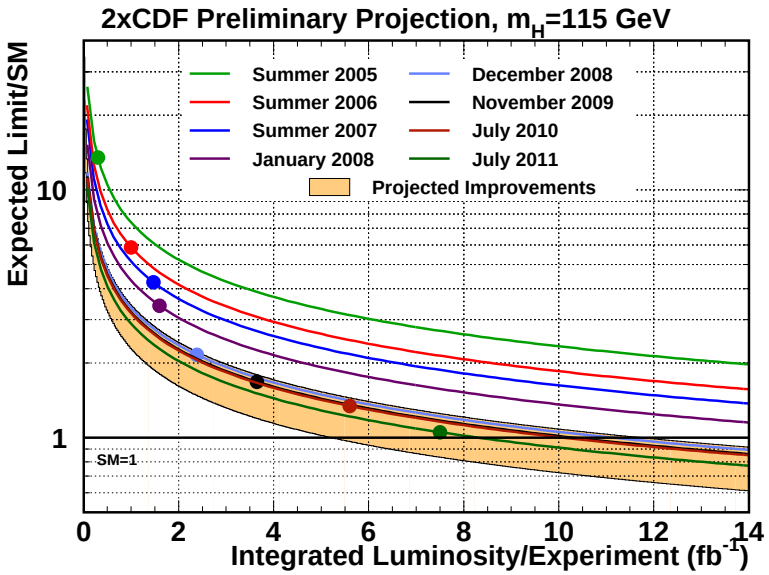


(a)

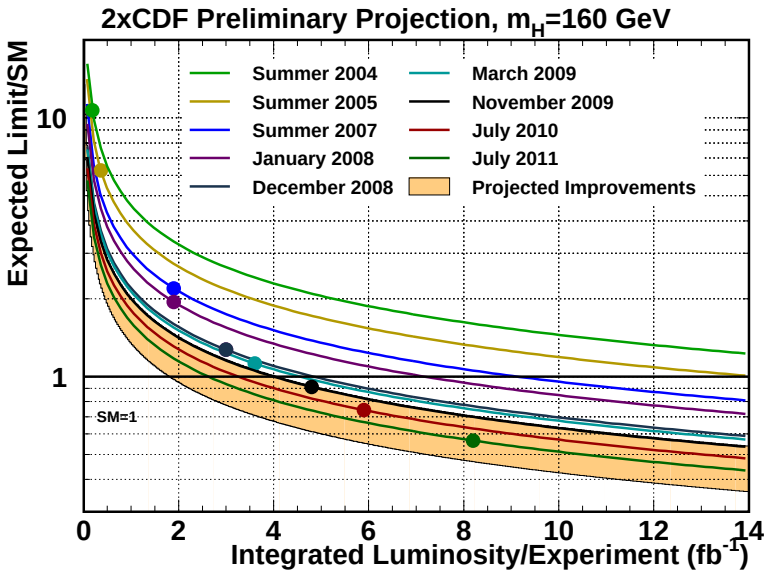


(b)

Figure 3.9.: Summer 2011 combination of SM Higgs searches at the Tevatron: (a) CDF-only, and (b) Tevatron (CDF+DØ). Source: [55]



(a)



(b)

Figure 3.10.: Achieved and projected median expected upper limits on the SM Higgs boson cross section, by date. The solid lines are $1/\sqrt{\mathcal{L}}$ projections. The top of the orange band corresponds to the Summer 2007 performance expected limit divided by 1.5, and the bottom of the orange band corresponds to the Summer 2007 performance expected limit divided by 2.25. The luminosity for the July 2011 point is 8.2 fb^{-1}, a sensitivity-weighted average of the contributing channels' analyzed luminosities. These plots are shown for (a) $m_H = 115$ GeV/c^2 and (b) $m_H = 160$ GeV/c^2. Source: [56]

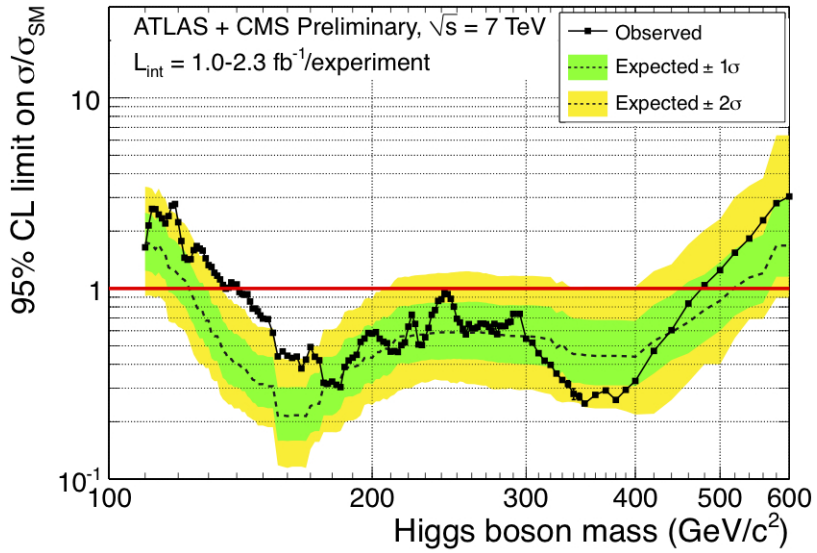


Figure 3.11.: The combined 95% C.L. upper limits on the signal strength modifier $\mu = \sigma/\sigma_{SM}$, obtained with the CLs method from using 1.0 to 2.3 fb^{-1} (per experiment) of CMS and ATLAS data, as a function of the SM Higgs boson mass in the range $110 - 600$ GeV/c^2. The observed limits are shown by solid symbols. The dashed line indicates the median expected $\mu^{95\%}$ value for the background-only hypothesis, while the green (yellow) band indicates the range expected to contain 68% (95%) of all observed limit excursions from the median. Source: [57]

Direct Searches at the LHC

Since March 2010, the LHC experiments have been actively looking of the SM Higgs boson, progressing at very quick pace, thanks to the excellent performance of the accelerator complex and the experimental setups. Although the production mechanisms are quite similar, the cross-sections involved are very different, as can be seen in figure 3.8b. We briefly mention the latest results from the CMS and ATLAS collaborations using 1.0 to 2.3 fb^{-1} per experiment [57]. Figure 3.11 shows the combined 95% C.L. upper limits on σ/σ_{SM} as a function of the Higgs boson mass, in the range $110 < m_H < 600$ GeV/c^2.

3.7 Physics Beyond the Standard Model

Even if we keep aside the description of the gravitational interaction, the observation of neutrino oscillations [1, 2] means that the SM is incomplete. In addition, Cosmic Microwave Background (CMB) [58] observations and other cosmological experimental data have shown that ordinary matter, as described by the SM, constitutes only 4% of the content of the Universe; the rest being attributed to dark matter (DM, 22%) and dark energy (76%), two terms expressing that we know very little about them.

There are currently many candidate theories to explain effects not predicted by the standard model. At the time of writing, no experimental evidence favors a particular theory.

3.8 Summary

In this chapter, we formally discussed the standard model of particle physics, and presented part of the experimental evidence supporting this model. This model, however, relies on the existence of a particle, the Higgs boson, that has yet to be observed. We reviewed the indirect constraints on the Higgs from precision electroweak and discussed direct searches at LEP, the Tevatron and the LHC. This thesis is about the search for the Higgs in the $\not\!\!E_T$+b-jets signature.

4. EXPERIMENTAL ENVIRONMENT

The analyses presented in this thesis were performed using data from the Collider Detector at Fermilab (CDF). This detector is located at one of the two interaction points (where the collisions take place) of the Tevatron accelerator. These facilities, among many others, are hosted at Fermilab, a US particle physics laboratory. In this chapter we describe the experimental setup providing the data we analyze. We first start with the description of the accelerator complex and proceed with that of CDF.

4.1 The Fermi National Accelerator Laboratory

Fermilab, or Fermi National Accelerator Laboratory (FNAL), located in Batavia, Illinois, USA (40 miles west of Chicago), was founded in 1967 to advance "the understanding of the fundamental nature of matter and energy by providing leadership and resources for qualified researchers to conduct basic research at the frontiers of high energy physics and related disciplines" [59]. As of 2011, Fermilab is the sole US national laboratory dedicated to particle physics.

Fermilab hosts a multi-stage accelerator complex, providing particle beams to many physics experiments both on and off site. The Tevatron[1] is the most powerful accelerator of this complex and provides beams to two experiments: CDF and DØ. In addition to high energy collider physics, Fermilab also hosts several smaller fixed-target and neutrino experiments. The beams required for these experiments are obtained by extracting protons from the main injector and colliding them with fixed targets, thereby producing secondary meson, muon and neutrino beams. The neutrino beam, NuMI (Neutrinos at Main Injector) [60], is used by several experiments.

[1]Although the Tevatron is no longer operating, we will refer to it at the present tense, for it has not yet been dismantled.

Two major components of the standard model of fundamental particles and forces were discovered at Fermilab: the bottom quark (May-June 1977) and the top quark (February 1995). In July 2000, Fermilab experimenters announced the first direct observation of the tau neutrino, the last fundamental particle to be observed [61].

4.2 The Tevatron, the Fermilab Accelerator Complex

The Tevatron is a superconducting proton-antiproton ($p\bar{p}$) synchrotron hosted at Fermilab with a circumference of 6.3 km. It produces $p\bar{p}$ collisions at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV. The production of the two beams is the result of a complex chain of accelerators, involving multiple stages spanning from the production of protons and antiprotons to their collision at the B0 and D0 interaction points, where the CDF and DØ experiments are respectively located. In between, there are multiple transfers to various storage and acceleration systems, two key aspects of any accelerator chain. Figure 4.1 shows a schematic view of the Fermilab accelerator complex, which we describe in the following. After a brief history of the Tevatron, we will describe the steps leading to the collision of the proton and anti-proton beams, leaving out other aspects of the complex.

4.2.1 A Brief History of the Tevatron

The Tevatron is the second highest energy particle collider in the world, after the Large Hadron Collider (LHC) at CERN[2]. The first $p\bar{p}$ collisions were produced in 1986. Since then, several upgrades were performed, leading to much improved overall performances. Run I lasted from 1992 to 1996, with a center-of-mass energy of 1.8 TeV.

[2]Until 2010, it was the most energetic particle collider build by mankind.

From 1997 to 2001, the accelerator complex underwent major upgrades aimed at increasing the energy and the luminosity, and at gathering large amounts of data[3]. The upgraded machine accelerates 36 bunches of protons and antiprotons, which is a factor of six more than the 6 bunches it accelerated before the upgrade. As a consequence, the time between two bunch crossings was reduced from 3.5 μs to 396 ns. Since 2001, the Tevatron has been providing collisions at a center-of-mass energy of 1.96 TeV. Run II ended on September 30, 2011, delivering a total of about 12 fb^{-1} to each experiment (CDF and DØ).

The proton beam originates in the Cockcroft-Walton pre-accelerator from which it proceeds through the Linac (linear accelerator) to the Booster Ring. The beam's energy is increased with each revolution before it is injected into the underground oval Main Injector where the energy is again raised until it reaches 120 GeV. Finally, the protons are passed to the Tevatron where they are accelerated to the energy of 980 GeV.

4.2.2 The Proton Source

The protons used in the Tevatron are extracted from a bottle of very pure hydrogen gas (at room temperature, with an energy of 0.04 eV). The gas is moved between two electrodes and a spark ionizes the hydrogen, yielding electrons and H^+ ions. Contrary to what one would assume *a priori*, the produced protons are not yet accelerated; they are first transformed to H^- ions. This step is necessary for multi-turn injection at a later stage of the acceleration process (section 4.2.5). The H^- ions are produced when the protons strike a cathode made of cesium – a very reactive material, that looses electrons easily – and occasionally pick up two electrons. These ions form a continuous beam of approximately 25 keV, which is extracted to the pre-accelerator[4].

[3]Back then, the goal was set to about 2 fb^{-1} of data; the final dataset for each Tevatron experiments will be of about 10 fb^{-1}. This confirms the excellent performance of both the accelerator and the detectors, which performed well beyond expectations.
[4]The motivation for this choice is presented in section 4.2.5.

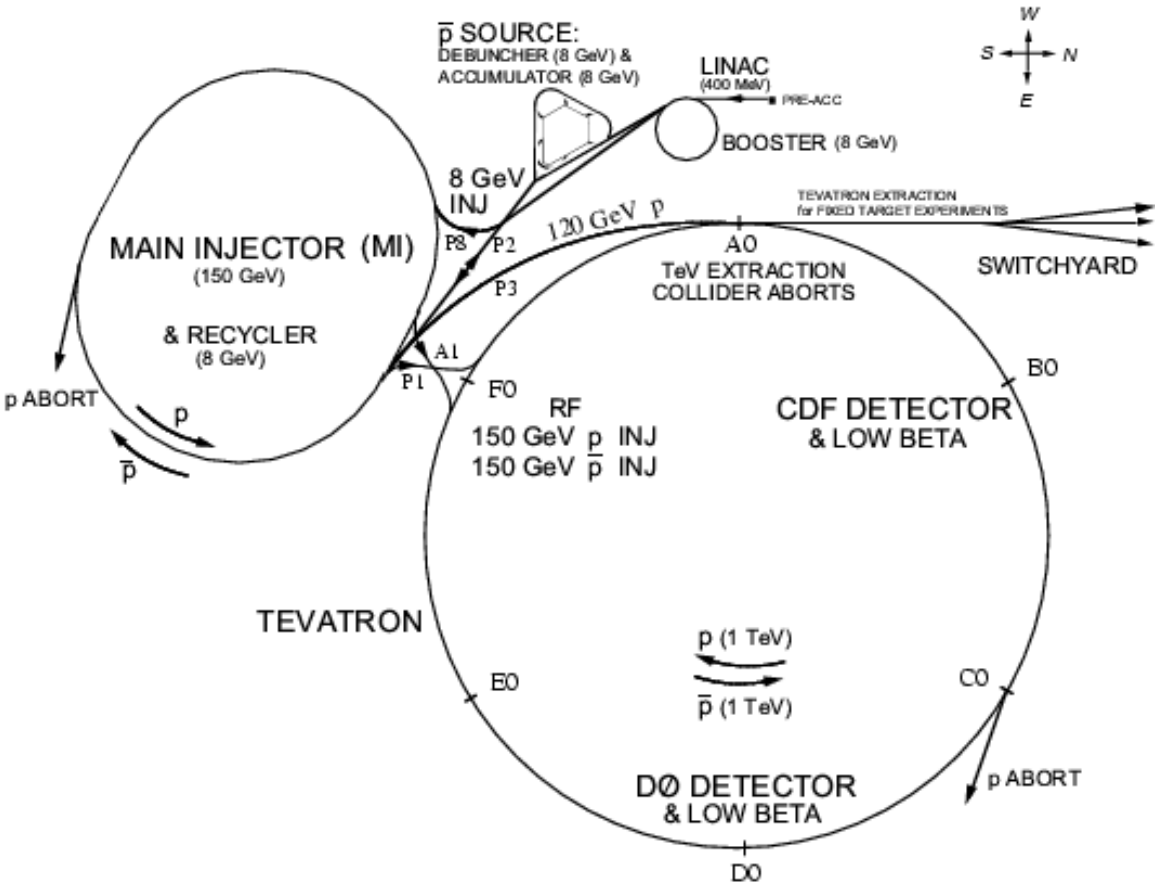


Figure 4.1.: A schematic view of the Tevatron accelerator complex.

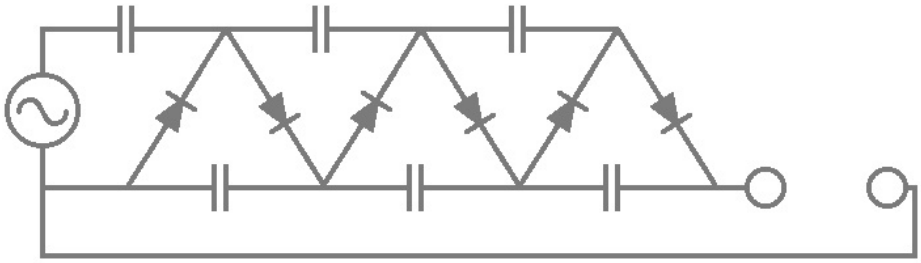


Figure 4.2.: Schematic view of a Cockcrof-Walton static accelerator.

4.2.3 The Pre-accelerator

The pre-accelerator is a Cockcroft-Walton (CW) [62] electrostatic generator. It applies a strong constant electric field that accelerates the H^- ion beam from the proton source to a final energy of 750 keV. This value is determined to be below the dielectric field strength of air, to avoid electric sparks. In a CW, the high voltages are generated from an alternative current (AC) power source using a voltage multiplier ladder network of capacitors and diodes (figure 4.2). Accelerated H^- ions are produced at a rate of 15 Hz (every 66 ms), and then focused and steered to a linear accelerator by magnets.

4.2.4 The Linac

The pre-accelerated H^- ions enter the Linac [63], a 130-meter-long linear accelerator, where they are accelerated to an energy of 400 MeV.

The Linac is made of radio-frequency (RF) cavities which provide time-varying electric fields for acceleration. Its first part consists of five Alvarez-style drift tube cavities (part of the original Linac from 1971) with a resonant frequency of 200 MHz; it accelerates the ions to 116 MeV. The purpose of the drift tubes is to shield the particles from the electromagnetic fields in the cavity; the particles are thus free to drift at a constant velocity. In the space between the tubes, the *accelerating gap*, the particles feel the electromagnetic field and are accelerated. As the ions accelerate, the size of the cavities and the gaps must be increased to provide a constant acceleration

(for the ions travel at higher velocity as they move along the accelerator). The second part of the LINAC, at the high-energy end, is made of seven side-coupled cavities, without drift tubes inside. Instead, couplers lying on top of the structure, rather than in the beam path, couple power from one resonant cavity to the next; the RF signals are amplified by klystrons (instead of traditional power tubes), and brought to the cavities through waveguides. The high-energy end of the LINAC operates at 805 MHz, and accelerates the ions to 400 MeV. The average accelerating gradient throughout the LINAC is 3MV/m.

Acceleration using RF cavities is the method of acceleration used for all subsequent accelerators of the Fermilab complex. The LINAC processes ions every 66 ms, when they arrive from the CW. The alternating electric field produced in the RF cavities groups the ions into bunches, which are sent to the next system in the chain. A typical LINAC pulse is made of $4,000$ bunches of 1.5×10^9 particles and is 20 ms long; the maximum duration is 50 ms ($10,000$ bunches). At this stage the beam power is about 18 MW.

4.2.5 The Booster

The the third stage of acceleration at Fermilab is a 475-meter circumference synchrotron called the Booster [64]. It is the first circular accelerator of the chain, since it would require prohibitively long – and expensive – linear accelerators to accelerate the H^- ions beyond 400 MeV[5].

The injected beam, with an energy of 400 MeV (and $\beta = v/c = 0.719$), takes about 2.2 ms to travel around the Booster. The typical LINAC pulse length, however, is 20 ms. Because of this significant time difference, the injection is a multi-turn process. In order to merge the incoming ion beam with the proton beam in the Booster, these had to be of opposite charge; this justifies the choice of accelerating

[5]A circular accelerator has the advantage that the accelerating cavities can accelerate the same particles more than once.

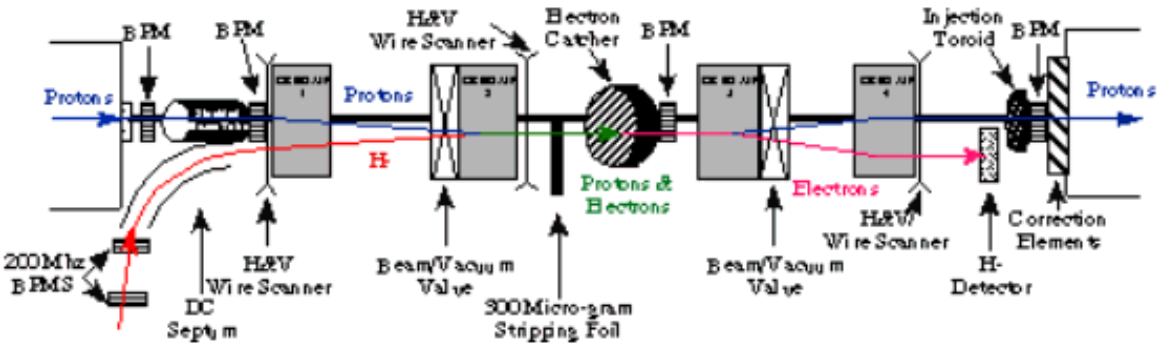


Figure 4.3.: Schematic drawing of the injection chain of the Booster. Credit: The Accelerator Division of Fermilab

H^- instead of H^+ in the earlier steps of the chain, which seemed odd at first. A schematic drawing of the injection process is shown on Figure 4.3.

The Booster is composed of a series of bending magnets, and has 18 accelerating RF cavities. After the injection phase is complete, the beam is accelerated, gaining 500 keV of energy at each turn. To keep a circular trajectory of constant radius, the magnetic field rises with the energy of the protons. After about 16,000 turns, the beam has acquired an energy of 8 GeV and is extracted from the Booster by a series of fast kicker magnets. The booster thus also operates in batches, at a rate of 15 Hz. The output batch contains 84 bunches of 6×10^{10} protons spaced by 18.9 ns. Part of the beam is sent to to the Mini Booster Neutrino Experiment (MiniBooNE)[6] [65]; the rest is sent to the Main Injector via a transfer line.

4.2.6 The Main Injector

The Main Injector (MI) [66] is a 3-kilometer-long circular synchrotron. It plays a central role among the accelerators at Fermilab, dispatching particles between the other facilities. The MI can accelerate or decelerate particles to energies anywhere between 8 and 150 GeV. It can operate in several modes.

[6]MiniBooNE is the first phase of the Booster Neutrino Experiment (BooNE); in this phase, neutrino oscillation measurements will be made with a single detector. If oscillations are observed, then MiniBooNE will be upgraded to stage two (BooNE) with a two-detector configuration.

Antiproton production mode The simplest mode is used to provide beam for antiproton production. In this mode, one batch of 8 GeV protons from the Booster is accelerated to 120 GeV. The batch is then extracted and sent to a nickel target to produce 8 GeV antiprotons (section 4.2.7). Recently, a new procedure, *slip stacking*, allows to merge two batches from the Booster in order to double the number of protons sent to the target, increasing the number of antiprotons produced [67].

Proton and antiproton acceleration mode The MI can accelerate protons from the Booster and antiprotons from the Recycler up to an energy of 150 GeV. The beams are then transferred to the Tevatron or sent to fixed target experiments [60].

Antiproton deceleration mode This mode is used to recycle the antiprotons after the end of a Tevatron collision run. This recovery allows to significantly reduce the amount of time needed to obtain a full store of antiprotons, since less of them need to be produced to fill the store (cf. section 4.2.7).

Tevatron collision mode This mode is the one in which the MI operates to provide protons and antiprotons to the Tevatron; it is much more complex than the previous ones, since on top of the production of antiprotons, the MI has to accelerate protons and antiprotons to an energy of 150 GeV, and to re-group the particles in bunches, a process known as *coalescing*. It is only then that they can be transfered to the Tevatron, for a final acceleration stage. A shot is defined as the steps needed to fill the Tevatron with 36 bunches of protons and 36 bunches of antiprotons at an energy of 150 GeV, and includes the following steps [68]:

1. One batch of 84 proton bunches arrives from the Booster with an energy of 8 GeV. Seven of these bunches are sent to the MI to be accelerated to 150 GeV.

2. At *flattop*, the bunches are coalesced to form a narrow, high density bunch; one of these in injected to the Tevatron.

3. Steps 1 and 2 are repeated 36 times.

4. Meanwhile, the 8 GeV antiprotons have been stored in the Accumulator (cf. section 4.2.7, waiting to be injected in the MI, in the direction opposite to that of the protons.

5. After injection, the antiprotons are accelerated to 150 GeV, and coalesced at *flattop*. They are then injected in the Tevatron, again in the direction opposite to that of the protons.

6. Step 5 is repeated with a new antiproton bunch, until 36 bunches have been delivered to the Tevatron.

The shot setup, which takes about one hour on average, is followed by the acceleration of the beam to 980 GeV (cf. section 4.2.8).

4.2.7 The Production of Antiprotons

The available number of antiprotons is an important limiting factor to producing the high luminosity wanted for collisions inside the Tevatron; producing antiprotons is indeed very time consuming.

The Fermilab antiproton source [69] consists in a target station, two rings – the Debuncher and the Accumulator – and the necessary transfer lines. To produce antiprotons, one batch of protons ($4 - 5 \times 10^{12}$ of them) is accelerated in the MI to an energy of 120 GeV.

Antiproton production After its extraction from the MI, the proton beam is focused (squeezed) using quadrupole magnets, and is directed onto a nickel target. The collisions produce a wide range of secondary particles, including antiprotons. After the target, a lithium collecting lens collects the negatively charged particles using a magnetic field produced by a solenoid. A pulsed dipole magnet follows next; its purpose is to deflect 8 GeV negative particles and send them to the Debuncher. Particles with a wrong charge to mass ratio are collected by a beam dump.

Antiproton cooling At this stage, we obtain about 1 to 2 antiprotons per 10^5 protons hitting the target. In the Debuncher (a 505 m long triangular-shaped synchrotron) , the momentum spread of the antiprotons in reduced using stochastic and momentum cooling [70, 71]. This step significantly improves the transfer efficiency from the Debuncher to the Accumulator, which has a limited momentum aperture at injection.

Antiproton storage Before the next antiproton pulse from the target, the Debuncher is emptied and its contents injected into the Accumulator for medium-term storage. The Accumulator, also a triangular-shaped synchrotron, is slightly smaller (474 m) than the Debuncher, and hosted in the same tunnel. When enough antiprotons have ben accumulated, they are transferred either to the MI or the Recycler. It takes up to one day to build up the stack of $3 - 4 \times 10^{12}$ antiprotons provided to the Tevatron for collisions.

The Recycler The Recycler [72] is a 3.3-kilometer-long storage ring located on top of the MI. Since it is made of permanent magnets and quadrupoles, it operates at a fixed energy of 8 GeV. It was designed to fulfill three missions: (a) it allows antiprotons from the Tevatron to be cooled and re-used[7]; (b) it operates as a cooler after the Accumulator, allowing it to operate more efficiently (the antiproton production rate decreases as the current in the Accumulator rises), and (c) it is a safe storage for antiprotons, the permanent magnets allowing the antiproton beam to survive a power cut of up to one hour.

4.2.8 The Tevatron

The Tevatron [73] is the last stage of the Fermilab accelerator complex. It is a circular synchrotron with a radius of 1 km. The acceleration of the protons and

[7]Because recycling the antiprotons at the end of a Tevatron run is lengthy and causes huge losses, this mode is no longer used.

antiprotons is achieved using 8 accelerating RF cavities. Dipole and quadrupole magnets steer and focus the beam. To accelerate the particles to TeV-scale energies, it is necessary to have large magnetic fields. It is more cost-effective to obtain such fields using superconducting magnets instead of resistive magnets, because of the large currents needed to obtain high fields. The Tevatron is therefore cooled to a temperature of 4.2 K using liquid helium.

The Tevatron accepts 150 GeV protons and antiprotons from the MI or the Recycler, and accelerate them to 980 GeV. The proton and antiproton beams circulate in opposite directions in the same magnetic field. Electrostatic separators generate a strong electric field preventing the two beams from interacting, except at the collisions points.

The Tevatron is composed of six sectors (A-F) with five service buildings each (0-4). These define points along the trajectory of the beams. The injection occurs at A0, which is also where one of two beam aborts is located. CDF (section 4.3) is located at B0. The second beam abort, for protons only, is located at C0. The DØ detector is in the building with the same name. The transfer lines from the Main Injector, as well as the RF stations putting back the beam into position when necessary are at F0.

The Tevatron also provides particles for fixed-target experiments. It can accelerate up to 3×10^{13} protons to an energy of 800 GeV, and deliver them in a single bunch, used in proton, meson and neutrino experiments. Table 4.1 shows the Tevatron accelerator parameters for Run I and Run II.

Luminosity As is the maximum energy, the luminosity ($\mathcal{L}$) is a key parameter in assessing the performance of an accelerator. It describes the number of collisions taking place at a given moment in time.

The luminosity is expressed as

$$\mathcal{L} = \frac{f N_b N_{\bar{p}} N_p}{2\pi[(\sigma_p^L)^2 + (\sigma_{\bar{p}}^L)^2]} \times F\left(\frac{\sigma_{p,\bar{p}}^L}{\beta*}\right), \tag{4.1}$$

Parameter	Run Ib	Run IIb
Number of bunches (N_b)	6	36
Protons per bunch (N_p)	2.3×10^{11}	2.7×10^{11}
Antiprotons per bunch ($N_{\bar{p}}$)	5.5×10^{10}	3.0×10^{10}
Total antiprotons	3.3×10^{11}	1.1×10^{12}
β^* [cm]	35	35
Bunch length (σ^L) [m]	0.60	0.37
Bunch spacing [ns]	3500	396
Interactions per crossing	2.5	2.3
Energy per proton or antiproton [GeV]	900	980
Maximum peak luminosity ($\mathcal{L}$) [$\mathrm{cm}^{-2}\mathrm{s}^{-1}$]	20×10^{30}	430×10^{30}
Integrated delivered luminosity ($\int \mathcal{L}dt$) [pb^{-1}]	112	10000

Table 4.1: Tevatron accelerator parameters for Run I and Run II.

where $f = 47713$ Hz is the revolution frequency, N_b is the number of bunches in each beam, N_p ($N_{\bar{p}}$) is the number of protons (antiprotons) per bunch, and σ_p ($\sigma_{\bar{p}}$) is the RMS beam width at the interaction point. F is a geometrical form factor that depends on the bunch longitudinal size ($\sigma^L_{p,\bar{p}}$) and the transverse beam width (β^*) at the interaction point, which is proportional to the beam's extent in phase space. The beam parameters are shown in 4.1. The Tevatron accelerator can provide an initial luminosity of the order of 10^{32} cm^{-2}s^{-1}.

As a consequence of (4.1), the instantaneous luminosity decreases with time, as particles are lost and beams heat up, i.e. spread in phase space. While collisions take place in the Tevatron, the luminosity decreases to about 50% of its original level in about seven hours and to about 35% in twelve hours. During that time, new antiprotons are stored in the Accumulator. After fifteen to twenty hours, it is best to dump the content of the accelerator and start a new cycle (eventually recovering the remaining antiprotons).

Figure 4.4 shows the performance of the Tevatron and the integrated luminosity collected by CDF during Run II. From this plot, it is clear how continuous improvements to the accelerator complex and its operation led to the rapid increase of the instantaneous luminosity, and thus to the increase of interesting physics events for the experiments.

It is worth mentioning that the instantaneous luminosity used to determine the collision rate for a given process is not derived from (4.1), but measured, as described in section 4.3.5. Moreover, the luminosity recorded by the experiments (shown for CDF in figure 4.4 as a function of time) is less than the luminosity delivered by the Tevatron. The average recording efficiency is $\sim 85\%$ (figure 4.5), which is due to: (a) unavoidable trigger dead-time (section 4.3.6); (b) unstable beam conditions (high particle losses) that do not allow a safe operation of the detector (which could be destroyed by irradiation); (c) small detector problems, or (d) operational decisions to dedicate part of a store to detector studies.

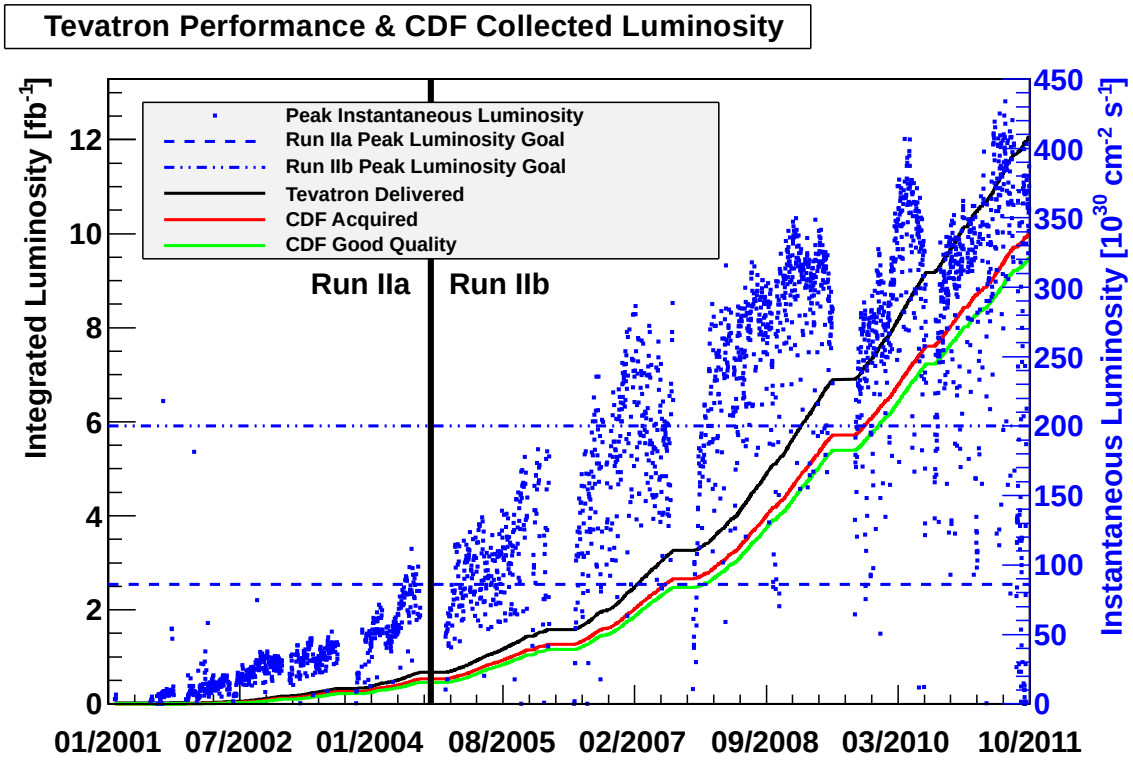


Figure 4.4.: The performance of the Tevatron and the luminosity collected by CDF during Run II. Left axis: integrated luminosity delivered by the Tevatron (black line), collected by CDF (red line) and analyzable (green line). Right axis: Peak instantaneous luminosity for each physics run (blue dots), together with the goals for Run I (dashed line) and Run II (dot-dashed line). The analyses presented in this dissertation use up to 7.8 fb^{-1} of data (collected till March 6, 2011).

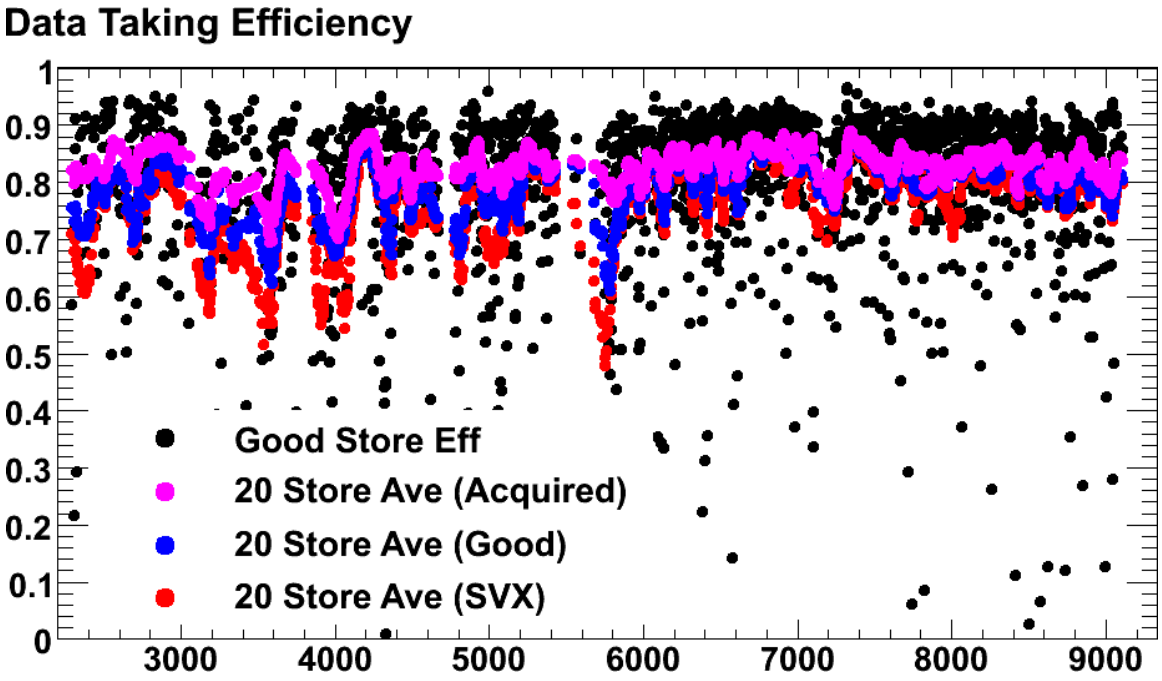


Figure 4.5.: Data taking efficiency as a function of time (store number). The average acquired efficiency is of the order of 85%. This factor is slightly lower for good quality data. In particular, the silicon detector (SVX, section 4.3.1) suffers from a relatively large unavailability. Credit: The CDF Collaboration

Luminosity leveling The performance of an experiment is not always better at higher values of luminosity. Depending on the physics goals and the intrinsic properties of the particle detectors, it might be wiser to reduce the collision rate to manageable values. The initial peak luminosity is often limited on purpose at the Tevatron, as the detectors suffer from high dead-times in a high luminosity regime rendering them almost useless for data acquisition, until the instantaneous luminosity decreases below $350 \cdot 10^{-3}$ cm^{2}s^{-18}.

4.3 The CDF II detector

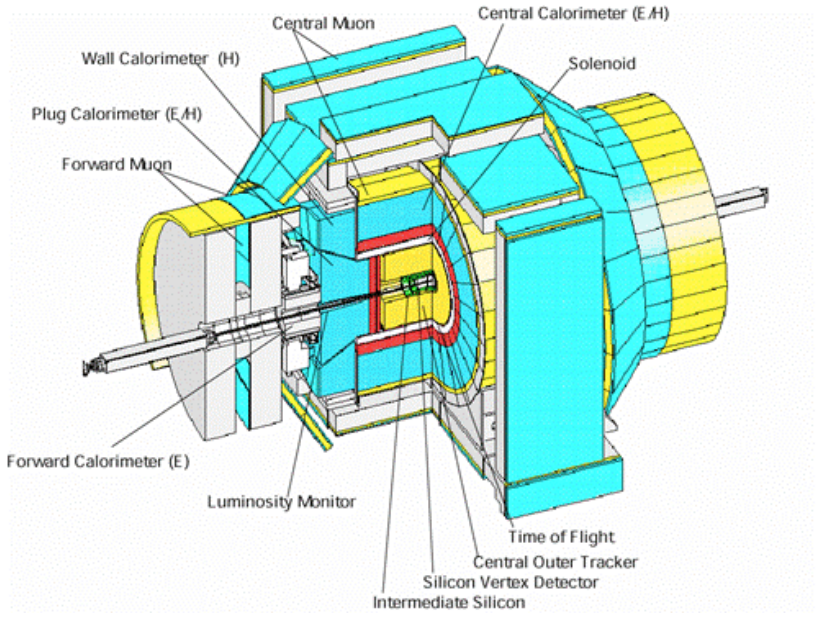


Figure 4.6.: The CDF II detector. Credit: The CDF Collaboration

The Collider Detector at Fermilab (CDF II) [74] is a multipurpose experiment designed to study high energy $p\bar{p}$ collisions. It has azimuthal and forward-backward symmetries. It features precision tracking of charged particles, fast projective calorimetry and muon detection.

[8]The author was on data acquisition shift on several of these records, though not for the absolute record. The systems had trouble coping with the high collision rate, well above design specifications.

The tracking system is inside a superconducting 4.8-meter-long by 1.5 m radius solenoid generating a magnetic field of 1.4 T parallel to the beam axis. The magnetic field is uniform within 0.1% in the entire tracking volume. The calorimeters and the muon system are located outside of the solenoid.

The coordinate system used follows the symmetries of the detector: the positive z-axis lies along the direction of the incident proton beam, ϕ is the azimuthal angle, θ is the polar angle measured from the detector centre. The rapidity

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}$$

is invariant under a Lorentz boost and is therefore useful in physics analyses of high energy (relativistic) particles. The pseudo-rapidity

$$\eta = f(\theta) = -\ln\left(\tan\frac{\theta}{2}\right)$$

replaces the polar angle since it is equal to the rapidity in the massless approximation, i.e. $E \gg mc^2$.

Figures 4.6 and 4.7 respectively show an overall view and a cross-sectional elevation of the CDF II detector. Figure 4.8 shows the η-coverage of some of its tracking and calorimetry subsystems.

4.3.1 The Tracking System

The purpose of the tracking system is to precisely measure the momentum of charged particles by reconstructing their trajectory (or *track*).

The identification of charged particle tracks, or *tracking*, is performed using an integrated system consisting of an inner and outer tracker, which are a silicon strip detector and a large gaseous drift chamber. The entire system is located close to the beam pipe, and in a region with a 1.4 T magnetic field generated by a solenoid parallel

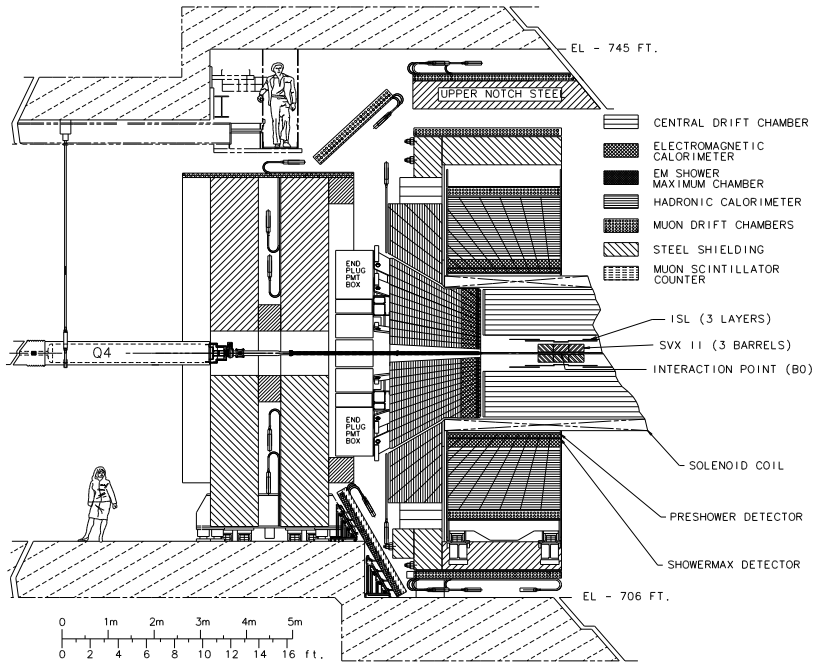


Figure 4.7.: A cross-sectional elevation view of the CDF II detector. Credit: The CDF Collaboration

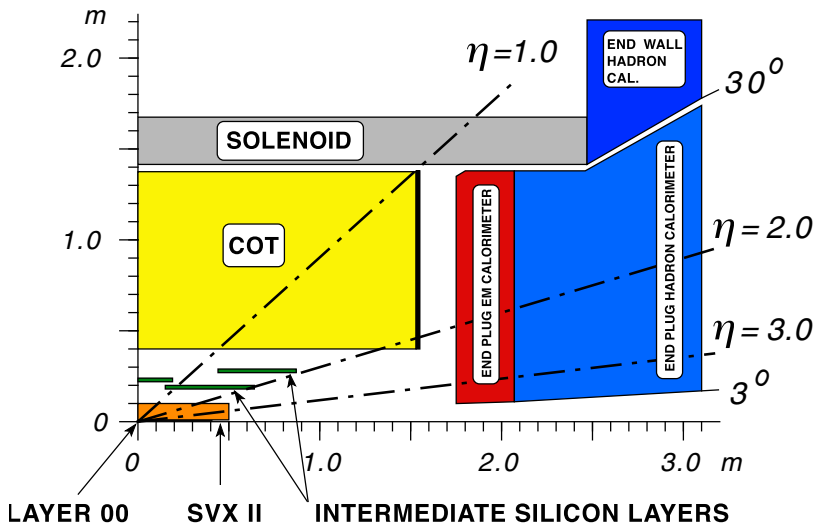
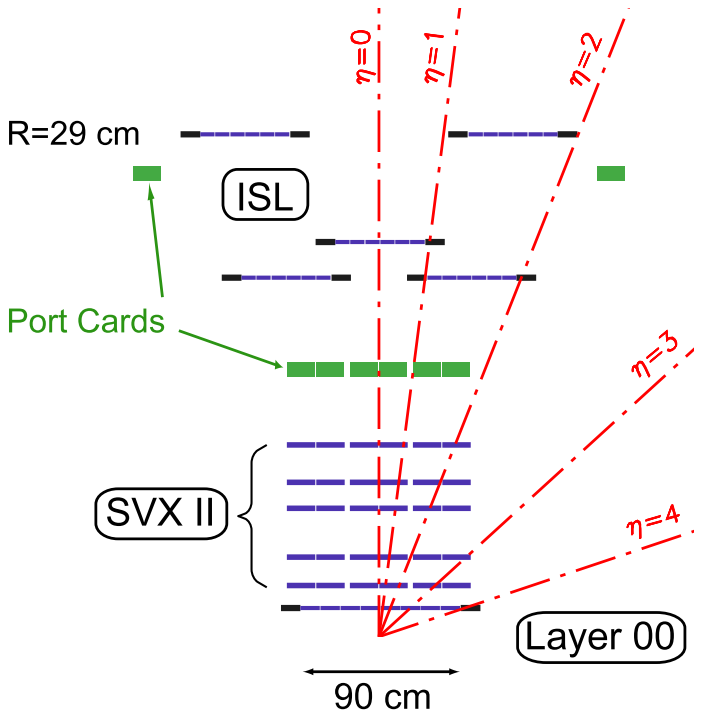


Figure 4.8.: A cross-sectional view of the CDF II tracking subsystems (orange, green and yellow). Also shown is a part of the calorimeter system (red and blue). The solenoid (gray) generates a 1.4 T magnetic field parallel to the beam axis. Credit: The CDF Collaboration

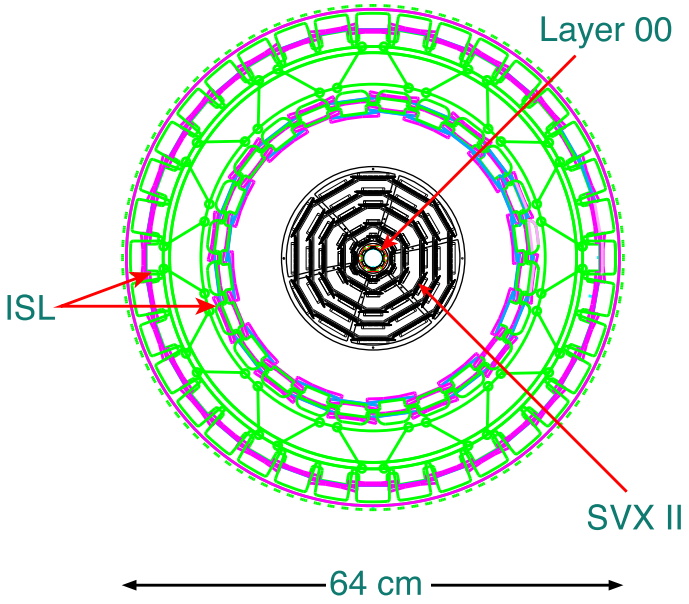
to the beam axis; this field bends the trajectory of a charged particle, allowing for the measurement of the particle's momentum (cf. section 5.2).

The Inner Tracker

The CDF II inner tracker, or silicon detector, is composed of three sub-detectors, built using different silicon sensor design and layout. These sensors consist essentially in a $p-n$ junction (diode) with a reverse-bias applied. As a charged particle passes through the semiconductor material, ionization occurs and produces electron-hole pairs. The electrons migrate to the anode (and the holes to the cathode) of the diode. The amount of charge collected there is proportional to the length of the path of the particle in the semiconductor material. By segmenting the p (or n) part of the junction into strips, and reading out the charge deposition separately for each strip, it is possible to measure the position of the (charged) particle passing-by. The (micro-) strips in the CDF silicon detector are separated by about 60 μm.



(a) Side view



(b) In the $r - \phi$ plane

Figure 4.9.: The CDF inner tracker. (a) Side and (b) transverse view.

SVX The Silicon Vertex Detector (SVX) [75, 76] is the core of the silicon tracking system. It is composed of a set of three cylindrical barrels, which are organized radially in five layers of double-sided silicon wafers, and extend from 2.5 (L0) to 10.7 cm (L4) from the beam. One side of the wafers hosts *axial* strips that run along the beam axis, while the other side hosts *stereo* strips perpendicular to the z axis (three such layers) or tilted by 1.2° with respect to the transverse direction (two such layers). Thus, hits on one side of the wafers provides a measurement of the hit in the $r - z$ plane, while the other side measures it in the $r - \phi$ plane; this layout allows for a rough three-dimensional reconstruction of tracks.

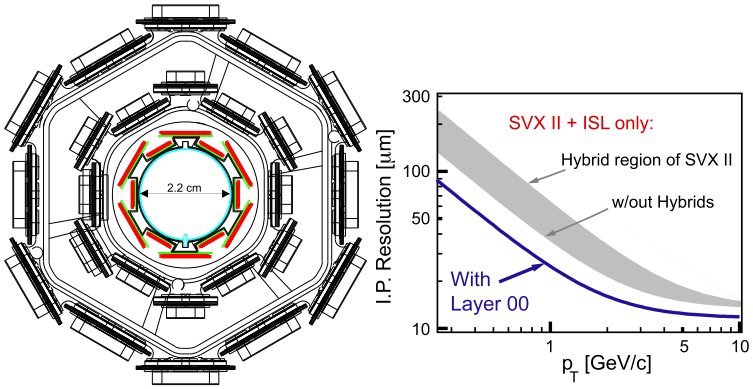
With a 90 cm active length, the SVX detector covers the (gaussian) longitudinal spread of the interaction point up to 3 standard deviations away from its nominal value ($z = 0$). It covers the region $|\eta| < 2.0$ with a resolution of 12 μm for a single hit.

L00 The Layer 00 (L00) [77] is a radiation-hard single sided silicon detector. It is mounted on the beam pipe, supported by a carbon fiber support with integrated cooling[9]. Its sensors are arranged in two 90 cm long sublayers at $r = 1.35$ and $r = 1.63$ cm (figure 4.10), covering the region $|\eta| < 4.0$.

L00 was designed to improve the resolution on the impact parameter of charged tracks, which was degraded due to multiple scattering off the passive material[10] of the SVX layers. It allows reaching a resolution of about 25 μm (from about 40 μm before it was installed) for tracks in the intermediate p_T range. Even though it only provides information in the transverse plane, L00 provides important information to identify the decay of long-lived hadrons containing a b-quark (chapter 6) because it is so close to the beam.

[9]The amount of readout electronics required to operate the silicon detector generates a lot of heat, even when in standby; the system must therefore be cooled constantly.

[10]The electronics inside the detector require the presence of cables and cooling tubes, which adds material in the active sensitive volume of the detector, increasing multiple scattering, and thus adversely affecting the resolution.



(a) (b)

Figure 4.10.: Layer 00 of the silicon detector. This layer is mounted directly on the beam pipe (cyan) and consists of an alternation of wide and narrow modules (red) mounted on a carbon fiber support. Credit: The CDF Collaboration

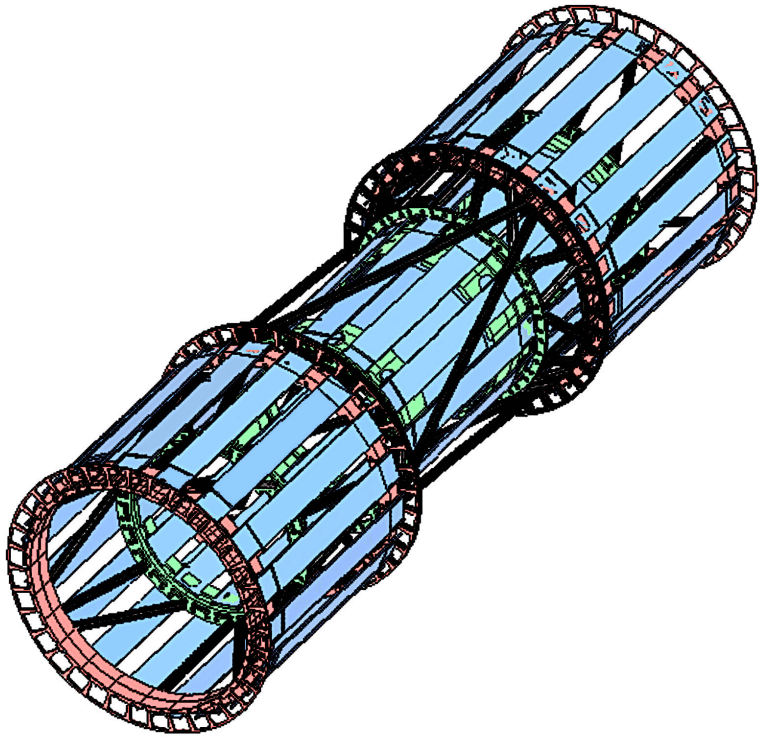


Figure 4.11.: The ISL detector: perspective view. Credit: The CDF Collaboration

ISL The Intermediate Silicon Layers (ISL) [78] serve as a link between the inner silicon tracking region and the outer wire tracker (COT, section 4.3.1). It is 1.9 m long and is made of five layers: (a) a central layer at $r = 22$ cm covering $|\eta| < 1.0$, and (b) on each side of the $z = 0$ plane, two layers, at $r = 20$ and $r = 28$ cm, covering the forward region ($1.0 < |\eta| < 2.0$). Each layer is double-sided and hosts radial strips on one side, and small amplitude stereo strips (tilted by 1.2°) on the other.

The ISL improves the tracking capabilities in the region $|\eta| < 1.0$ by providing additional hits to those from the SVX and the COT, and allow a standalone tracking outside the COT pseudo-rapidity coverage.

The information provided by all these sub-detectors are used offline by the track reconstruction algorithms. However, SVX plays a crucial role online, and especially

to trigger for the presence of B hadrons, providing crucial information to the Silicon Vertex Trigger (SVT) [79], described in section 4.3.6.

The Central Outer Tracker

The Central Outer Tracker (COT) [80] is a cylindrical open-cell drift chamber. It is located between the Time of Flight (TOF) detector and the inner silicon tracker, in the radial region from 40 to 138 cm. It extends longitudinally from -155 to +155 cm (z axis). It provides up to 96 measurements of the track position with alternating axial and $\pm 2^\circ$-stereo superlayers of 12-wire layers each (four of each, cf. figure **??**). This layout provides a full coverage for $|\eta| < 1$ and maximal acceptance for $|\eta| < 2$. The COT provdes a very precise measurement in the $r-\phi$ plane, and thus of the transverse momentum p_T, but is less precise for measurements along the z axis (longitudinal momentum p_z).

To form the drift cells, each sense wire is surrounded by potential wires. As charged particles travel through the COT, they ionize the gas mixture. The electrons created drift towards the sense wires, and create an avalanche of electrons through secondary ionization. The nominal spacing between the sense and potential wires is ~ 3.6 mm. The sense (potential) wires have an applied voltage of $2.6-3$ kV ($1-2$ kV). The chambers are filled with a mixture of Argon-Ethane (1:1 ratio) gas mixture in isopropyl alcohol (1.7%). This choice yields a constant drift velocity along the cell width, and ensures that the maximum drift time (less than 200 ns) is sufficient for all the electric charges produced by ionization in the drift chamber to drift away before the next bunch crossing (which arrives 396 ns later).

In the COT, the drift of the electrons is approximately perpendicular to the wires; this is achieved by tilting the COT cells by 35°, which corresponds to the Lorentz

angle in the COT[11]. The hit resolution of the COT is ~ 140 μm, which provides a momentum resolution of

$$\frac{\sigma(p_T)}{p_T^2} = 0.17\% \cdot (\mathrm{GeV/c})^{-1}. \tag{4.2}$$

The COT also provides useful information for particle identification. The ionization per unit unit length of the track, dE/dx, can be used to separate kaons K, pions π, and protons.

4.3.2 The Calorimeter

The CDF II electromagnetic and hadronic calorimeters [81–84] surround the tracking system and measure the energy of charged and neutral particles produced during the interaction. Therefore, they can also provide a measure of the energy of particles leaving the detector without interacting, by measuring the transverse energy imbalance, also known as missing transverse energy $\not\!\!E_T$.

When passing through the detector material, electrons lose their energy mostly through bremsstrahlung, while photons produce electron-positron pairs. A shower is formed in the detector by a cascade of these two processes. The amount of energy deposited by these showers as particles traverse the detector volume increases until the particles in the shower are not energetic enough to initiate these processes. This corresponds to a maximum depth, referred to as the *shower maximum*. Hadrons can lose their energy by ionization or secondary nuclear interactions, forming showers similar to electrons and photons.

The calorimeters are designed to be thick enough to fully capture all of the energy that the particle possesses. The energy lost by a particle passing through the calorimeter depends on the type and initial energy of that particle. The radiation length X_0 is the distance over which a traversing electron loses on average a fraction

[11]Because the COT is enclosed in a 1.4 T magnetic field, the electrons experience a Lorentz force that rotates their path.

$1-1/e$ of its energy in a given material. The nuclear interaction length λ_0 is the mean free path necessary for a hadron to undergo a nuclear inelastic interaction. X_0 and λ_0 are the main characteristics of the detector. The number of interaction lengths hadrons need to lose their energy is less than the number of radiation lengths electrons or photons need. This is due to a higher multiplicity in nuclear interactions. However, the interaction length λ_0 is usually larger than the radiation length X_0 [7].

The electromagnetic and hadronic calorimeters are respectively lead-scintillator and iron-scintillator sampling devices, which consist of alternating layers of passive absorber (made from a heavy element) and active scintillator tiles for shower sampling. The coverage of the calorimeters is nearly 4π solid angle, and covers all of ϕ and the region $|\eta| < 3.6$. In the central region ($|\eta| < 1.1$), they are segmented in towers of 15° in azimuth and 0.1 in η, and the forward region ($1.1 < |\eta| < 3.6$) in towers of 7.5° for $|\eta| < 2.11$ and 15° for $|\eta| > 2.11$. The electromagnetic calorimeters are instrumented with proportional and scintillating strip detectors that measure the transverse profile of electromagnetic showers at a depth corresponding to the expected shower maximum.

The CDF calorimeter system consists of the Central and Plug Electromagnetic Calorimeters (CEM and PEM) [81], and the Central, Wall and Plug Hadronic Calorimeters (CHA, WHA and PHA) [84]; their properties are presented in table 4.2. The Central Electromagnetic Showermax (CES) [81] and the Plug Electromagnetic Showermax (PES) [83] detectors are used for photon and electron identification, and are located at around $6X_0$ distance into the detector.

The measured energy resolution for electrons in the electromagnetic calorimeters are $13.5\%/\sqrt{E_T} \oplus 2\%$[12], in the central and $16\%/\sqrt{E_T} \oplus 1\%$,in the forward where the units of E_T are GeV. We also measure the single-particle (pion) energy resolution in the hadronic calorimeters to be $50\%/\sqrt{E_T} \oplus 3\%$ for the central and $80\%/\sqrt{E_T} \oplus 5\%$ for the forward detector [74].

[12]We define the transverse energy as $E_T = E \sin\theta$ and transverse momentum as $p_T = p \sin\theta$, where E is the energy measured in the calorimeter and p is the magnitude of the momentum measured by the spectrometer.

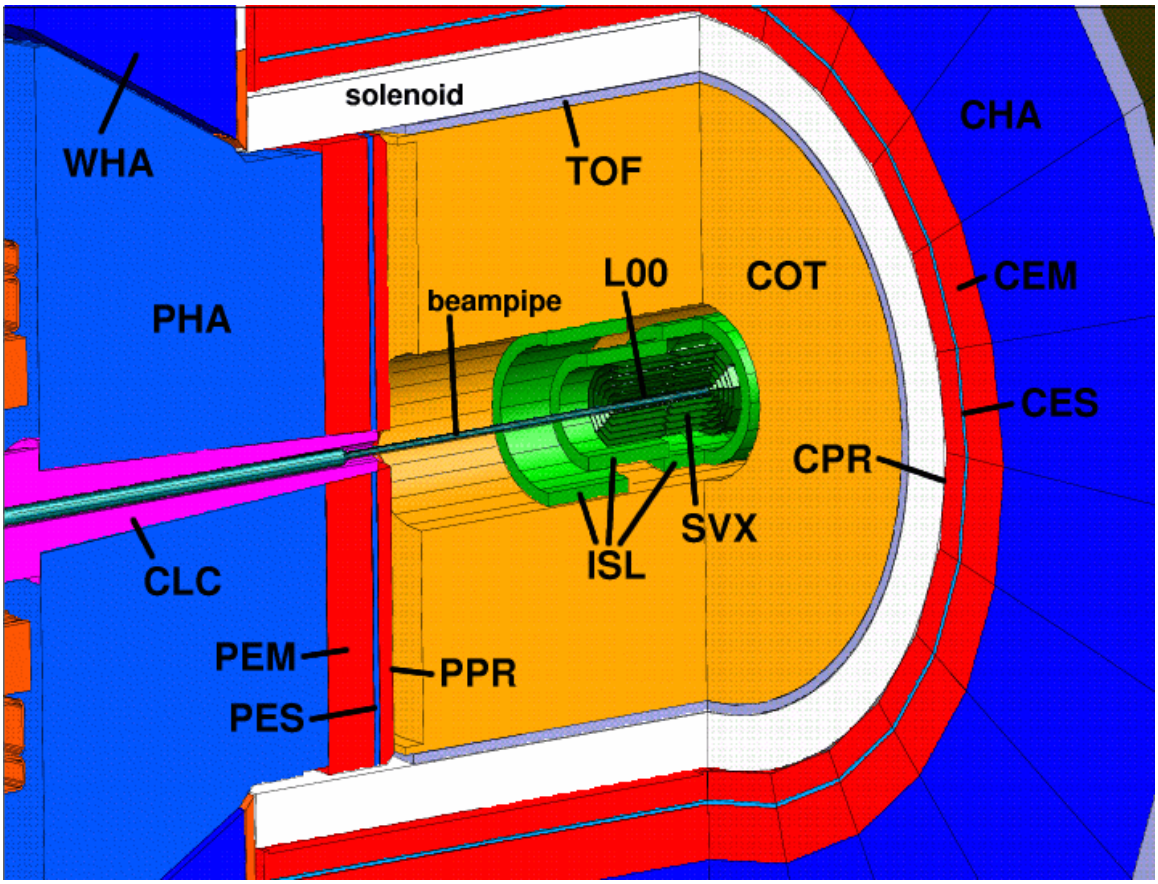


Figure 4.12.: Schematic view of the electromagnetic and hadronic calorimeters of CDF surrounding the tracking system. Credit: The CDF Collaboration

Table 4.2: Main characteristics of the electromagnetic and hadronic calorimeters of CDF. Source: [81–84]

Subsystem	Coverage	Depth	Resolution
CEM	$\lvert\eta\rvert \leq 1.0$	$18X_0$	$13.5\%/\sqrt{E_T} \oplus 2\%$
PEM	$1.1 \leq \lvert\eta\rvert \leq 3.6$	$23.2X_0$	$16\%/\sqrt{E_T} \oplus 1\%$
CHA	$\lvert\eta\rvert \leq 0.9$	$4.7\lambda_0$	$50\%/\sqrt{E_T} \oplus 3\%$
WHA	$0.9 < \lvert\eta\rvert \leq 1.3$	$4.7\lambda_0$	$75\%/\sqrt{E_T} \oplus 4\%$
PHA	$1.3 < \lvert\eta\rvert \leq 3.6$	$6.8\lambda_0$	$80\%/\sqrt{E_T} \oplus 5\%$

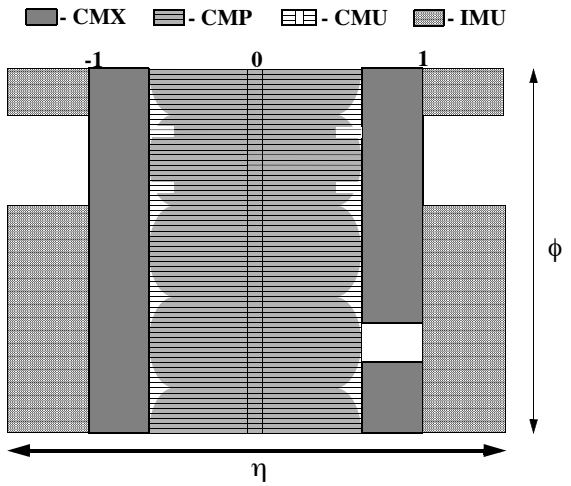


Figure 4.13.: Muon coverage as a function of η and ϕ. Credit: The CDF Collaboration

4.3.3 The Muon Detector

Muons traverse large amounts of material before loosing a significant amount of energy, leaving a minimal amount of energy in the calorimeter. This is because they are are minimum ionizing particles in a large energy range (from a few hundreds of MeV to a few TeV) and loose much less energy in bremsstrahlung than lighter particles like the electron. In contrast, the majority of the particles produced in $p\bar{p}$ collisions are absorbed by the calorimeter material. Therefore, the muon system is the outermost layer of the CDF II detector. To improve the purity of the muon sample, additional layers of steel absorber are located right after the calorimeter and the magnet return yoke.

The muon system provides a combined coverage in the region $|\eta| < 1.5$. It consists of four sub-detectors: the Central MUon detector (CMU), the Central Muon uPgrade (CMP), the Central Muon eXtension (CMX) and the Intermediate MUon system (IMU). The latter consists of a barrel-shaped array of muon chambers (BMU) and scintillators (BSU), mounted parallel to the beam line, and one ring-shaped array of scintillators (TSU) [85,86]. Figure 4.13 shows the muon coverage as a function of the pseudo-rapidity η and the azimuthal angle ϕ.

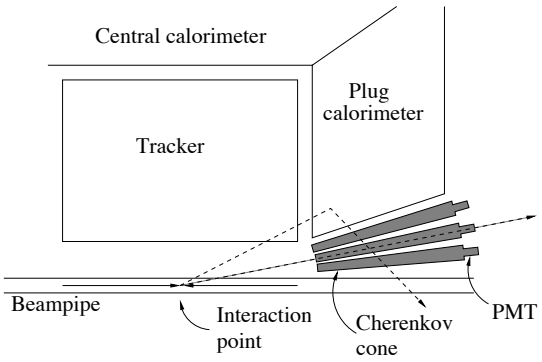


Figure 4.14.: A schematic view of the luminosity monitor. Credit: The CDF Collaboration

4.3.4 The Time of Flight Detector

Particle identification is obtained from the ionization energy loss, measured through dE/dx in the COT. The dE/dx measurement provides about one standard deviation separation between charged kaons and charged pions for momenta greater than 2 GeV/c.

The purpose of the Time of Flight (TOF) detector [87] is to improve particle identification, in combination with the momentum measurement p from the tracking system. It consists of a barrel of scintillators located at 140 cm from the beam line (about 5 ns flight time for the fastest particles). The TOF resolution of 100 ps allows for at least two standard deviation separation between $K^{\pm}$ and $\pi^{\pm}$ for momenta $p < 1.6$ GeV/c, complementing the measurement from the COT.

4.3.5 The Luminosity Counter

At CDF II, the luminosity is measured using low pressure gaseous Cherenkov Luminosity Counters (CLC) [88, 89] placed at small angles relative to the direction of the beam. The luminosity monitoring detector is composed of a segmented array of such counters, as shown in figure 4.14.

4.3.6 The Trigger and Data Acquisition

The interaction rate at CDF is orders of magnitude higher than than what the data acquisition system can handle. Moreover, the majority of the $p\bar{p}$ collisions are not of interest, since they are due to inelastic scattering. Indeed, the cross-section for inelastic scattering is about 60 mb, while the top pair production has a cross-section which is of the order of 7 μb. The CDF data acquisition system can operate at a maximum rate of 18 MB/s. Given an average size of 170 kB per event, this corresponds to a maximum rate of 100Hz. Since the collision rate at the Tevatron is at about 1.7 MHz, we are forced to reject 99.9% of the events.

This requires a trigger system that preselects events online and decides whether the detector data corresponding to the event should be written to tape or discarded. The trigger evaluates if a given event should be read out, reconstructed and stored. The design of a trigger highly depends on the physics goals, as its task is to select interactions of possible interest for physics analyses. The CDF trigger consists of three levels, at which decisions are made based on increasingly more complex information. Figure 4.15 shows a block diagram of flow of data in the the CDF II trigger system.

The two first levels, L1 and L2, are hardware based, while the level 3 (L3) consists of a computer farm implementing filters that resemble the offline reconstruction code. L1 and L2 use only a subset of detector information to make their decision to either pass the event on for subsequent processing or to reject it. L3 is the only level at which the entire detector data is available and the event is fully reconstructed before the trigger decision is made. Events passing the L3 trigger are transferred to the mass storage. The first two trigger levels reduce the output rate by a factor of 20,000 ; the L3 farm reduces it further by about a factor of four.

The Level 1 Trigger

The L1 trigger operates using only a subset of the detector. Three parallel systems examine each event, as shown in figure 4.16: (a) the L1-CAL trigger boards

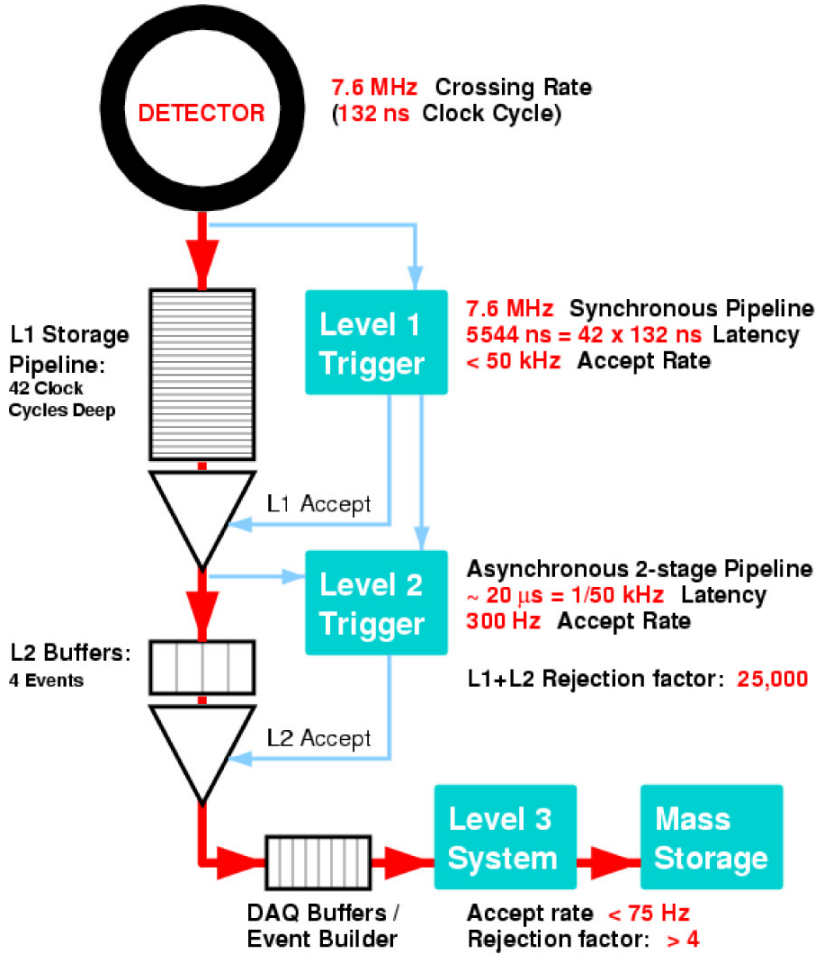


Figure 4.15.: A block diagram of the dataflow in the CDF II trigger system. Credit: The CDF Collaboration

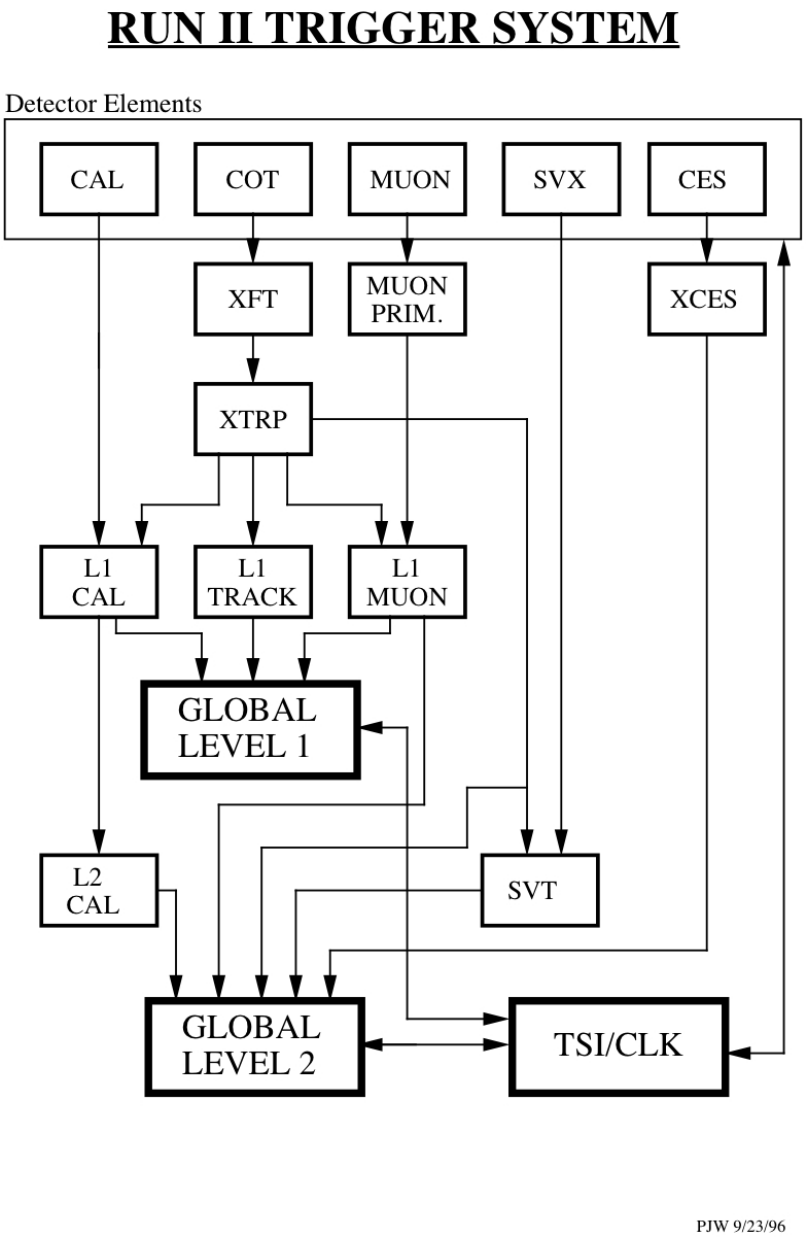


Figure 4.16.: A block diagram of the CDF II hardware trigger system. Credit: The CDF Collaboration

find objects in the calorimeter; (b) the L1-MUON trigger cards identify muons, and (c) the L1-TRACK trigger implements the eXtremely Fast Tracker (XFT) [90,91] to reconstruct tracks in the COT. The output of the XFT is fed to the extrapolation module (XTRP) [92], whose goal is to distribute the tracks and information derived from them to the L1 and L2 systems; this allows to match the tracks to energy deposits in the calorimeter towers, or hits in the muon chambers. The information from all three systems is used independently to determine whether to pass the event to L2 or not. The decision is based on the number and energy of tracks, electron, photon, μ, τ, and jet candidates, as well as on the total energy and $\not\!\!E_T$ in the event. If any of these requirements is met (logical OR), the event is passed to L2.

The calorimeters, COT, muon and TOF systems provide information to the L1 trigger. The Trigger Supervisor System (TSI) is responsible for maintaining synchronization and allocates buffer space for each event accepted at L1. The typical acceptance rate of the L1 system is 25 kHz.

The Level 2 Trigger

All the front-end readout components move the data corresponding to an event accepted by the L1 trigger to one of four asynchronous L2 buffers (figure 4.15). These buffers are sufficient for a 25 kHz L1 accept rate with about 5% dead-time. At L2, information from the calorimeter, COT, muon, CES and SVX systems is used to reconstruct physical objects (section 10.2). Jet candidates are formed by the L2 cluster-finding algorithm by combining contiguous regions of the calorimeter. Additionally, the CES shower-max detectors improve the spatial resolution by providing more positional information on the calorimeter clusters. The Silicon Vertex Trigger (SVT) [79,93,94] combines SVX information with the L1 tracking information, which allows selecting tracks with large impact parameter (often produced by the decay of B hadrons, cf. chapter 6). The typical acceptance rate of the L2 system is 350 Hz.

The L2 Calorimeter Trigger Upgrade

The initial L2 calorimeter trigger algorithm for Run II of CDF was implemented in hardware, and based on a simple clustering algorithm from Run I. The global transverse energy information was directly passed from L1. While this system worked well at low luminosity, this simplistic approach yielded to large growth terms in the cross-section as the luminosity increased, with serious implications on the L2 trigger (high dead-time). For this reason, the L2 calorimeter trigger system was upgraded in order to make the full calorimeter trigger tower information directly available to the L2 decision CPU. This system allows running more sophisticated algorithms, and to significantly improve the purity and efficiency of the jet and $\not\!\!E_T$-related triggers, for the L2 jets and $\not\!\!E_T$ are very close to the offline quantities[13]. This upgrade was a big step forward to improve the L2 triggering capabilities. Clearly, this dissertation wouldn't be as it is without this upgrade.

The Level 3 Trigger

When L2 accepts an event, the full detector is read out, and the data from each front-end element is assembled into an event. The latter is fed to one of the processors of the L3 computing farm. The L3 processor reconstructs events and filters them, using the full event reconstruction (very similar to the offline code). The typical acceptance rate at L3 is about 100 Hz. These events are written to permanent storage.

4.4 Summary

In this chapter, we have reviewed the experimental apparatus, namely the Fermilab accelerator complex, the CDF detector and its data acquisition system. The information provided by the detector is analyzed in the next chapter to determine the kinematic properties of each collision event.

[13]In addition, quantities such as the dijet mass, the angular separation between jets (or $\not\!\!E_T$and jets), and the scalar sum of the jet clusters (H_T) were made available at L2.

5. EVENT RECONSTRUCTION AND OBJECT IDENTIFICATION

In order to properly identify the physical processes taking place when colliding the proton and antiproton beams, the energy deposits in the various sub-detectors of CDF are transformed into electronic signals and digitized. After the L3 trigger decides to keep an event, the raw data are calibrated and processed by the reconstruction software. The latter uses the digitized information to reconstruct physical objects (high-level objects) corresponding (ideally) to particles that have been produced during the collision. For this reason, these objects are sometimes referred to as *experimental observables*.

Particle identification consists in comparing the properties of an object against identification criteria, which reflect *a priori* knowledge about features of tracks, leptons, etc. Event reconstruction is the process by which the properties of a given collision are determined based on the combination of these objects to obtain high-level *event information*.

This chapter introduces the experimental observables used in this dissertation: charged particle tracks, the primary interaction vertex, leptons, jets, and missing transverse energy ($\not\!\!E_T$), among others. We will describe how these are obtained from low-level information such as silicon hits, energy deposits, etc.

5.1 Data Model and Analysis Software

In Run II, the CDF experiment uses a fully object-oriented data model [95, 96]. The ROOT [97] data format is used to handle and store data in an efficient way,

allowing to cope with the rate at which the information from the various sub-detectors has to be recorded and stored[1].

The analyses in this dissertation were performed on data stored in the *Standard Ntuple* (StNtuple) [98] format, which was developed for storing CDF data. In this format, data is organized in blocks containing a collection of same-type objects (in the object-oriented sense of the term). We use the ROOT analysis framework – based on the C++ programming language – along with custom made code to process the data.

5.2 Charged Particle Tracks

The trajectories of charged particles, called *charged tracks* or simply *tracks*[2], are reconstructed by combining *hits*, i.e. measurements, along these trajectories using the tracking system of the CDF experiment (section 4.3.1).

Located inside the tracking system is a uniform magnetic field generated by a solenoid magnet (section 4.3.1), which forces charged particles to follow a helical trajectory. The curvature of the helix depends on the momentum and charge of the particle. We use five parameters to describe these helices in the transverse plane: (a) the z coordinate of the point P_C of closest approach to the z axis (z_0); (b) the signed impact parameter (d_0), i.e. the distance between P_C and the z axis; (c) the direction in ϕ of the transverse momentum of the particle at P_C (tangent to the helix), ϕ_0; (d) the helix pitch (λ), which is equal to the cotangent of the polar coordinate of P_C, $\cot\theta$, and (e) the signed helix (half-)curvature (C).

The analytical expressions for the impact parameter and the curvature are given by

$$d_0 = q \cdot \left(\sqrt{x_C^2 + y_C^2} - \rho \right) \tag{5.1}$$

[1]The Tevatron experiments' data load paved the way to new tools for handling them. Since then, the LHC has taken this to a much higher level.

[2]Neutral particles leave no tracks; their mass, energy, and momentum can be found only from their decay products, assuming energy and momentum conservation, or from the energy deposits in the calorimeters.

and

$$C = \frac{q}{2\rho}, \tag{5.2}$$

where q is the particle charge in units of the charge of the electron, (x_C, y_C) is the center of the helix projected on the transverse $(x - y)$ plane, ρ is the radius of the helix in the transverse plane, and the computations are made along the (OP_C) axis. Figure 5.1 shows these parameters graphically.

The transverse and longitudinal components of the particle momentum (p_T) are then derived to be

$$p_T = 0.15\frac{qB}{|C|} \tag{5.3}$$

$$p_z = p_T \cdot \cot\theta, \tag{5.4}$$

where the magnetic field B is expressed in Tesla, C in m^{-1}, and p_T in GeV/c.

Several algorithms are used at CDF to reconstruct tracks [99], depending on which component of the detector the particle travels through:

COT Stand-Alone Tracking The COT stand-alone tracking algorithm only uses hits[3] from the central outer tracker (COT, section 4.3.1).

The algorithm identifies circular paths[4] in the axial super-layers starting with *segments* of four or more[5] hits that can be fit to a straight line. These segments are then combined into $r-\phi$ tracks using two methods [99]: *segment linking* and *histogram linking*. Hits common to two tracks are assigned to the track with the largest number of hits[6]. The histogram linking generates for each segment a circular trajectory, centered on the beam line, and processes the hits found within 1 cm of the generated circle. The r coordinate of each of these hits is plotted in a histogram (with 200 μm bin width). When one of the bin reaches a count of 10, a track with the corresponding radius is reconstructed. A new

[3]A hit is an interaction inside of a cell of the detector.
[4]The projection of a helical trajectory on the transverse plane is a circular path.
[5]After four hits have been found, additional hits within 1 mm are added to the fit.
[6]At high luminosity, the high activity yields many tracks that can be close to each other; hits from one track can then shadow hits from other tracks, which reduces the reconstruction efficiency.

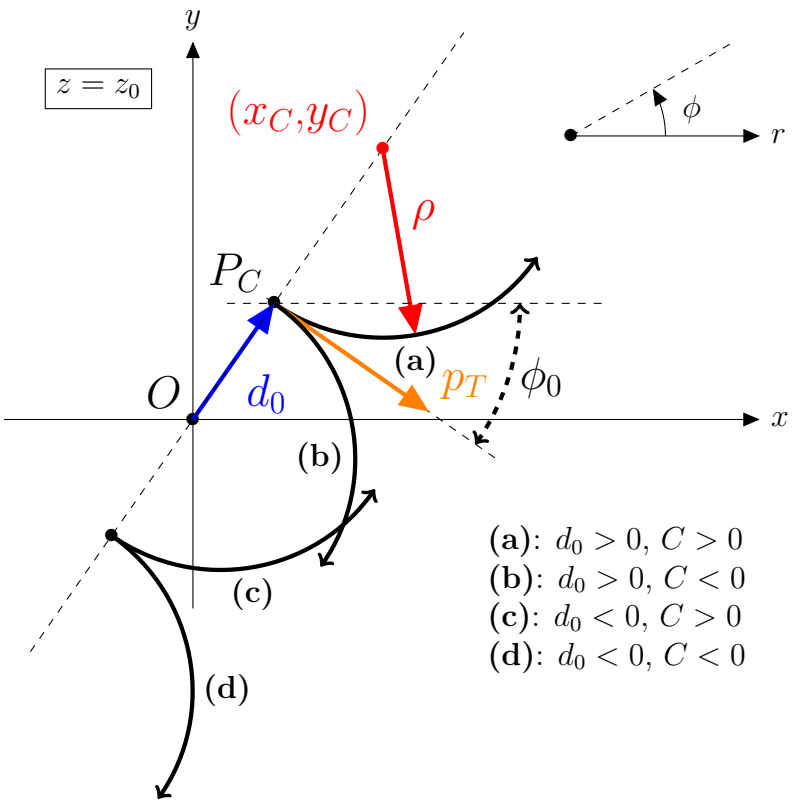


Figure 5.1.: Parameterization of a track's trajectory in the $r - \phi$ plane: a positively charged track (a) is shown with its radius (ρ) and its impact parameter (d_0). The curvature (C), derived from these according to equation 5.2, is positive. Also shown are the sign of d_0 and that of C for three other categories of tracks (b)-(c).

fit is performed with all the hits within 750 μm of the new radius. These two algorithms are complementary, so that when one fails to reconstruct a track, the other usually succeeds. Valid tracks must contain more than 20 hits.

In a second phase, these $r-\phi$ tracks are combined using stereo information. Starting from the outer stereo layer, stereo segments are matched to tracks, re-fitting at each step. The newly estimated trajectory is used to determine which remaining stereo segments are consistent with belonging to the track.

The track parameters are corrected for material effects, variation in the magnetic field, energy losses, etc. The COT tracking efficiency is very high, and of about 97% for tracks with $p_T > 10$ GeV/c^2 passing through all super-layers.

Outside-In Tracking The Outside-In (OI) algorithm, which exploits informations from both the silicon detector (section 4.3.1) and the central drift chamber (COT), is used to track the particles in the central region ($|\eta| < 1.0$) of the detector. As its name suggest, this algorithm reconstructs the tracks first in the COT, where the larger radius yields a lower track density (and thus better separated tracks) and then moves inwards towards the beam pipe.

Silicon Stand-Alone Tracking The remaining[7] silicon hits that are not associated with COT tracks are used to construct silicon stand-alone (SSA) tracks. The SSA algorithm extends the tracking coverage to the region $|\eta| < 2.0$ [100].

Inside-Out Tracking The Inside-Out (IO) algorithm [101] extrapolates standalone silicon tracks into the COT, and allows recovering tracks that traverse the COT because they partially are in the intermediate rapidity region.

Track quality A track is required to have more than two (one) axial (stereo[8]) COT segments with at least five hits each. The track z_0 must be within 60 cm from the

[7]To drastically reduce the number of combinations possible, hits that are already attached to another track are not considered in this step.

[8]When the track is matched to a stub-less muon (cf. section 5.5.2), the number of stereo segments must also be larger than two. In this case, and because the probability of muon mis-identification is higher than for stub muons, the tighter requirement helps suppressing mis-reconstructed muons.

center of the detector, in order to originate from the luminous region and fall in the tracking volume. The impact parameter of the track, d_0, should be within 0.2 mm of the primary vertex if the COT track is matched to a track in the silicon tracker; otherwise, d_0 is required to be within 0.2 cm of the primary vertex. This helps remove tracks from cosmic rays and non-prompt decays. The track is constrained to originate from the beam spot, and the track parameters are adjusted accordingly. During the track reconstruction process, the quality of the fit to the hits, measured by $\chi^2/n.d.f$ is required to be less than 2.75. An angular dependent curvature correction is needed to take into account mis-alignment in the COT

5.3 Beam Position and Primary Interaction Vertex

For each event, we determine the primary vertex from fitting high quality, low impact parameter tracks associated to the vertex with the highest total scalar sum of transverse track momentum. It is worth noting that, because of the lack of identified leptons in our selection, we cannot assign the primary vertex to be the vertex nearest the identified high-momentum electron or muon, as is generally done in analyses with leptons [102].

The position of the primary vertex is found by a fit using all tracks within the z window and with an impact parameter significance (relative to the beam line) of $|d_0/\sigma(d_0)| < 3$, where $\sigma(d_0)$ includes the uncertainty on both the track and the beam line positions. The transverse profile of the beam line – 30 μm at $z = 0$, rising to $50 - 60$ μm at $|z| = 40$ cm – is used to constrain the fit. We remove tracks with large χ^2 contribution. The fit is repeated without the removed tracks until a vertex with no tracks with $\chi^2 > 10$ is found. In the case where no tracks survive this pruning, the beam line profile is used to estimate the primary vertex position. The resulting transverse vertex position uncertainty ranges from 10 to 32 microns, depending on the number of reconstructed tracks and the topology of the event [102].

5.4 Calorimeter Clusters

The information contained in the many calorimeter cells is represented by a set of calorimeter clusters. For an event, each cluster is assigned a vector $\vec{E}$ whose magnitude is equal to the energy deposit in the tower, and which is pointing away from the event primary vertex position to the center of the calorimeter tower. The projection on the transverse plane of the measured energy, $\not\!\!E_T$, has a magnitude of $E \sin\theta$.

Particles with high momentum (electrons, photons, or the decay products of jets) deposit energy in contiguous groups of calorimeter towers, which are identified as an energy cluster by a calorimeter clustering algorithm. First, *seed* clusters are identified (with an energy higher than a certain threshold), and adjacent towers with an energy higher than a second (lower) threshold are added iteratively to form a cluster. The position of this cluster is by convention defined by the energy-weighted mean position of the towers in the cluster, and the energy assigned to it is the sum of the energies of each tower that is part of the cluster.

These clusters are then used to form higher-level objects, i.e. jets (cf. section 5.6.1).

5.5 Lepton Identification

Although the analyses described in this dissertation do not require a precise lepton identification, since we reject events containing identified leptons, we present in this section the identification process. This is to satisfy the curious reader, and allows to ensure that the looser requirements we make to reject leptons (presented in the next section) are preserving the orthogonality between the $\not\!\!E_T$+jets sample and the samples containing identified leptons.

Electrons are reconstructed as charged particles reconstructed in the tracking system, striking the calorimeter, and leaving the majority of their energy in the electromagnetic section of the calorimeter. Muons are identified as charged particles in the

Electrons must have been detected in the active region of the CES/CEM
$E_T \geq 20$ GeV/c^2
$E_{\rm HAD}/E_{\rm EM} \leq 0.055 + 0.00045 \times E$
More than three COT axial segments with five hits each
More than two COT axial segments with five hits each
Track $\lvert z_0 \rvert < 60$ cm
Track $p_T > 10$ GeV/c
$E_{\rm ISO}/E < 0.1$
$E/p \leq 2$ for $p_T \leq 50$ GeV/c
$L_{\rm SH} < 0.2$
$\chi^2_{\rm CES} \leq 10$
$\lvert \Delta z \rvert < 3$ cm
$-3 < q \cdot \Delta x < 1.5$ cm

Table 5.1: Offline requirements for central electron candidates.

tracker which leave hits in the muons chambers located outside the calorimeter. No specific tau identification algorithm was used. τ leptons decaying hadronically are reconstructed as jets, and leptonic τ decays are not identified, since we veto the electrons and muons reconstructed according to the identification algorithms described in this section.

5.5.1 Electron identification

Electrons deposit most of their energy in the electromagnetic calorimeters. An electron candidate is typically a high-p_T track isolated from other activity in the tracking systems which is matched to an electromagnetic calorimeter cluster. The isolation requirement ensures that the candidate originates from the primary interaction vertex, and does not, e.g., originate from a B hadron semi-leptonic decay.

Details on the selection criteria for central electron candidates (CEM, in the region $|\eta| < 1.1$) are summarized in table 5.1. The trigger requires a transverse energy of the calorimeter cluster of $E_T > 18$ GeV, but we require an electron to have $E_T > 20$ GeV (and thus veto all events with an candidate electron having $E_T > 20$ GeV). $E_{\rm HAD}/E_{\rm EM}$ is the ratio between the energies deposited in the hadron-

$p_T > 20$ GeV/c
$E_{\rm HAD} < 6 + \max(0, 0.0280 \cdot ({\rm p}-100))$ GeV
$E_{\rm EM} < 2 + \max(0, 0.0115 \cdot ({\rm p}-100))$ GeV
$E_{\rm ISO}/E \leq 0.1$
More than three COT axial segments with five hits each
More than two COT axial segments with five hits each
Track $\lvert z_0 \rvert < 60$ cm and $p_T > 10$ GeV/c
$d_0 < 0.2$ cm (0.02 cm with silicon hits)
$\chi^2_{\rm COT} \leq 2.3$
$\Delta x < 7$ cm (CMU), $\Delta x < 5$ cm (CMP), or $\Delta x < 6$ cm (CMX)
$\Delta\rho^{\rm COT}_{\rm CMU} > 140$ cm

Table 5.2: Offline requirements for muon candidate (CMUP and CMX) identification.

ic calorimeters (CHA, WHA or PHA) and in the electromagnetic calorimeters (CEM); this ratio is small for electron candidates, typically of the oder of 5%. We define isolation as the ratio between the energy deposited in the additional towers locate in a cone of radius $R = 0.4$ around the calorimeter cluster ($E_{\rm ISO}$), and the energy of the cluster itself; it is required to be smaller than 0.1. $L_{\rm SH}$ measures the matching between the theoretical electron shower profile and the distribution of energy in the calorimeter clusters. The χ^2 is that from the fit of the shower maxumum profile (from CES or PES) with respect to electron test beam data. Δx ($|\Delta z|$) is the signed difference in x (distance in z) between the track extrapolated to the position of the shower maximum and the calorimeter cluster.

5.5.2 Muon identification

Muons, due to their large rest mass (~ 206 times that of the electron), are minimum ionizing particles, and therefore deposit only small amounts of energy in the calorimeters. Muon candidates are identified by matching a high-p_T track reconstructed from the hits (ionization deposits) in the muon chambers, i.e. a *stub*, with a track from the COT, and requiring small energy deposits in the calorimeters along the trajectory of the particle. The selection requirements for CUMP and CMX muons are

$p_T > 0.5$ GeV/c
At least 10 (20) COT stereo (axial) hits
ΔR(track, track candidate) < 0.4
Δz_0(track, track candidate) < 5 cm

Table 5.3: Good quality track criteria for isolated track selection.

$p_T > 20$ GeV/c
$\lvert\eta\rvert < 1.2$
$\lvert z_0\rvert < 60$ cm
$T_{\text{ISO}} \geq 0.9$
More than 24 COT axial hits
More than 20 COT stereo hits
$\chi^2 > 10^{-8}$
$d_0 < 0.2$ cm (< 0.2 mm with silicon hits)

Table 5.4: Offline requirements for isolated track candidates

summarized in table 5.2. CMUP muons are reconstructed in the region $|\eta| < 0.6$ with the requirement that the track matches stubs in both the CMU and CMP detectors. This requirement reduces considerably the fake muon fraction of the collected sample. CMX muon candidates are reconstructed in the region $0.65 < |\eta| < 1.0$, with a stub in the CMX detector.

5.5.3 Isolated Tracks

Isolated tracks are high-p_T tracks that are isolated from energy deposits in the tracking systems, but with no associated requirements in the calorimeter or muon systems. This recovers charged leptons that traveled in non-instrumented regions of the calorimeter or muon detectors. An isolated tracks event category has been recently added to the WH analysis [103]. In order to maintain orthogonality between the two selections, we remove this category of events.

Starting from a set of good quality tracks, shown in table 5.3, we define the track isolation

$$T_{\rm ISO} = \frac{p_T(\text{track})}{p_T(\text{track}) + p_T(\text{other tracks})} \tag{5.5}$$

where *other tracks* refers to all the tracks within a cone radius of 0.4 of the candidate isolated track. The track is fully isolated if it has $T_{\rm ISO} = 1.0$. This case, however, is very seldom found in a hadronic collision. We therefore consider a track to be isolated if it has $T_{\rm ISO} \geq 0.9$. Additional requirements for isolated tracks are described in table 5.4.

5.5.4 Tau identification

The τ lepton has a very short lifetime (about 290 fs), and decays quickly. Therefore, it must be identified based on its decay products. Several analyses involve τ's, and use dedicated τ reconstruction algorithms [104, 105]. In the present dissertation, however, we make no attempt to reconstruct them.

The τ lepton decays into another lepton (e or μ, with a branching fraction of about 17% for each type) or into quarks, for the large mass of the τ (1.7 GeV/c^2) permits such decays. The latter decay mode is reconstructed as a jet (next section) and contributes to the pool of events with three jets (cf. chapter 11); the former mode is rejected using the electron and muon identification criteria described above.

5.6 Jet Reconstruction

After the hard scattering process has taken place, the quark and gluons coming from that interaction undergo fragmentation, during which they create partons via a cascade of gluon emissions and decays. This process continues until the p^2 of the partons is of the order of the infrared cut-off scale. Then, during hadronziation, the partons form colorless particles which interact with the detector material. A shower

of particles is detected as the result of this process; it appears as a cluster of energy deposited in the calorimeter. We call these particle showers *jets*.

The challenge with analyses dealing with jets is to recover as much detector information as possible to determine the energy, momentum, and type of the initial parton. In other words, the analysts must infer the *parton level* information, which is what theory provides, from the measured calorimeter information. Jets can be reconstructed in many ways, and several algorithms exist for that purpose. There is no universal definition for a jet, for there is no knowledge about how the hadronization process takes place.

From a theoretical point of view, several properties are sought [106], such as insensitivity to soft or collinear radiation (infrared and collinear safety), and to the details of the final state (boundary stability). The algorithm should be equivalent at the parton, particle and detector levels (order independence), and fin the same jets independent of boosts along the jet axis (boost invariance). In addition, the algorithm should be straightforward to implement in perturbative calculations.

On the other hand, there are also *experimental* attributes for jet algorithms [106], among which insensitivity to detector segmentation, energy response, or resolution (detector independence) or insensitivity to multiple hard scatterings at high beam luminosities (stability with luminosity). In particular, jets should not grow to excessively large sizes because of additional interactions. The algorithm should identify all physically interesting jets (i.e. those associated with partons), and not amplify the (inevitable) effects of resolution smearing and angular bias. It should be straightforward to implement, easy to calibrate and should identify jets using minimal computer time. Finally, all details of the algorithm should be fully specified. This includes specifications for clustering, definitions for energy and angles, and jet splitting/merging.

We proceed with the description of the jet reconstruction algorithm used at CDF: JetClu.

5.6.1 The JetClu fixed cone algorithm

At CDF, the most common algorithm is JetClu [107, 108]; it is an iterative fixed cone jet reconstruction algorithm which is based solely on energy deposits in the calorimeter (calorimeter clusters, section 5.4). This is the algorithm used to reconstruct the jets in the analyses presented in this dissertation.

In this family of algorithms, jets are formed by associating calorimeter towers lying within a specific radius (cone) in the $\eta - \phi$ plane. In this dissertation, we use a cone size of

$$R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} = 0.4, \tag{5.6}$$

where $\Delta\eta = \eta_C - \eta_i$ and $\Delta\varphi = \varphi_C - \varphi_i$. The coordinates of the i^{th} calorimetric tower are given by (η_i, φ_i), while the centroid of the jet, defined as the E_T-weighted average of the tower's location within the jet, is located at (η_C, φ_C).

The JetClu algorithm starts by listing all towers with a transverse energy (E_T) larger than 1 GeV/c^2; these are called the *seed* towers. In this context, the transverse energy is defined as

$$E_T = E_{\text{EM}} \cdot \sin\theta_{\text{EM}} + E_{\text{HAD}} \cdot \sin\theta_{\text{HAD}} \tag{5.7}$$

where E_{EM} (E_{HAD}) is the energy measured in the electromagnetic (hadronic) portion of the calorimeter, and θ_{EM} (θ_{HAD}) is the polar angle with respect to the origin (the primary vertex of the event, section 5.3).

The *seed* towers are ordered by decreasing energy. The algorithm then loops over the seeds, starting from the most energetic. It forms a *pre-cluster* by clumping together adjacent seed towers within a cone of $R = 0.4$ in the $\eta - \phi$ plane, such that: (a) every seed tower is assigned to exactly one pre-cluster, and (b) each pre-cluster contains at least one seed tower. For each pre-cluster, the algorithm computes the E_T-weighted centroid, along with a cone of radius R around it.

In a second phase, *clusters* are formed around each centroid derived from a pre-cluster, adding the remaining towers that have $E_T > 0.3$ GeV/c^2 one by one until a stopping criterion is met. At each step the centroid is recomputed, and a new cone defined. The process is stopped when the tower list is un-changed with respect to the previous iteration, or when a certain number of iterations are reached.

Overlapping jets

Since it is possible for clusters to overlap, the algorithm must decide whether to merge or separate the overlapping clusters. If the towers belonging to one cluster are contained within another cluster, the smaller one is dropped. If there is partial overlap between the towers of two clusters, the *overlap fraction* is determined from the ratio of the energy contained in the common towers to that of the smaller cluster. Depending on the value of the overlap cutoff, if it is larger than 0.5, the clusters are merged; otherwise, the clusters are kept separate, and each tower in the overlap region is assigned to the cluster with the nearest centroid. After this process, each remaining cluster is promoted to a jet, which are defined by their tower lists.

Limitations of the JetClu algorithm

The JetClu algorithm was chosen for its execution speed and simplicity. However, there are several drawbacks associated to it. For example, once a tower is assigned to a cluster, it is never removed from it. After several re-computation steps, it is possible for the center of the cone to drift away, so that the tower is no longer within the cluster's (and thus the jet's) cone. This effect is called *ratcheting*. Another important remark is that the JetClu algorithm is not infrared-safe, since it is possible that two jets separated by $R < \Delta R < 2R$ be merged into a single one if a soft gluon between them causes the centroids to drift closer.

The jet four-vector, (E, p_x, p_y, p_z), is then computed from the towers associated to it. The electromagnetic and hadronic parts of each tower are assigned a massless

four-vector with a magnitude corresponding to the energy measured in that tower and a direction pointing from the primary vertex to the location of the shower maximum for that tower. Thus,

$$E = \sum_{i=1}^{N_{\text{TOWERS}}} \left(E^i_{\text{EM}} + E^i_{\text{HAD}}\right) \tag{5.8}$$

$$p_x = \sum_{i=1}^{N_{\text{TOWERS}}} \left(E^i_{\text{EM}} \sin\theta^i_{\text{EM}} \cos\phi^i_{\text{EM}} + E^i_{\text{HAD}} \sin\theta^i_{\text{HAD}} \cos\phi^i_{\text{HAD}}\right) \tag{5.9}$$

$$p_y = \sum_{i=1}^{N_{\text{TOWERS}}} \left(E^i_{\text{EM}} \sin\theta^i_{\text{EM}} \sin\phi^i_{\text{EM}} + E^i_{\text{HAD}} \sin\theta^i_{\text{HAD}} \sin\phi^i_{\text{HAD}}\right) \tag{5.10}$$

$$p_z = \sum_{i=1}^{N_{\text{TOWERS}}} \left(E^i_{\text{EM}} \cos\theta^i_{\text{EM}} + E^i_{\text{HAD}} \cos\theta^i_{\text{HAD}}\right) \tag{5.11}$$

Other algorithms for jet reconstruction There are several other algorithms to reconstruct jets, some of which are much more advanced, and therefore are used in more recent LHC experiments. For the Run II of Tevatron, the Jet Algorithms Working Group was formed to present recommendations to develop a standard set of jet algorithms that are well-defined for both experiment (CDF and DØ) and theory [109].

5.6.2 Jet Energy Corrections

Once a jet is reconstructed following the procedure described in the previous section, its energy needs to be corrected to take into account detector response and multiple interactions [108, 110]. There are several levels for jet correction; these are reported below. A overview of the levels is presented in table 5.5.

Relative Energy (L1) The relative energy correction (C_{RE}) is the first level of correction, and accounts for the non-uniform response of the detector in pseudo-rapidity. The η dependence of the calorimeter output is partly due to the difference in clustering performance between the central and plug calorimeters. Another aspect of this dependence are the un-instrumented regions, or *cracks*,

Level 0	–	Online/Offline calibrations
Level 1	–	Eta-dependent
Level 2	–	Not in use
Level 3	–	Not in use
Level 4	–	Multiple Interactions
Level 5	–	Absolute
Level 6	–	Underlying event
Level 7	–	Out-of-cone
Level 8	–	Splash-out (uncertainty only)

Table 5.5: CDF Jet Energy Correction Levels.

located around $|\eta| = 0$, where the east and west parts of the calorimeter join, and at $|\eta| = 1.1$, where the central joins the plug calorimeter.

The correction for these effects consists in scaling the response of the forward (plug) calorimeter to the scale of the central one, which is better calibrated and understood. The C_{RE} also corrects for the transverse spread of calorimeter showers outside the jet cone, and for any η dependance in multiple parton interactions and gluon radiation.

The correction is obtained from p_T balancing studies in di-jet events. These are selected to have one jet, the *trigger jet*, in the region $0.2 < |\eta| < 0.6$ (i.e away from detector craks), and another jet, the *probe jet*, which is constrained to the region $|\eta| < 3.0$. The p_T balancing fraction

$$f_b \equiv \frac{\Delta p_T}{\langle p_T \rangle} = \frac{p_T^{\mathrm{probe}} - p_T^{\mathrm{trigger}}}{(p_T^{\mathrm{probe}} + p_T^{\mathrm{trigger}})/2} \tag{5.12}$$

is formed and the correction defined as $1/\beta(\eta)$, where

$$\beta(\eta) = \frac{2 + \langle f_b \rangle}{2 - \langle f_b \rangle}, \tag{5.13}$$

and where p_T^{trigger} (p_T^{probe}) is the transverse momentum of the trigger (probe) jet. Mathematically β is equal to $p_T^{probe}/p_T^{trigger}$, but we reduce the sensitivity

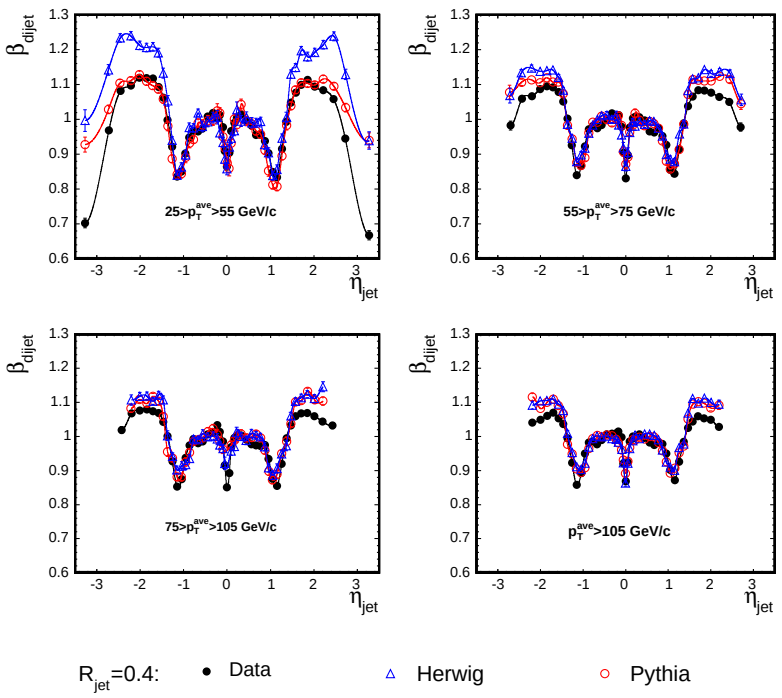


Figure 5.2.: Relative Energy Correction (C_{RE}). Distribution of $\beta(\eta)$ for data, HERWIG and PYTHIA MC samples for four transverse momentum regions.

to non-Gaussian tails by deriving it from f_b, which is approximately Gaussian distributed, contrary to the ratio of momenta.

The correction is derived separately for MC and data in different p_T^{jet} bins. Figure 5.2 shows the distribution of $\beta(\eta)$ for data and MC for four transverse momentum regions before and after the correction is applied. The data samples used for this purpose contain at least one jet with 20, 50, 70, or 100 GeV/c^2. It is applied to the raw jet energies of both data and simulated events, to ensure an homogeneous response across the detector. The uncertainty associated with this correction is estimated between 1% for central jets (figure 5.4b) and 8% for forward jets.

Multiple Interaction (L4) The correction for multiple interactions (C_{MI}) accounts for the extra energy deposited in the calorimeter as the instantaneous luminosity

increases[9]. As the average number of $p\bar{p}$ collisions increases, the hadrons arising from the additional interactions deposit extra energy in the calorimeter, which biases the measured jet energies. The C_{MI} factor subtracts that extra energy.

This effect is estimated from minimum bias events (triggered by the CLC). A random tower in the region $0.2 < |\eta| < 0.6$ of the calorimeter is chosen as a seed tower to form a cluster whose energy is measured. The correction is parameterized as a function of the number of vertices of the event (cf. figure 5.3b), which is an indication of the number of interactions in an event. For a luminosity of 10^{32} cm^{-2}s^{-1}, there is an average number of three collisions (figure 5.3a). This technique works well for events with less than seven vertices, because of the finite reconstruction efficiency of the vertices. The uncertainty on this correction is of the order of 15%.

Absolute Energy (L5) (C_{ABS})

Contrary to L1 and L2 corrections, which operate at calorimeter level, the absolute energy corrections operate at particle level. The procedure to estimate the L5 correction factor [108] uses a PYTHIA [111] MC sample of dijet events to compare *particle jets* (at generator level) to *calorimeter jets* (reconstructed from the detector simulation). These jets have to be within $\Delta R < 0.1$ to ensure theta the same jet is being dealt with. The correction factor is taken to be the maximum probability of measuring a value of p_T^{jet} given p_T^{particle}. The uncertainty on this corrections is of the order of 3.5% (15% at the edge of the calorimeter).

Underlying Event (L6) (C_{UE})

The reconstruction of the energy of the jets from hard $p\bar{p}$ interactions may contain contributions from particles originating from soft spectator interactions, from initial state radiation gluons, or from hadrons from the break-up of the

[9]The number of interactions taking place during the collision of one proton and one antiproton bunch follows a Poisson distribution whose mean increases with instantaneous luminosity.

colliding proton or antiproton. These constitute the underlying event. The $C_{\rm UE}$ accounts for the energy deposited in the jet cone due to these effects.

Out of Cone (L7) ($C_{\rm OOC}$)

A fraction of the energy of the parent parton can be lost from the jet cone due to low-p_T particles bending in the magnetic field, or final state gluon radiation at large angles with respect to the parent parton. This effect is due to the finite size of the jet cone algorithm. We estimate $C_{\rm OOC}$ by studying γ+jet events in data and MC, and comparing the energy deposited in a ring with inner and outer radiuses of 0.4 and 1.3 in the $\eta - \phi$ plane between the data and the simulation. The largest difference is taken as a systematic uncertainty.

Splash Out (L8) The *splash-out* contribution refers to the lost energy that is not accounted for at L7, i.e. that falls outside of the $R = 1.3$ jet cone. Only a rate systematic uncertainty is determined for this effect. In MC samples, this contribution was determined to be small ($\sim$ 0.5 GeV), and this correction is therefore not necessary.

The corrected transverse jet energy is obtained as follows

$$p_T^{\rm parton} = (p_T^{\rm jet} \times C_{\rm RE} - C_{\rm MI}) \times C_{\rm ABS} - C_{\rm UE} + C_{\rm OOC} \tag{5.14}$$

where $p_T^{\rm parton}$ is the transverse momentum of the parent parton, and $p_T^{\rm jet}$ that of the calorimeter jet. The transverse jet energy resolution is given approximately by

$$\frac{\sigma(E_T)}{E_T} = \frac{1.0}{E_T\ [\rm GeV]} \oplus 0.1. \tag{5.15}$$

Figure 5.4b shows the individual fractional systematic uncertainties as a function of jet p_T in the region $0.2 < |\eta| < 0.6$. For $p_T > 60$ GeV/c, the largest contribution arises from the absolute jet energy scale which is limited by the uncertainty of the calorimeter response to charged hadrons. Each of these steps has an individual

systematic uncertainty that is added in quadrature to derive the total uncertainty which decreases from 8 to 3% with increasing jet energy. After these corrections the jet energy provides a good estimate of the initial parton energy. This can be verified comparing the jet energy to the energy of an electromagnetic object such as a prompt photon or a Z boson produced in the same event.

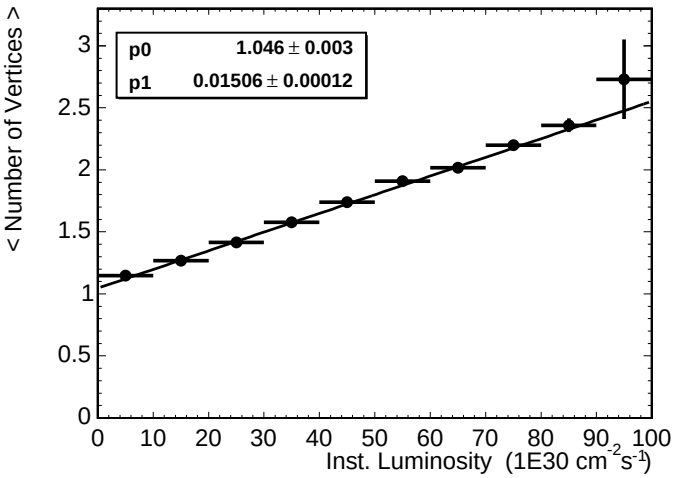
5.6.3 The H1 Algorithm

The jet energy corrections described in the previous section rely on calorimeter information. The calorimeter jet reconstruction is improved by combining momentum measurements of charged particles in the tracking detector with the information provided by the calorimeter energy deposits [112, 113]. Indeed. the particles composing hadronic jets ($\pi^{\pm}$, π^0, $K^{\pm}$, K_S, K_L, protons and neutrons) have mostly low momenta, and thus the tracking system is able to measure their momenta with higher precision.

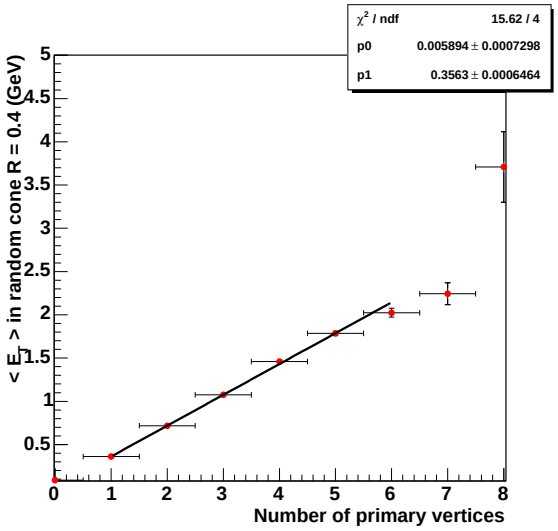
Using the tracking system to reconstruct the hardonic particles, and the EM calorimeter for electromagnetic particles improves the estimate of the true energy of the jet, because the resolution of the hadronic calorimeter is worse than that of its electromagnetic counterpart. However, the insufficient separation between hadronic and electromagnetic showers leads to moderate improvements in jet energy resolution. In this dissertation, jet energies are also corrected using the algorithm developed by the H1 collaboration [113]. This approach provides an improvement of $\sim$ 10% in resolution [114]. The resolution improvement is not very high for Monte Carlo studies, but the benefit comes from a better agreement with data.

5.7 Missing Energy Reconstruction

Neutrinos interact with the surrounding material only through the weak force. They thus escape detection, carrying away some amount of energy. This energy cannot be directly measured, and creates an energy imbalance in the detector. Because the z-component of the momentum of the interacting partons is unknown on an event

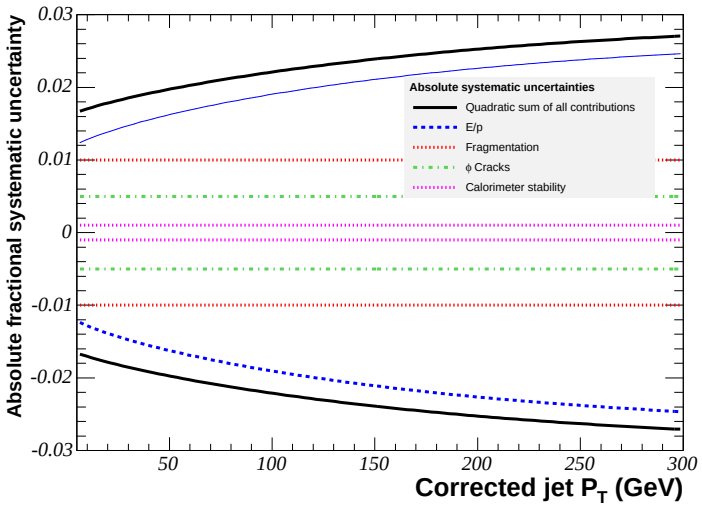


(a)

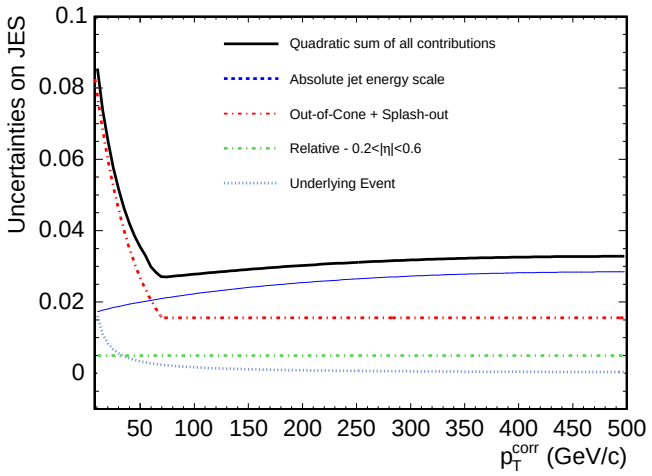


(b)

Figure 5.3.: Distributions of (a) the number of vertices as a function of instantaneous luminosity, and of (b) the $\langle E_T \rangle$ in a random cone of radius $R = 0.4$ as a function of the number of vertices.



(a) Absolute Energy (C_{ABS})



(b) Total JES Correction

Figure 5.4.: Systematic Uncertainties on the JES Corrections

basis, we cannot precisely determine the net energy imbalance. However, the momentum of the partons in the transvese plan is zero in good approximation. This allows to define the transverse component of the missing energy as

$$\vec{\not\!\!E}_T^{\,\mathrm{RAW}} = -\sum_i E_T^i \cdot \hat{n}_i \tag{5.16}$$

where E_T^i is the transverse energy measured in the ith calorimeter tower, and $\hat{n}_i$ is the projection on the plane perpendicular to the beam axis of the vector pointing from the event vertex to the ith calorimeter tower. For first approximation, $\vec{\not\!\!E}_T^{\,\mathrm{RAW}}$ gives the energy and direction in the transverse plane of an undetected particle[10].

In the SM, $\not\!\!E_T$ originates from neutrinos produced during electroweak decays. However, energy mis-measurements or particles entering a crack in the detector also yield apparent missing transverse energy; these effects need to be dealt with.

5.7.1 Jet Energies and $\not\!\!E_T$ Correction

After correcting the jet energies, the $\not\!\!E_T$ needs to be recomputed using the corrected values of the jet energies. The correction is performed according to

$$\not\!\!E_x^{\mathrm{CORR}} = \not\!\!E_x^{\mathrm{RAW}} - \sum_{i=1}^{N_{\mathrm{jets}}} \left(\not\!\!E_x^{\mathrm{CORR},i} - \not\!\!E_x^{\mathrm{RAW},i}\right) \tag{5.17}$$

$$\not\!\!E_y^{\mathrm{CORR}} = \not\!\!E_y^{\mathrm{RAW}} - \sum_{i=1}^{N_{\mathrm{jets}}} \left(\not\!\!E_y^{\mathrm{CORR},i} - \not\!\!E_y^{\mathrm{RAW},i}\right) \tag{5.18}$$

where CORR and RAW refer respectively to the corrected quantities and to those measured directly by the CDF calorimeter, $\not\!\!E_x^{\mathrm{RAW}}$ and $\not\!\!E_x^{\mathrm{RAW}}$ are computed using

$$\not\!\!E_x^{\mathrm{RAW}} = -\sum_{i=1}^{N_{\mathrm{towers}}} E_T^i \cdot \cos\phi_i \;\; \not\!\!E_y^{\mathrm{RAW}} = -\sum_{i=1}^{N_{\mathrm{towers}}} E_T^i \cdot \sin\phi_i. \tag{5.19}$$

[10]This is the case for a single neutrino, as well as for a Z decaying to a pair of neutrinos. The case where two undetected particles are back to back is not discussed here.

The corrected $\not\!\!E_T$ is then recomputed using

$$\not\!\!E_T{}^{\text{CORR}} = \sqrt{\not\!\!E_x^{\text{CORR}\,2} + \not\!\!E_y^{\text{CORR}\,2}} \tag{5.20}$$

and its direction in ϕ obtained from

$$\phi_{\not\!\!E_T{}^{\text{CORR}}} = \tan^{-1}\left(\frac{\not\!\!E_x^{\text{CORR}}}{\not\!\!E_y^{\text{CORR}}}\right). \tag{5.21}$$

5.7.2 Muons and $\not\!\!E_T$ Correction

Muons have only ionization energy losses in the calorimeter and hence leave an apparent missing energy in the detector. For analyses in which this muon is identified, it is necessary to correct the $\not\!\!E_T$ by subtracting the expected amount of energy left in the calorimeter, and adding back the measured transverse momentum of the muon track. But because this analysis only addresses events with no identified muons, we do not need this correction.

5.8 Summary

In this chapter, we have presented the event reconstruction, which provides high-level physics objects such as leptons, jets, and missing energy (i.e. neutrinos). In the the remaining chapters, especially in chapter 11, we use this information to select events according to selection criteria.

The identification of the decay products of the Higgs bosons, i.e. two b quarks, was left out of the discussion on purpose. In the next chapter, we describe the identification of jets originating from b hadrons.

6. HEAVY FLAVOR TAGGING

Once the jets have been identified and properly reconstructed, they can be classified according to their flavor. In this context, light flavor jets originate from the hadronization of an up (u), down (d), or strange (s) quark, or a gluon, while heavy flavor jets originate from a charm (c) or bottom (b) quark. For most physics processes taking place at a collider, the majority of events contains only light quarks in the final state. The Higgs signal we are after, however, consists mainly of two b quarks in the final state – $\mathcal{B}(H \to b\bar{b}) \sim 70\%$ at $m_H = 115$ GeV/c^2; this is also the case for signals with top quarks ($t \to Wb$) or a Z boson ($Z \to b\bar{b}$). Identifying events with b quarks is thus a key step in reducing the signal to background (S/B) ratio. Several algorithms make use of various properties of the decay of b (and c) quarks to identify (tag) heavy flavor jets. This process is commonly referred to as *b-tagging*.

As soon as they are produced, b quarks hadronize almost immediately to form B hadrons (meson or baryon), e.g. B_0 ($d\bar{b}$) or Λ_B (udb). Because the b quark decays through the weak force, these hadrons have a relatively long lifetime, of the order of a few picoseconds. Therefore, they travel a few millimeters away from the primary vertex before decaying through a cascade of particles (cf. figure 6.1). The charged decay products are reconstructed as displaced tracks, which intersect at secondary vertices (where the B hadrons decayed).

This chapter expands on the two algorithms used in the searches presented here, and commonly used at CDF, to identify jets originating from the hadronization of b quarks: the secondary vertex (SecVtx) algorithm and the jet probability (JetProb) algorithm. Alternative techniques can be used to distinguish heavy flavor from light flavor jets. One of these, called *soft lepton tagging* (SLT), is to look for the presence of a soft electron or muon within the jet cone, which indicates a semileptonic decay

of a heavy flavor hadron. However, we do not expand further on these, and direct the interested reader to the appropriate documentation [115, 116].

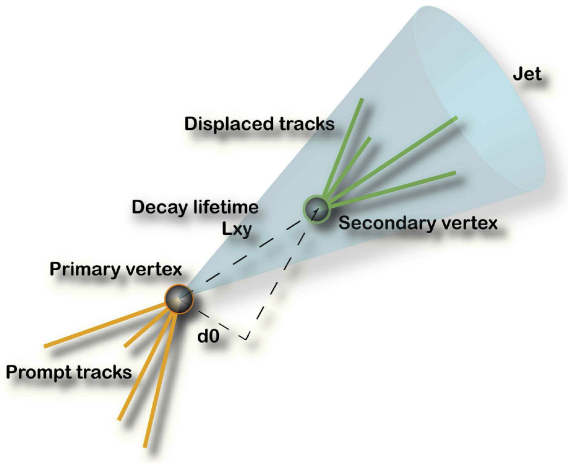


Figure 6.1.: Schematic view of the decay of a B hadron. Credit: The CDF Collaboration

6.1 The Secondary Vertex Algorithm

The CDF secondary vertex (SecVtx) detection algorithm [102] identifies the decay of long-lived hadrons by looking for secondary vertices displaced from the primary vertex of the event. The Run II algorithm is essentially unchanged from that of Run I used to discover the top quark [117], but the track selection has been re-tuned for the CDF II detector [102]. Requiring that one of the jest in the event be tagged by SecVtx keeps more than half of the signal while removing approximately 95% of the background.

In order to achieve its goal, the algorithm needs a precise knowledge of the collision point, or primary vertex (section 5.3). To ensure the rejection of poorly reconstructed tracks, SecVtx requires that all tracks satisfy the following baseline cuts: $p_T > 0.5$ GeV/c, $|d_0| < 0.3$ cm, and $|z_{trk} - z_{PV}| < 5$ cm, where d_0 (Figure 6.2) is the impact parameter of the tracks corrected to the primary vertex, and z_{PV} is the z-position of the event primary vertex. In addition to the baseline cuts, SecVtx makes

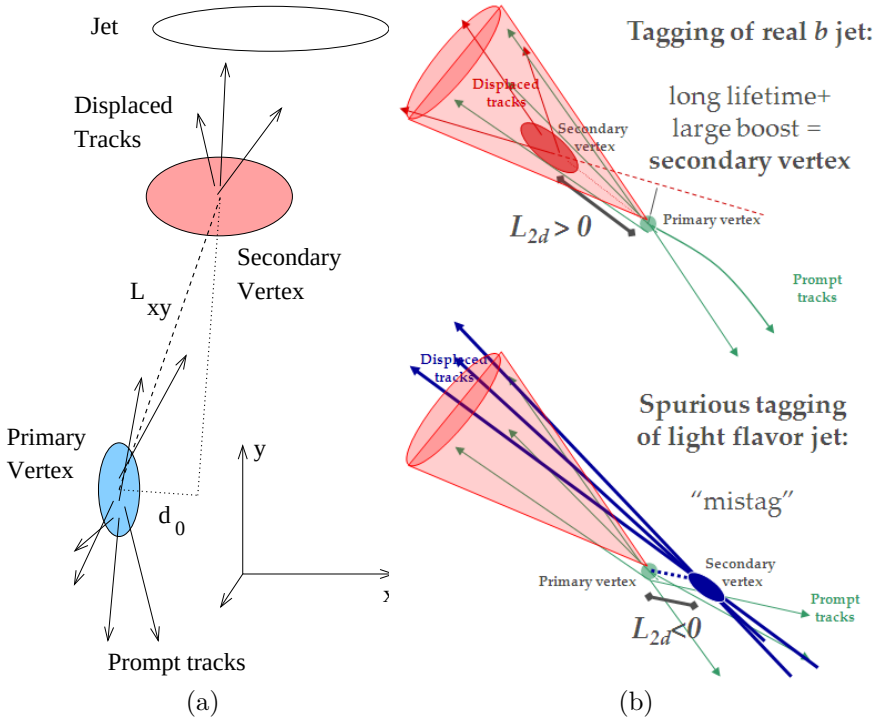


Figure 6.2.: Diagram of a secondary vertex tag, showing the prompt and displaced tracks, with the associated primary and secondary vertices. Credit: The CDF Collaboration

specific requirements on the number and type of COT (section 4.3.1) and SVX (section 4.3.1) hits associated with the tracks, the track fit χ^2/n.d.f. and other track qualities that can discriminate fake or poorly reconstructed tracks. Only jets with at least two of these good tracks can produce a displaced vertex; a jet is considered *taggable* if it has $E_T^{\mathrm{RAW}} > 10$ GeV/c^2, $|\eta| < 2.4$, and has at least two good tracks [102].

The SecVtx algorithm operates on a per-jet basis, considering, for each jet in the event, only tracks within the jet cone ($\Delta R < 0.4$). The selected tracks are ordered in p_T, and a secondary vertex is sought among these tracks. Using silicon tracks confirmed by a track in the COT, SecVtx attempts to locate secondary vertices using a two pass approach. In the first pass, the algorithm uses loose track selection criteria ($p_T > 0.5$ GeV/c and $|d_0/\sigma(d_0)| > 2.5$) and searches for a secondary vertex which includes at least three tracks, of which at least one has $p_T > 1$ GeV/c^2. If this

step doesn't succeed, a second pass attempts to reconstruct a two-track vertex using tighter requirements (one track with $p_T > 1.5$ GeV/c, the other with $p_T > 1$ GeV/c, and both satisfying $|d_0/\sigma(d_0)| > 3$) [102].

The two dimensional decay length of the vertex with respect to the primary vertex, L_{xy}, is the principal variable to discriminate b jets, and is fivent by

$$L_{xy} = \vec{d} \cdot \vec{p_T} \tag{6.1}$$

where $\vec{d}$ is the displacement of the secondary vertex and $\hat{p_T}$ is the unit vector in the direction of the jet momentum. The sign of L_{xy} indicates the position of the secondary vertex relative to the primary vertex, and labels the type of the b-tag. It is positive when the angle between the jet axis and the vector joining the primary vertex to the secondary vertex is less than $\pi/2$; it is negative otherwise. True b hadron decays typically yield a positive value for L_{xy}. A negative value can happen as the result of resolution effects that cause tracks to be mis-measured; negative tags are usually high-p_T light flavor jets (see figure 6.2b). To reduce the contribution from the false secondary vertices, a good secondary vertex is required to have $|L_{xy}/\sigma(L_{xy})| > 3$, where $\sigma(L_{xy})$ is total estimated uncertainty on L_{xy} including the error on the primary vertex position and is roughly 190 microns. Positively tagged light flavor jets are referred to as *mistags*; the negative tags are useful for estimating the rate of fake b jets (cf. section 6.1.2).

6.1.1 Measurement of the tagging efficiency

The detector simulation overestimates the tracking resolution [102]. Therefore, the tagging efficiency is higher in MC than in data. We apply a weighting factor to MC events to compensate for this effect.

A crucial aspect of any selection algorithm is its efficiency. In this context, we need to know how often at least one of the jets in a $VH \to \not\!\!E_T b\bar{b}$ event is positively tagged by SecVtx. Since it is not possible to measure this directly using the signal

sought for, we adopted an different strategy [102]. We use a sample of jets whose heavy flavor fraction can be measured to derive the per-jet tagging efficiency in the data for that sample.

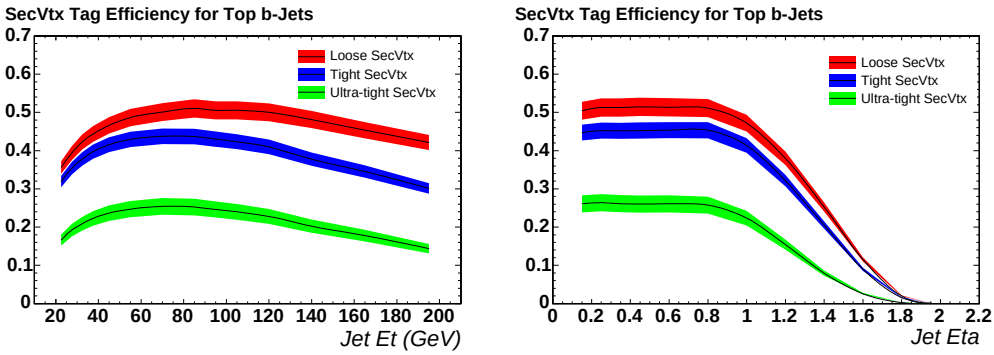
The heavy flavor is a mixture of c and b. We suppress the charm component by requiring a second tagged jet in the event, so that the measured tagging efficiency is dominated by the contribution from b hadrons[1]. Because the jets in the signal have in general different properties (energy, pseudo-rapidity, track multiplicity, etc.), the efficiency cannot be propagated to the region of interest for the final analysis. Instead, we use a matching sample of MC jets and determine the ratio of efficiencies between data and MC[2]. This ratio, the b-tagging *scale factor*, is then used to correct the tagging efficiency in the MC samples. We thus rely on the simulation for the geometrical acceptance and energy dependence of the tagger, and only derive the overall normalization from the data.

We determine the efficiencies from a low-p_T inclusive electron data sample which is enriched in semi-leptonic decays from b and c hadrons. The corresponding MC sample is generated by the HERWIG [118] (cf. section 7.3.3) program. The generated $2 \to 2$ parton events are passed through a filter requiring an electron with $p_T > 7$ GeV/c and $|\eta| < 1.3$, and then processed using the CDF detector simulation.

Electrons in the event are identified (cf. chapter 5) with lower thresholds, i.e. $E_T > 9$ GeV/c^2 and track $p_T > 8$ GeV/c. Electrons are required to be non-isolated and conversions are not removed. The electron track must also pass trough every layer of the SVX detector. Along with the electron, we require two jets in the event: the *electron jet* and the *away jet*. The *electron-jet* is required to have an total energy of $E_T > 15$ GeV/c^2 and be within the jet cone ($\Delta R < 0.4$), and presumably contains the decay products of a heavy flavor hadron. The *away jet* must be energetic ($E_T > 15$ GeV/c^2), within COT coverage ($|\eta < 1.5$) and approximately back-to-back

[1] A variation of the double-tagging technique using events with a single tag yields results compatible with the double-tagged case [102].

[2] For this reason, the scale factor derived from the double-tagging technique can be used as an efficiency for the single tagged sample (within uncertainties).



(a) (b)

Figure 6.3.: Tagging efficiencies for jets matched to b quarks in $t\bar{t}$ MC samples for the three operating points of SecVtx as a function of the jet E_T and η. At high jet E_T, the efficiency decreases as a result of the declining yield of silicon tracks passing the quality cuts. Source: [102]

with the *electron jet* ($\Delta\varphi > 2$). Details on the derivation of the efficiency are provided in ref. [102]. For the dataset corresponding to 7.8 fb^{-1} of data, the scale factor for the tight operating point is measured to be $\epsilon_{MC}^{SV} = 0.96 \pm 0.05$.

Three predefined operating points for SecVtx are used at CDF, with an increasing purity of the selected b jets at the cost of decreased efficiency (cf. next section); these are the Loose, Tight and Ultra Tight operating points. Here, we operate SecVtx at the Tight level, which achieves a good compromise between purity and efficiency. The efficiencies to tag jets which have been matched to b in $t\bar{t}$ MC samples for these operating points are shown in Figure 6.3. The efficiency is obtained by multiplying the tag rate for such jets in the MC by appropriate data/MC scale factors. The bands represent the systematic uncertainty on the scale factors. At high jet E_T, the efficiency decreases as a result of the declining yield of silicon tracks passing the quality cuts.

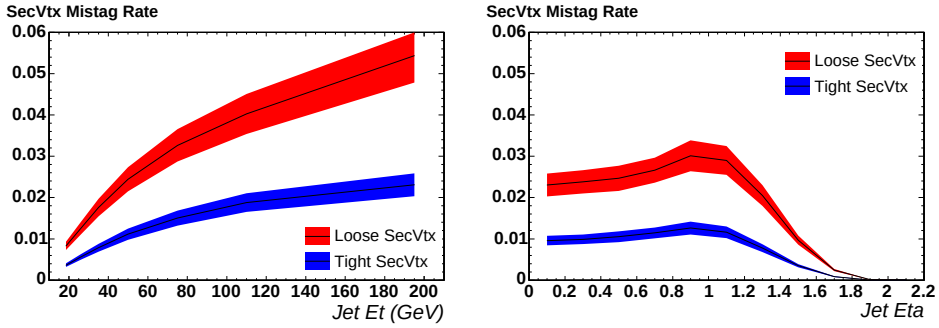
6.1.2 Measurement of the mis-tag rate

A *mistag* is defined to be a light flavor jet, i.e. a jet not resulting from the fragmentation of a heavy quark, with a SecVtx tag. It is often caused by random overlap of tracks which are displaced from the primary vertex due to tracking errors, although there are also other sources, such as K_S and Λ decays, or nuclear interactions with the detector material (beam-pipe, inner silicon layers). The contribution from mistags is directly determined from inclusive jet data samples[3], and does not rely on the detector simulation.

Because SecVtx treats $\sigma(d_0)$ and $\sigma(L_{xy})$ in a symmetric way, the mistags due to tracking errors should occur at the same rate for $L_{xy} < 0$ and $L_{xy} > 0$, and thus the negative mistag rate can provide an estimation of the positive mistag rate. The rate of negative tags for taggable jets is determined in an inclusive sample of jet triggers. It is parameterized as a function of four jet variables – E_T, track multiplicity, η and φ – and one event variable, $\sum E_T$, the scalar sum of the transverse energy every jet in the event with $E_T > 10$ GeV/c^2 and $|\eta| < 2.4$. The parameterized rates give the probability that a given jet will be negatively tagged [102].

This naive estimation of the mistag rate must be corrected for negative tags occurring in jets originating from heavy flavor, and that contribution must be subtracted. In addition, the negative tag rate does not account for lifetime effects nor for interactions with the detector material. The sum of the necessary corrections is found to be $20 \pm 10\%$, consisting in the subtraction of 20% for removal of the heavy flavor negative tags, and the addition of 40% to accound for the mistags due to lifetime and material interactions [102]. Figure 6.4 shows the mistag rate for jets originating from the fragmentation of a light quark for the tight and loose operating points.

[3]This inclusive jet sample is the union of the 20, 50, 70 and 100 GeV/c^2 jet samples (cf. section 5.6.2).



(a) (b)

Figure 6.4.: Mistag rates for jets originating from the fragmentation of a light quark, as a function of the jet E_T and η. These are obtained from inclusive jet data. Source: [102]

6.2 Jet probability algorithm

The jet probability b-tagging algorithm [119] uses tracks associated with jets to determine their probability to originate from the primary interaction vertex. The calculation is based on the impact parameter (d_0) of the tracks in the jets and their uncertainties. The sign of the impact parameter is assigned depending on the position of the track's point of closest approach to the primary vertex with respect to the jet direction; d_0 is positive (negative) if the angle φ between the jet axis and the vector connecting the primary vertex to the track's point of closest approach to the primary vertex is smaller (bigger) than $\pi/2$, as shown in Figure 6.5. By construction, the probability for a track to originate from the primary vertex is uniformly distributed between 0 and 1. For a heavy flavor jet, however, the distribution peaks at 0, since from the long-lived particles have a large impact parameter with respect to the primary vertex.

Particles inside a light flavor jet originate from the primary vertex, but the associated tracks are recomputed with a nonzero impact parameter due to the finite resolution of the tracker, and have an equal probability to be positively or negatively signed. Heavy flavor jets, on the contrary, contain tracks displaced in the jet direction (from the long-lived hadrons); these tracks preferentially populate the positive side of the signed impact parameter distribution.

The tracking resolution is extracted from the data by fitting the negative side of the signed impact parameter distribution of tracks from prompt jets, the dominant component of inclusive jet data. The tracks are divided into 72 different categories according to the number and quality of SVX hits, detector η, and p_T. The signed impact parameter significance, $S_{d_0} = d_0/\sigma(d_0)$, is parameterized for each track category separately, to minimize the contribution from badly measured tracks with a large impact parameter. The tracks are fitted to a helix, and the impact parameter is corrected for beam offsets in order to take into accounts any displacement of the primary vertex from its nominal position. The impact parameter significance is

parameterized for good quality tracks associated with jets with $E_T > 7$ GeV/c^2 and $|\eta| < 2.5$. Good quality tracks are those with $p_T > 0.5$ GeV/c, $|d_0| < 0.1$ cm (to reject long-lived K's and Λ's), 3 to 5 SVX axial hits, at least 20 (17) axial (stereo) COT hits, and $|z_{trk} - z_{PV}| < 5$ cm. The vertex with highest sum of transverse momentum of all tracks is chosen to be the primary vertex, and is used to determine d_0.

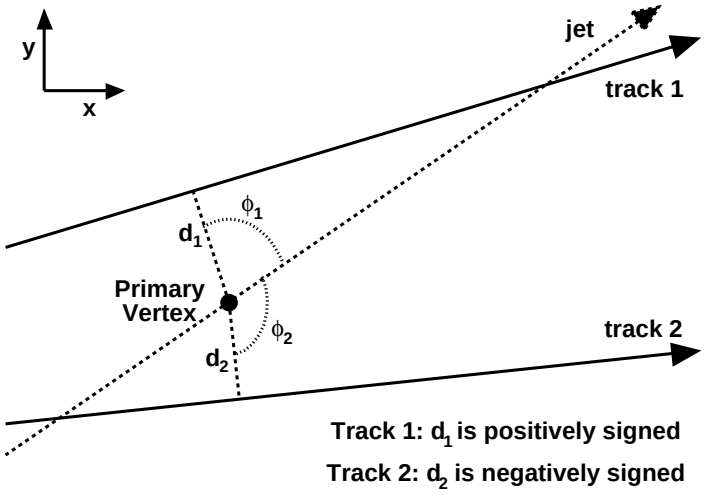


Figure 6.5.: Definition of the sign of the impact parameter of a track. It is positive (negative) if the angle φ between the jet axis and the vector connecting the primary vertex to the track's point of closest approach to the primary vertex is smaller (bigger) than $\pi/2$. Source: [119]

Figure 6.6 shows the distribution of the impact parameter significance for tracks with at least 5 good SVX hits, $p_T > 5$ GeV/c, and $|\eta| < 0.6$. The negative side of this distribution is fitted to the following resolution function

$$R(\Sigma) = \sum_{i=1}^{4} \frac{p_i}{\sqrt{2\pi}\sigma_i} \exp^{-\Sigma^2/2\sigma_i} \tag{6.2}$$

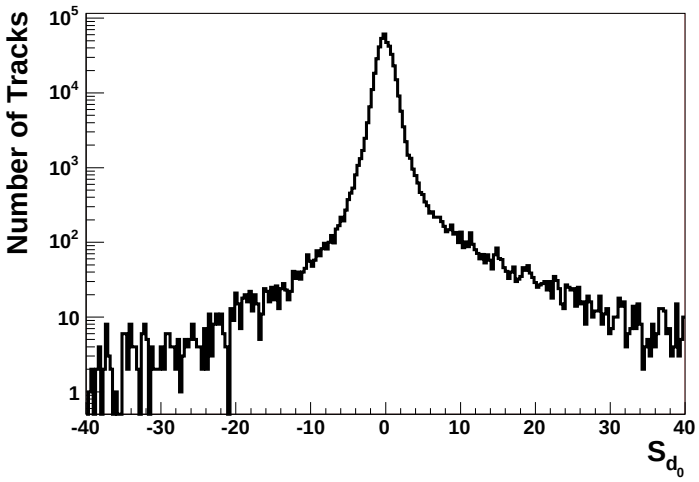


Figure 6.6.: Distribution of the impact parameter significance for tracks in an inclusive jet sample with at least 5 good SVX hits, $p_T > 5$ GeV/c, and $|\eta| < 0.6$. Source: [119]

which is found to fit well the distributions for all the 72 track categories. From it, the probability that the impact parameter significance of a given track is due to detector resolution is defined as follows:

$$P_{\rm tr}(\sigma(d_0)) = \frac{\int_{-\infty}^{-|\sigma(d_0)|} R(\Sigma)d\Sigma}{\int_{-\infty}^{0} R(\Sigma)d\Sigma}. \tag{6.3}$$

Because the tails of the distribution of the impact parameter significance are rather long, we improve the statistics and obtain a better fit use nonlinear bins from applying the transformation $\sigma(d_0) \to X = \ln|\sigma(d_0)|$ and using only the negative part of the distribution [119].

The jet probability P_J that a jet is consistent with a zero lifetime hypothesis is defined as

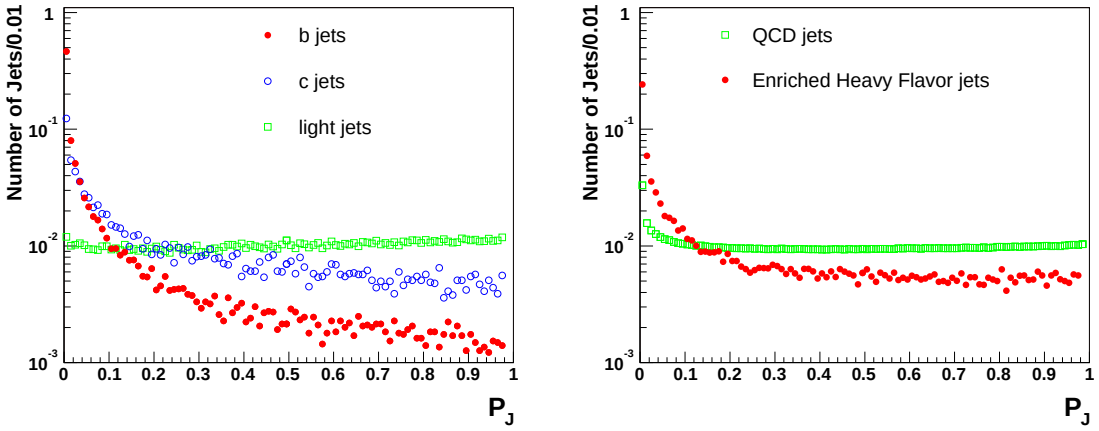
$$P_J = \prod_{1}^{N_{\rm trk}} P_{\rm tr} \times \sum_{k=0}^{N_{\rm trk}-1} \frac{\left(-\ln \prod_{1}^{N_{\rm trk}} P_{\rm tr}\right)^k}{k!} \tag{6.4}$$

where $N_{\rm trk}$ is the number of good quality tracks with positive impact parameters.

Jets are required to have at least two good quality tracks with positive impact parameter to be *taggable*. Tracks with negative impact parameter are used to define a negative P_J, which is used to estimate the misidentification rate (section 6.2.2).

The JetProb algorithm considers a jet to be positively (negatively) tagged when the positive (negative) P_J is less than a given cutoff. Jets with a positive tag are expected to be enriched in heavy flavor. Several operation points are defined for the JetProb algorithm, of which the 0.5%, 1% and 5% points are the most commonly used. The 1% level has similar performance to the SecVtx tagger (section 6.1) while the looser 5% level was chose near the point where P_J becomes flat. It is a feature of this algorithm that the b-tagging is performed using a continuous variable; it allows one to move continuously along the efficiency curve and select the optimal point for a specific analysis. This method can potentially be used to separate b and c heavy

flavor contributions, as shown in Figure 6.7a. In the following, however, we do not make use of these features.



(a) (b)

Figure 6.7.: (a) JetProb distributions for jets matched to b (full circles), c (empty circles) and light (empty squares) quarks in MC simulated events; (b) JetProb distributions for electron jets in inclusive electron data (full circles) and for generic QCD jets in Jet50 data (empty squares). Source: [119]

6.2.1 Measurement of the tagging efficiency

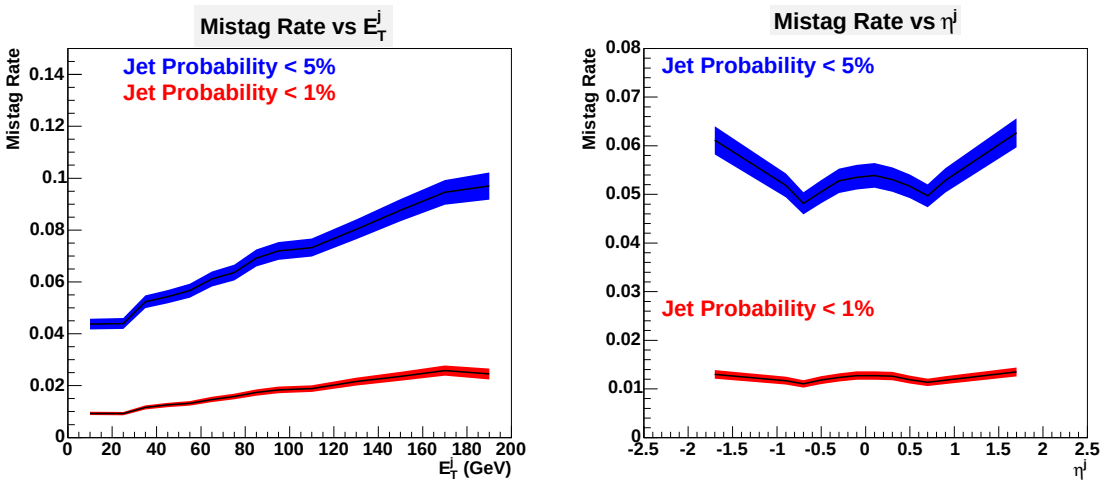
The method to estimate the tagging efficiency for JetProb follows that used for SecVtx (see section 6.1.2 and [102]). We use a calibration data sample of jets whose heavy flavor fraction can be measured: a sample triggered on low-p_T electrons, which is enriched in semileptonic decays of b and c hadrons. Again, we do not determine the tagging efficiency directly, but rather a *scale factor* to apply to the MC simluation. This factor was measured to be $\epsilon_{MC}^{JP} = 0.78 \pm 0.04$.

6.2.2 Measurement of the mistag rate

A mistag is defined as a light flavor jet wrongly tagged as heavy flavor. The probability for assigning a positive tag to a light jet is closely related to the negative

tag rate. We assume that the negative tags are solely due to detector resolution effects, while positive tags have a contribution from real heavy flavor. Thus, under this assumption, the mistage rate is roughly requal to the negative tag rate. In reality, there is a smal contribution from heavy flavor jets to the negative tag rate, and contributions from K's, Λ's, and nuclear interactions with the detector material to the positive tag rate.

The tag rate has considerable dependence on jet kinematics, and is therefore parameterized as a 6-dimensional tag rate matrix of the E_T of the jet, the number of good quality tracks ($N_{\rm trk}^{JP}$), the $\sum E_T$ taken over all the jets in the event, the detector η of the jet, the z vertex position $z_{\rm vtx}$, and the φ of the jet. As was the case with SecVtx we use four inclusive jet sample obtained by requiring the E_T of the leading jet to be greater than 20, 50, 70 and 100 GeV/c^2 respectively. Figure 6.8 shows the negative tag rates for two cut values on P_J as a function of the jet E_T and η.



(a) (b)

Figure 6.8.: Mistag rate for JetProb cuts of 1% and 5% as a function of jet (a) E_T and (b) η in inclusive jet data sample. The bands represent the statistical and systematic uncertainties added in quadrature. Source: [119]

To properly estimate the positive mis-tag rate, we need to determine the corrections for effects from K's and Λ's, or interactions with the detector material. For

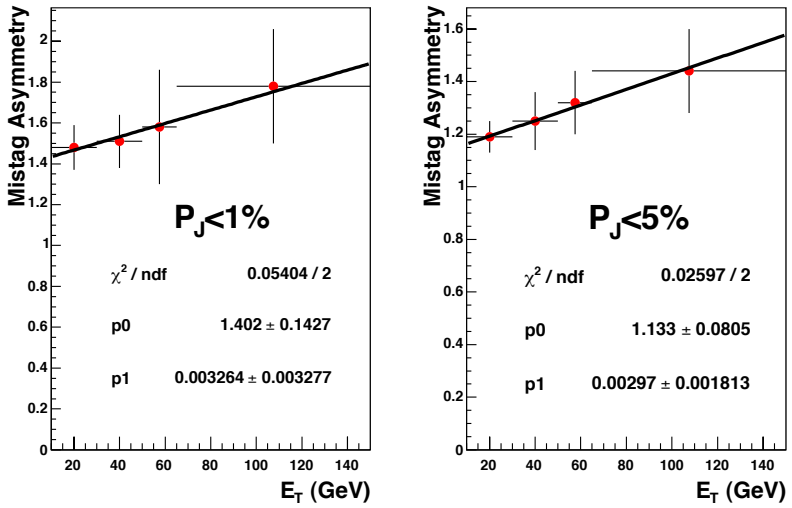


Figure 6.9.: Assymmetry factor for the JetProb mis-tag rate. Source: [119]

.

this purpose, we study the flavor composition of tagged jets in data. This is done by building a variable sensitive to the flavor content of the jet itself, and obtaining the relative contribution from heavy and light partons in data by fitting the distribution of this variable for tagged jets in data to MC simulated templates for b, c, and light jets. The data sample is selected by requiring a jet with $E_T > 50$ GeV/c at the trigger level; the MC templates are obtained from Herwig [118] (cf. section 7.3.3), generating $2 \to 2$ processes with an outgoing parton with $p_T > 40$ GeV/c. The fit is made more robust by fitting the positive excess only – for which the distributions for negative tags are subtracted from the positive side – to remove contributions to the mis-tags due to detector resolution effects, which could be poorly simulated. For the 5% JetProb cut level, we obtain a correction factor of 1.27 ± 0.17 [119].

The asymmetry exhibits a small dependence with jet E_T, as shown in Figure 6.9. We do not account for this small dependence.

6.3 Recent advancement in b-tagging

Today, the most advanced b-tagging algorithm employ multivariate analysis to combine multiple sources of information, in order to improve the efficiency of b-quark

identification. There are two categories of known implementations: *flavor separators* and *multivariate taggers*.

Flavor separators, such as KIT [120] and CHAOS [121] separate light flavor and heavy flavor jets in order to improve the purity of the b-tagged samples. Additionally, CHAOS separates c quarks from b quarks, which is useful because c quarks tend to be more often b-tagged. We investigated the possibility of using a loose operating point for SecVtx together with a cut on the CHAOS output, but the improvement in sensitivity is limited.

Multivariate taggers, such as the NN-tagger used in ref. [122, 123] and the BDT-tagger used by DØ [124] combine information from multiple sources to improve the b quark identification. As was the case for the output of the JetProb algorithm discussed in section 6.2, the output of multivariate taggers can be provided as is to multivariate discriminants, which is expected to improve their sensitivity. In practice, it is common to first define *operating points* to remove events which are not expected to contain b-jets. However, the possibility of using the output of the tagger leads the analyzers to choose looser selections, which increase the efficiency but also the mis-tagging rate. It is not clear whether this approach is best, for the backgrounds subject to mis-tags have a cross-section that is much larger than that of the signal.

6.3.1 Higgs-Oriented B Identification Tagger

In 2011, CDF developed a multivariate tagger, the Higgs-Oriented B Identification Tagger (Hobit) [125], which is based on the inputs to the SecVtx and JetProb algorithms, but also on additional variables. The main motivation is to improve the efficiency of the current CDF taggers, while keeping similar mis-tag rates. Several operating points have been defined. Because this tool was not fully validated at the time of writing, it is not included in this dissertation. We briefly discuss the benefits to this analysis in terms of sensitivity improvement.

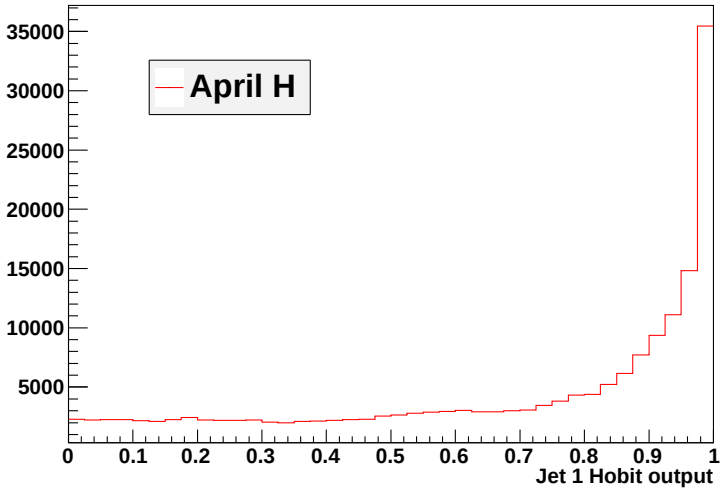


Figure 6.10.: Output of the preliminary version of the HOBIT tagger for $ZH \to \nu\nu b\bar{b}$ MC events in the event selection of the analysis.

Figure 6.10 shows the output of a preliminary version of the HOBIT tagger. We define the tight (loose) operating point by requiring $Hobit > 0.865$ ($Hobit > 0.375$) and assign it the label H (h). Table 6.1 shows the effect of replacing the SECVTX (JETPROB) tagger by the tight (loose) operating point. We thus replace the 1S, SS and SJ categories (defined in section 11.2.5) by the 1H, HH and Hh categories, respectively. The signal acceptance increase using these new categories is of the order of 45% for HH and 10% for Hh and $1H$. The increase in significance $S/\sqrt{B}$ is about 10% for 1H and HH. The Hh category sees a decrease of 8% in significance; this is due to the fact that the h operating point is too loose, and collects too much light flavor mis-tags. A new loose operating point will be chosen for inclusion to the next iteration of the analysis, together with an improved version of the tagger.

6.4 Summary

In this chapter, we have described the two algorithms used in this dissertation to identify jets originating from b quarks. These algorithms are combined to define

Table 6.1: Effect of replacing the SecVtx (JetProb) tagger by the tight (loose) Hobit tagger (see text for details).

$\not\!\!E_T$+b-jets 7.8 fb^{-1}: Signal Region [CDF II Preliminary]

Sample	SS	HH	HH/SS
$ZH \to \nu\nu b\bar{b}$[1]	3.63	5.33	1.46
WW	0.93	1.20	1.29
WZ	15.43	21.91	1.41
ZZ	17.55	25.53	1.45
$t\bar{t}$	197.33	286.85	1.45
Data	5018	8443	1.68
Sample	SJ	Hh	Hh/SJ
$ZH \to \nu\nu b\bar{b}$[1]	2.82	3.12	1.10
WW	3.29	2.57	0.78
WZ	13.80	15.06	1.09
ZZ	15.68	17.08	1.08
$t\bar{t}$	180.41	185.87	1.03
Data	10037	13905	1.38
Sample	1S	1H	1H/1S
$ZH \to \nu\nu b\bar{b}$[1]	8.68	9.50	1.09
WW	183.07	209.74	1.14
WZ	90.90	103.59	1.13
ZZ	62.69	70.13	1.11
$t\bar{t}$	942.23	1096.00	1.16
Data	157494	158917	1.00

[1] $m_H = 125$ GeV/c^2

flavor categories (cf. section 11.2.5) which group events according to their b quark content in order to increase the signal significance in the corresponding sample.

We also discussed HOBIT, a multivariate tagger that has the potential to improve the sensitivity of the current analysis. It will be used in the next iteration of this analysis, using the final CDF dataset.

Now that we have established the experimental procedures, we proceed with the description of the event simulation, which provides the events to be compared to the experimental data.

7. EVENT SIMULATION

Event simulation consists in the prediction of the distribution of signal and background events for different observables. These distributions are compared statistically with the collected data, e.g. using a Bayesian likelihood method (cf. chapter 12). Depending on the outcome of this comparison, we either strengthen our confidence in the underlying assumptions of the event simulation, i.e. a given theory and its parameters, or reject them.

But before the statistical interpretation, we must present the many steps that bring us from a given theory (a Langrangian and a set of rules) to the predicted distributions of signal and background events.

7.1 Monte Carlo Simulation

Monte Carlo (MC) simulation is a fundamental tool in high energy physics experiments, and in many other fields of research[1] The general idea is to generate a large number of $p\bar{p}$ collisions according to the prescriptions of a given theory, in order to mimic the collisions taking place in the detector. The programs used for this purpose are *event generators*. These simulated events, however, cannot yet be compared to data, for they do not carry information about how the detector components (really the act of measuring something) modify the observed distributions. Detector effects are added during the phase of *detector simulation*. Note that for convenience, the term *event simulation* is often used to refer either to event generation or detector simulation.

[1]In this chapter, we will restrict the definition of MC simulation to generating collisions, either beam-beam ($p\bar{p}$) or particle-detector. In chapter 12, we present another use for MC techniques.

Event generation and detector simulation involve many complex aspects. Not only is it impossible (hard, at best) to fully include *all* known (and unknown) effects into the simulation – since there are many holes in our understanding of the complete structure of collision events – but it is often necessary to use reasonable simplifications to model some effects in a limited amount of time. Therefore, choices and compromises must be made.

In the next section, we describe several aspects that need to be considered when generating MC events and performing a detector simulation.

7.2 The Generation of Simulated Events

The generation of simulated events is conveniently separated into several steps: the event generation and the detector and trigger simulation.

7.2.1 Event generation

To first approximation, physics processes involving the interaction between the fundamental objects of nature (quarks, leptons, and gauge bosons) have a simple structure, e.g. $q\bar{q} \to Z^0 \to e^+e^-$. But as often, corrections to this structure are needed.

Phenomenological approaches are needed in order to overcome the complexity of the inherent physics of $p\bar{p}$ scattering (i.e. QCD). The subdivision of the simulation task in different parts as done in MC event generators is theoretically justified by the factorization theorem for hard processes [126, 127], which states that a scattering process with a hard scale can be factorized into distinct stages: (a) the *hard sub-process*; (b) the *initial- and final-state radiation*; (c) the *hadronization (fragmentation)*, and (d) the *beam-remnant fragmentation*.

The hard sub-process involves the hard scale and can be computed by perturbative expansion; it determines the main characteristics of the event. High order

corrections can be added to the picture. These are usually a combination of loop graphs and the soft parts of the initial-state radiation needed to cancel some divergences. The perturbative computations are usually very difficult, and it is hard to include all aspects of more than one non-trivial order (i.e. more than one loop).During the last decade, much work has been done to incorporate higher-order effects, some of which are included in the analyses in this dissertation. The inclusion of these high-order terms is imperative for high-precision studies. When these are not available explicitly, phenomenological models are used to add them in the further simulation stages.

Initial- and final-state radiation involves bremsstrahlung-type effects, namely the emission of particles by branchings such as $e \to e\gamma$ or $q \to qg$. These effects can happen in every process containing charged and/or colored objects in the initial or final state. They give rise to large corrections to the topology of the event [111].

The bulk of these corrections is universal, for they do not depend on the details of the process under study; they depend instead on the momentum transfer scale of the process. Exact derivations may be carried out order by order in perturbation theory but they become more and more complicated. Instead, it is common to use consecutive branchings ($q \to qg$, $g \to q\bar{q}$, $g \to gg$, etc.), generating a *parton shower*.

During this process a single initial parton may give rise to a large number of partons. In particular, the QCD emission of quarks an gluons is very prolific, due to the large value of the strong coupling constant α_s and triple gluon vertices.

When considering QCD radiation, the final-state radiation stage provides the final partonic configuration of the event, and tends to simulate high-order effects to the cross-section and often includes partly-known high-order terms. In the case of QED radiation, the showers only contribute to the global properties of the event, but do not affect the partonic content.

Hadronization is the process by which colored partons become (colorless) hadrons. This stage is often subdivided into a fragmentation step, i.e. the process of breaking a

high-mass system into ones with lower mass (hadrons) through the creation of quark-antiquark pairs, and a decay step, modeling the decay of the hadrons. Fragmentation is a non-perturbative process that has yet to be fully understood from first principles, and therefore only phenomenological models are available to obtain the hadronic final state from the partonic configuration of the event.

The top-quark, with a mean lifetime of ~ 0.5 ys ($\sim 5 \times 10^{-10}$ fs), is the only quark that decays before hadronizing. It is the only quark that can be studied from its decay products, for the other quarks become part of hadrons, which in turn decay.

Beam-remnant fragmentation refers to the process by which the beam-remnants are fragmented along with the output of the hard interaction. The initial-state radiation algorithm reconstructs the shower initiator by backward evolution from the hard sub-process. This initiator only takes some fraction of the beam energy, leaving behind a *beam-remnant*. For instance, a u-quark initiator leaves behind a ud diquark remnant. The latter has an anti-triplet color charge and is color-connected to the hard interaction. It is hence part of the same fragmenting system, and needs to be connected with the rest of the event.

More complex, or even multiple, remnants can also occur. For example, a gluon initiator (quite common at the Tevatron and much more so at the LHC) leaves a more complicated remnant, namely a uud proton remnant system in a color octet state, which is conveniently subdivided into a color anti-triplet diquark and a color triplet quark, both color-connected to the hard interaction. The energy and momentum sharing between these two parts are additional degrees of freedom, which are not understood from first principles.

Since a hadron collider smashes particles of composite nature, there is a possibility that several parton pairs undergo hard scattering separately. At the present there is limited understanding of these mechanisms, and generators resort once more to phenomenological models and parameterization based on existing data to include these effects [111].

7.2.2 Detector and trigger simulation

The events generated with the above considerations are passed through the detector and trigger simulation, incorporating the response of the experimental apparatus. The output has the same format as the recorded data, but also includes information about the underlying truth, or right answer (the generated event). This information is stored in the form of a list of the particles that where generated by the simulation.

Using the same storage format greatly simplifies the analysis process, and means that every reconstruction algorithm and selection criterion is applied on both the real and simulated data. The only difference in treatment is the application of a parameterization of the trigger system (chapter 10) efficiency to the simulated events (in the form of a per-event weight, corresponding to the probability to be accepted by the real trigger). Additionally, the presence of the simulated truth allows tuning of the analysis technique, and to cross-check whether a tool works as expected; something that cannot be done so easily with real data.

7.3 Event Generators

Multi-particle production is the most characteristic feature of current high-energy physics. Today, observed particle multiplicities are typically between ten and a hundred, and with future machines this range will be extended upward. The bulk of the multiplicity is found in jets, i.e. in collimated bunches of hadrons (or decay products of hadrons) produced by the hadronization of partons, i.e. quarks and gluons.

Several code libraries exist to model multi-particle production. They stem from various approaches to solve the complex problem of event generation. We briefly describe the codes used to obtain the results presented in this dissertation.

7.3.1 Pythia

Pythia [111] generates multi-particle production in collisions between elementary particles. It contains theoretical and phenomenological models for hard and soft interactions, parton distributions, initial- and final-state parton showers, multiple interactions, fragmentation and decay.

For the treatment of the hadronization process, the Lund string model [128,129] is implemented. In this model the color field between final state quarks and antiquarks is implemented by color flux tubes or strings. If the potential energy in such a tube is high enough, it can be transformed in new quark-antiquark pair, producing colorless hadrons. Pythia additionally provides several sets of parameters, or *tunes*, which are optimized to model several aspects of collision data like underlying event structure or the contribution from pile-up collisions.

7.3.2 Alpgen

The Alpgen [130] event generator belongs to the class of parton-level event generators, which are a subclass of the matrix element even generators. It is based on the exact leading order evaluation of tree-level matrix elements for a fixed number of partons in the final state. It generates standard model processes in hadronic collisions, with emphasis on final states with large jet multiplicity. We therefore use this generator to model the production of a W/Z boson in association with jets.

Alpgen includes the b and top quark masses (c quark masses are implemented when necessary), as well as top quark and gauge boson decays with helicity correlations. It does not include any form of hadronization and thus the final states only contain leptons, quarks and gluons. Another event generator, Pythia, is used for the showering and hadronization, which produce the observable particles through fragmentation, gluons splitting and particle decay.

Because it is possible to obtain a kinematic configuration with n final state partons starting from $n-m$ partons generated by the tree-level matrix element generator and

generating m partons in the shower, we ought to avoid double counting of certain parts of the phase space. We follow the MLM prescription [131] for the matching of matrix element hard partons and jets generated by the shower.

7.3.3 HERWIG

HERWIG [118] is a general purpose event generator for the simulation of lepton-lepton, lepton-hadron, and hadron-hadron collisions; it includes a large range of hard scattering processes. It implements initial- and final-state radiation using the angular-ordered parton shower, hadronization and hadron decays, and underlying event simulation.

HERWIG is particularly advanced in its treatment of the decay of unstable resonances, and includes full spin correlations for most processes. The program contains a large library of hard $2 \to n$ scattering processes for both the standard model and its supersymmetric extension. Particular emphasis lies thereby on the detailed simulation of QCD parton showers.

Additional features of HERWIG is the QCD jet evolution with soft gluon interference considered via angular ordering. The cluster model used by HERWIG exploits the preconfinement property of perturbative QCD [132] to form color-neutral clusters which decay into colorless hadrons.

7.3.4 MADEVENT

MADEVENT [133,134] is a multi-purpose tree-level generator which is powered by the matrix element generator MADGRAPH [135]. Given a standard model process, MADGRAPH automatically generates the amplitudes for all relevant sub-processes and produces the mappings for the integration over the phase space. This process-dependent information is passed to MADEVENT and a stand-alone code is produced that allows the user to calculate cross sections and to obtain simulated events.

Because MadEvent is designed to produce events at parton level, a parton showering software is needed to compute all desired final state (colorless) particles. For this purpose Pyxtra was used, a software interface that passes on the MadEvent output to Pythia. MadGraph/MadEvent provides several approaches of jet-parton matching, including the MLM [131], CKKW [136] prescriptions [137].

7.3.5 Powheg

The Powheg method is a prescription for interfacing NLO calculations with parton shower generators. It was first suggested in ref. [138] and is described in great detail in ref. [139]. Powheg produces parton level events with positive (constant) weight with NLO accuracy. The Powheg-Box is a framework for implementing NLO calculation in shower MC programs according to the Powheg method. This method was developed with the goal to solve the negative event weight problem of MC@NLO [140], another method to achieve the systematic inclusion of complete NLO corrections for a fixed multiplicity.

It is easily interfaced with any Monte Carlo showering program, such as Pythia. The method is very similar to traditional matrix-element re-weighting, and thus is relatively simple to implement in a process-independent way; it is well suited for the systematic inclusion of NLO QCD corrections to arbitrary processes [141].

It currently implements the following processes: (a) Z pair hadro-production; (b) heavy-flavor production; (c) e^+e^- annihilation into hadrons and top pairs; (d) W' production; (e) Drell-Yan vector boson production; (f) Higgs boson production (via gluon fusion, vector boson fusion, and in association with a vector boson); (g) single--top production (s-channel, t-channel and associated tW channel), and (h) recently vector boson plus one jet production.

7.3.6 MCFM

MCFM is a parton-level Monte Carlo program which gives NLO predictions for a range of processes at hadron colliders [142,143]. The program is designed to calculate cross-sections for various fb-level processes at hadron-hadron colliders. For most processes, matrix elements are included at next-to-leading order and incorporate full spin correlations.

7.4 Detector and Trigger Simulation

Once the final state (colorless) particles have been generated, their propagation through the detector, i.e. their interaction with the detector material, is simulated using the GEANT3 (Geometry and Tracking) package [144]. This package allows to describe any experimental setup by a structure of geometrical volumes. Each volume is assigned a medium, defined by several parameters characterizing how particles interact in it. The code allows to describe the transport of the particles through the various regions of the experimental setup, taking into account the volume geometrical boundaries and the physical effects inherent to the nature of the transported particles, and characterizing their interactions with matter and the magnetic field. Finally, it records the trajectories of the particles and the response of the sensitive detectors. Originally designed for high-energy physics experiments, it has today found many applications outside this domain. The packages makes heavy use of Monte Carlo methods to randomly generate the particle paths inside the detector volume.

Charge deposition models are used to describe the response of the tracking detectors. The interaction with the silicon detector is simulated using a simple geometrical model, based on the ionizing particle path length and on a Landau distribution. The drift of particles in the COT (section 4.3.1) is modeled by a GARFIELD [145] simulation tuned to COT data.

As modeling the interaction of every particle (and the secondary particles it produces) is computationally intensive, the simulation of the calorimeter response con-

sists in a parameterization performed using the GFLASH [146] package interfaced with GEANT3[2], which is tuned to test beam data. The outcome is a rapid and (quite) accurate response of the individual calorimeter towers to the energy deposited by the incoming particles.

As already mentioned, the reconstruction algorithm is applied on both the real and simulated data. The detector simulation takes into account the state of the detector, and the events are reconstructed using different run configurations, which include the state of all sub-detectors (and their calibrations). At the same time, the luminosity conditions are updated in order to model the low, medium, and high instantaneous luminosity regimes provided by the Tevatron during Run II. The main goal is to model the number of primary interaction vertices accurately.

The simulation of the trigger efficiency includes, for the first time in this signature, a complete multivariate parameterization, described in great detail in chapter 10.

7.5 Summary

In this chapter, we presented the event simulation and discussed the steps required for a proper modeling of proton-antiproton collisions. In the next chapter, we describe implementation of these steps to describe the processes contributing the $\not\!\!E_T$+b-jets dataset.

[2]However, recent studies of detector effects, mainly aimed at deriving adequate parameterizations of detector non-linearities, use the more recent GEANT4 [147]

8. THE $\not\!\!E_T$+JETS DATASET

8.1 The $\not\!\!E_T$+jets Signature

The $\not\!\!E_T$ plus jets signature was originally used in searches for supersymmetry (SUSY) or dark matter, where large missing energy is expected. More recently, it was also used in searches for the standard model Higgs boson [114, 148]. Indeed, the $\not\!\!E_T$ plus two jets is one of the most favorable signatures for searches for the standard model Higgs boson with mass below 135 GeV, since the Higgs boson then decays mostly to two b-quarks.

In the analysis presented here, we look at events with large missing transverse energy ($\not\!\!E_T$), two or three high-p_T jets, and with one (or both) of the two leading jet(s) is (are) identified as originating from a b-quark. These are detected in the calorimeters as sprays of particles called jets. Jets are identified as a group of electromagnetic and hadronic calorimeter clusters within a cone of radius $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \leq 0.4$ [107]. Since we are looking at events with misidentified electrons and muons, or hadronically decaying taus, we use a trigger relying on the presence of two calorimetric jets and large missing transverse energy.

This signature is mimicked by several standard model processes, which we shall review in detail later on. Due to its huge cross-section, the biggest background is by far the QCD production of multiple jets, referred to as QCD multijet production. In those events, the $\not\!\!E_T$ is not due to undetected particles but rather to mis-measurements and resolution effects. Then follow the other standard model processes involving the vector bosons or the top quark: the top pair production, the production of W or Z bosons together with heavy flavor jets (i.e. originating form b- or c-quarks), and the production of two vector bosons (referred to as diboson production).

8.2 The $\not\!\!E_T$+jets Trigger Path

For the purpose of collecting events with large $\not\!\!E_T$, it is natural to chose a dedicated trigger. Three $\not\!\!E_T$-based triggers are implemented at CDF:

1. MET_BJET, requiring $\not\!\!E_T > 20$ GeV and two displaced tracks;
2. MET45, requiring $\not\!\!E_T > 45$ GeV;
3. MET+JETS, requiring $\not\!\!E_T > 35$ GeV and two clusters with $E_T > 10$ GeV.

We discuss the details of these trigger paths in depth in chapter 10.

8.3 Quality Criteria for $\not\!\!E_T$+jet Analyses

All $\not\!\!E_T$+jets analyses rely on the presence of $\not\!\!E_T$ in the event. This $\not\!\!E_T$ can originate from neutrinos or other particles not interacting with the detector (SUSY particles, WIMPS, etc.) but also from instrumental (or detector) effects. These were studied at CDF by the $\not\!\!E_T$ working group, which issued a set of recommended selections to use in all $\not\!\!E_T$-based analyses: the $\not\!\!E_T$ *clean-up cuts*. These cuts aim at removing the instrumental sources of $\not\!\!E_T$, including beam halo muons. They are summarized below; a detailed description is presented in ref. [149].

1. Pass 1 cuts
   - At least one central jet ($|\text{detector } \eta| < 0.9$, cone size 0.4) and $E_T > 10$ GeV;
   - At least one COT track with $p_T > 0.5$ GeV and one axial super-layer with six or more hits on it;
   - An Event Electromagnetic Fraction (EEMF) > 0.1, where

$$\text{EEMF} = \frac{\sum_{j=1}^{N_{\text{jets}}} E_T^j \cdot \text{EMF}_j}{\sum_{j=1}^{N_{\text{jets}}} E_T^j}$$

     and EMF_j is the fraction of the jet energy deposited in the electromagnetic calorimeter (only jets with $E_T > 10$ GeV are considered).

2. Pass 2 cuts

   - At least one good primary vertex in the event;
   - An Event Charge Fraction (ECHF) > 0.1, where

$$\mathrm{CHF} = \frac{\sum_{j=1}^{N_{\mathrm{jets}}} \mathrm{CHF}_j}{N_{\mathrm{jets}}}$$

     and CHF_j is the jet charge fraction, defined as the ratio between the sum of the p_T of the tracks matched to the jet and the jet E_T.

3. Additional cuts (pass 3)

   - Event not in the *chimney*, a geometrical region of the calorimeter where jets are mis-measured. This region, $\phi \in [60°, 100°]$ and $\eta \in [0.5, 1.0]$, hosts cryogenic and instrumental connections to the inner detector. We exclude any event with a jet with $\not\!\!E_T > 10$ GeV falling into this region, since the latter is almost certainly mis-measured;
   - The primary vertex must fall within $|z| < 60$ cm of the nominal interaction point at the detector center.

The events failing any of the above selection criteria are discarded and are not analyzed.

8.4 The Event Reconstruction

The events not discarded by the selections described above are fully reconstructed using the complete offline CDF software, as described in chapter 5.

8.5 The Description of Processes Yielding $\not\!\!E_T$+jets

Before looking at the data, one has to know what to expect. For this purpose, we use either Monte Carlo simulations or data-driven methods to predict the distribu-

Table 8.1: Monte Carlo simulation path used to model the electroweak processes.

Process	Event Generator	Hadronization (PS)
VH	PYTHIA	PYTHIA
Single Top	POWHEG[1]	
Diboson	PYTHIA	
Top Pair	PYTHIA	
W/Z+jets	ALPGEN	

[1] We cross-check the simulation with MADEVENT.

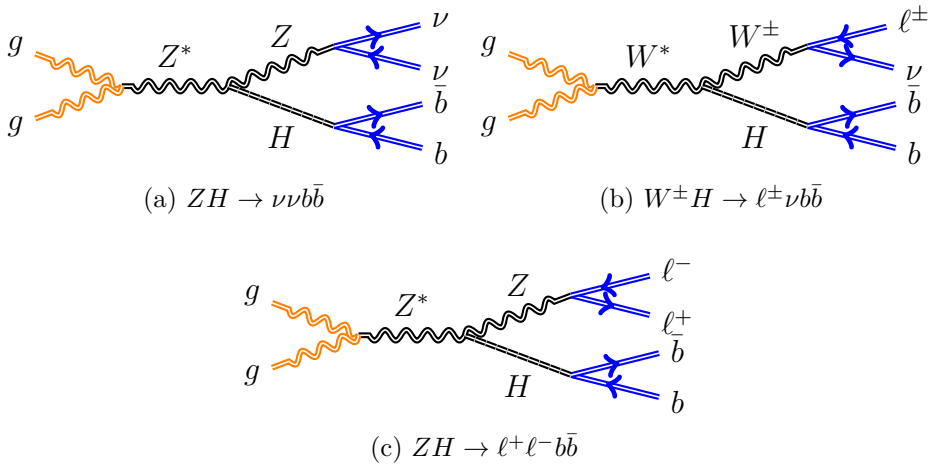
tions of various kinematic properties, ensuring though that we do not introduce any excessive bias[1], especially when using the latter methods[2].

The events analyzed in this dissertation are characterized based on the specificities of associated Higgs production (VH, $V = W, Z$) decaying to $\not\!\!E_T$ + b-jets, i.e.: (a) a large intrinsic missing transverse energy (> 30 GeV); (b) a relatively low jet multiplicity (2 or 3 jets), and (c) the lack of (detected) isolated leptons. There are numerous standard model (SM) processes that can produce this signature (in order of increasing cross-section at $\sqrt{s} = 1.96$ TeV): VH, electroweak single top, WZ/ZZ, top pair ($t\bar{t}$), WW, Z+jets, and W+jets production. Additionally, multi-jet QCD production (MJ) with mis-measured jet energies produces this signature, and is by far the most abundant at a hadron collider. In this chapter, we review these, along with the tools used to estimate their properties.

8.6 The Simulation of Electroweak Processes

As already mentioned, the $\not\!\!E_T$+b-jets signature is characterized by events with high $\not\!\!E_T$ and two or three high-p_T jets. We denote as electroweak processes events that involve the decay of a carrier of the electroweak interaction (W or Z boson). We model these processes exclusively using several MC event generators, interfaced to PYTHIA for the parton showering (PS). We briefly describe these simulated samples

[1]Since any model choice is driven by assumptions, it is not possible to be totally unbiased.
[2]Analysts must be careful when deriving data-driven models from regions in which the quantities under investigation behave differently than in the region where the model is to be applied.



(a) $ZH \to \nu\nu b\bar{b}$ (b) $W^{\pm}H \to \ell^{\pm}\nu b\bar{b}$

(c) $ZH \to \ell^{+}\ell^{-}b\bar{b}$

Figure 8.1.: Leading order Feynman diagrams for VH associated production. A similar set of diagrams is obtain from a $q\bar{q}$ initial state instead of a gg one. The diagrams are ordered by their respective contribution to the $\not\!\!E_T$+b-jets sample: (a) $ZH \to \nu\nu b\bar{b}$; (b) $W^{\pm}H \to \ell^{\pm}\nu b\bar{b}$, and (c) $ZH \to \ell^{+}\ell^{-}b\bar{b}$.

below. Table 8.1 summarizes the event generators used throughout this dissertation, unless mentioned otherwise.

8.6.1 SM Higgs

In the main result presented in this dissertation, the signal processes correspond to the associated production of a vector boson V (W or Z) in association with a Higgs boson H, where the weak force carriers decay to undetectable or undetected particles. Figure 8.1 shows three leading order Feynman diagrams for this production mode. This signature is sensitive to the following processes: $ZH \to \nu\nu b\bar{b}$, $WH \to \ell\nu b\bar{b}$, and, to a much lesser extent, $ZH \to \ell\ell b\bar{b}b$ when the leptons are undetected.

Since the mass of the Higgs boson, m_H, is not predicted in the SM, we produce samples for several assumed masses, ranging from 100 to 150 GeV/c^2 in 5 GeV/c^2 steps. This range has been chosen based on the fact that the branching $H \to b\bar{b}$ is dominant only up to $m_H = 135$ GeV/c^2 and falls rapidly thereafter, but also on indirect and direct constraints, respectively from electroweak precision measurement

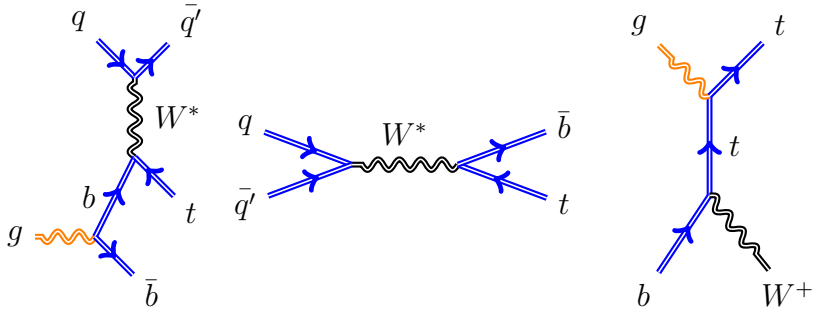
Table 8.2: The production cross sections and decay branching fractions for the SM Higgs boson assumed in this dissertation. The cross-sections for alternative Higgs-production mechanisms are given for comparison. Source: [55], and references therein

m_H (GeV/c^2)	$\sigma_{gg\to H}$ (fb)	σ_{WH} (fb)	σ_{ZH} (fb)	σ_{VBF} (fb)	$\sigma_{t\bar{t}H}$ (fb)	$B(H\to b\bar{b})$ (%)
100	1821.8	291.90	169.8	97.2	8.000	79.1
105	1584.7	248.40	145.9	89.7	7.062	77.3
110	1385.0	212.00	125.7	82.7	6.233	74.5
115	1215.9	174.50	103.9	76.4	5.502	70.5
120	1072.3	150.10	90.2	70.7	4.857	64.9
125	949.3	129.50	78.5	65.3	4.279	57.8
130	842.9	112.00	68.5	60.4	3.769	49.4
135	750.8	97.20	60.0	55.9	3.320	40.4
140	670.6	84.60	52.7	51.8	2.925	31.4
145	600.6	73.70	46.3	48.1	2.593	23.1
150	539.1	64.40	40.8	44.6	2.298	15.7

and past experimental searches [53]. The samples were generated using PYTHIA which treats the Higgs boson as a resonance [111]. The production cross-section and decay branching fraction depend on the assumed m_H, and include the latest higher order QCD and electroweak corrections [150]. The values assumed in this dissertation are shown in table 8.2.

8.6.2 Single Top

The electroweak production of a single top quark per event is the last standard model process to be observed [31, 32], and is a key milestone towards the Higgs. In single top events, a top quark is produced via the electroweak interaction (W^*tb vertex); there are three production modes: the t-channel, the s-channel and the tW associated production (figure 8.2). In t-channel production, a virtual space-like W boson strikes a b quark inside the proton or antiproton. In s-channel production, or W^* production, a time-like W boson is produced by the quark-antiquark annihilation



(a) t-channel (b) s-channel (c) tW-channel

Figure 8.2.: Feynman diagrams associated with (a) the t-channel, (b) the s-channel; and (c) the tW associated single top production modes.

process. Finally, the tW production, in which the top quark is produced in association with a real W boson, is also possible, but at a very low rate at the Tevatron [3]

Electroweak single top production was first observed at the Tevatron in 2009 [31, 32], 15 years after the observation of the top quark (via pair production) [28,29]. This process is not only important as a background to the SM Higgs (with a cross-section about 100 times higher), but is also interesting in its own right. Indeed, the single top cross-section is directly proportional to the square of the V_{tb} element of the CKM matrix (section 3.5.2). A precise determination of the former allows to constrain the latter (a loosely constrained parameter of the SM), without any assumption of unitarity (only the SM cross-section ratio between the different production mechanisms is assumed); this constraint on the CKM matrix is complementary to those from, e.g., B physics. Additionally, and because the top quark decays before hadronizing, the single top production allow studying the polarization of the top quark [151,152] and its decay width [153]. Finally, it is also sensitive to several kinds of new physics [153,154].

For single top, we use the NLL t-channel calculation of Kidonakis [155], which has been updated using the MSTW2008NNLO PDF set [45]. For the s-channel process we use [156], again based on the MSTW2008NNLO PDF set. Both of the cross section

[3]This process is however significant at the LHC.

values below are the sum of the single t and single $\bar{t}$ cross sections, and both assume $m_t = 173 \pm 1.2$ GeV.

$$\sigma_{t-\text{chan}} = 2.10 \pm 0.027(\text{scale}) \pm 0.18(\text{PDF}) \pm 0.045(\text{mass})\text{pb}. \quad (8.1)$$

$$\sigma_{s-\text{chan}} = 1.05 \pm 0.01(\text{scale}) \pm 0.06\ (\text{PDF}) \pm 0.03\ (\text{mass})\ \text{pb}. \quad (8.2)$$

Other calculations of $\sigma_{\text{s+t}}$ are similar for our purposes [157]. The cross-section for associated tW production, $bg \to tW$ (figure 8.2c), is $\sigma_{bg\to tW} \sim 0.22 \pm 0.08$ [158] (based on the CTEQ6.6M parameterization), which is of the order of the uncertainty on the t-channel cross-section, and therefore is small enough to be neglected for our purpose.

Both of the published Run I CDF single top searches used the PYTHIA Monte Carlo generator. It turned out to be inadequate because the top quarks were un-polarized, and the t-channel wasn't modeled well (too soft and too far forward distributed second b-quarks). For CDF Run II, we used MADEVENT to generate the default single-top samples. Compared to PYTHIA, MADEVENT offers two main advantages: it fully includes spin effects and it allows to use a matching procedure to derive a correct t-channel sample. These samples were used for the observation [31, 32].

Since the observation, CDF has put focus on a more precise direct measurement of V_{tb}. Using MADEVENT requires to re-do a heavy matching procedure (next section), which, additionally, was questioned by theorists [159, 160]. For these reasons, it was decided to look for an alternative, which turned out to be POWHEG.

The MADEVENT Samples

The CDF single top cross-section measurement in the $\not\!\!E_T$+b-jets signature[4] [161] is performed using events produced using the MADEVENT [134] matrix element generator, interfaced to the CTEQ5L [41] parameterization of the parton distribution functions (PDF). PYTHIA [111] is used for the parton showering and hadronization.

[4]This was the first analysis the author made, improving upon the techniques developed for searching for the SM Higgs in the $\not\!\!E_T$+b-jets signature [114].

Until 2011, this was the official sample used by CDF, as it allows the description of the polarization of the outgoing top quark. This analysis uses another approach, namely POWHEG [139] samples (cf. next section) to model single top production. The MADEVENT samples are used to cross-check that simulation.

The leading-order (LO) process for top quark production is a $2 \to 2$ process with a b-quark in initial state: $b+u \to d+t$ or $b+\bar{d} \to \bar{u}+t$. The antitop quark production implies the conjugate processes.

As several authors have pointed out [159, 160], the distribution of observable jets is not adequately represented by the LO contribution to the t-channel production of single top and it is better predicted by next-to-leading-order (NLO) calculations. In the latter, the b-quark stems from a gluon splitting into a $b\bar{b}$ pair. The $\bar{b}$ quark required by flavor conservation of the strong interaction is created by LO parton shower programs through backward evolution following the DGLAP scheme [162–164]. The high-p_T tail of the transverse momentum distribution of the $\bar{b}$ quark is not adequately modeled but the the low-p_T is. The mismodeling is estimated by comparing with a NLO calculation [160].

The modeling of the t-channel single top quark process can be improved by producing simulated events with a matrix element generator, followed by the simulation of the production of observable particles by parton shower Monte Carlo. For this, two samples are used: one for the leading $2 \to 2$ process; and one for the $2 \to 3$ process with a gluon in the initial state $g + q \to q + t + \bar{b}$. In the latter process, the $\bar{b}$ quark is directly produced in the hard scattering described by the matrix element. It also describes the most important NLO high-p_T tail of the $\bar{b}$ quark p_T distribution.

The construction of a Monte Carlo sample following the NLO predictions is done by matching the $2 \to 2$ and $2 \to 3$ processes. Additional details on the matching procedure are provided in ref. [165]. This technique provides reasonable agreement with NLO predictions.

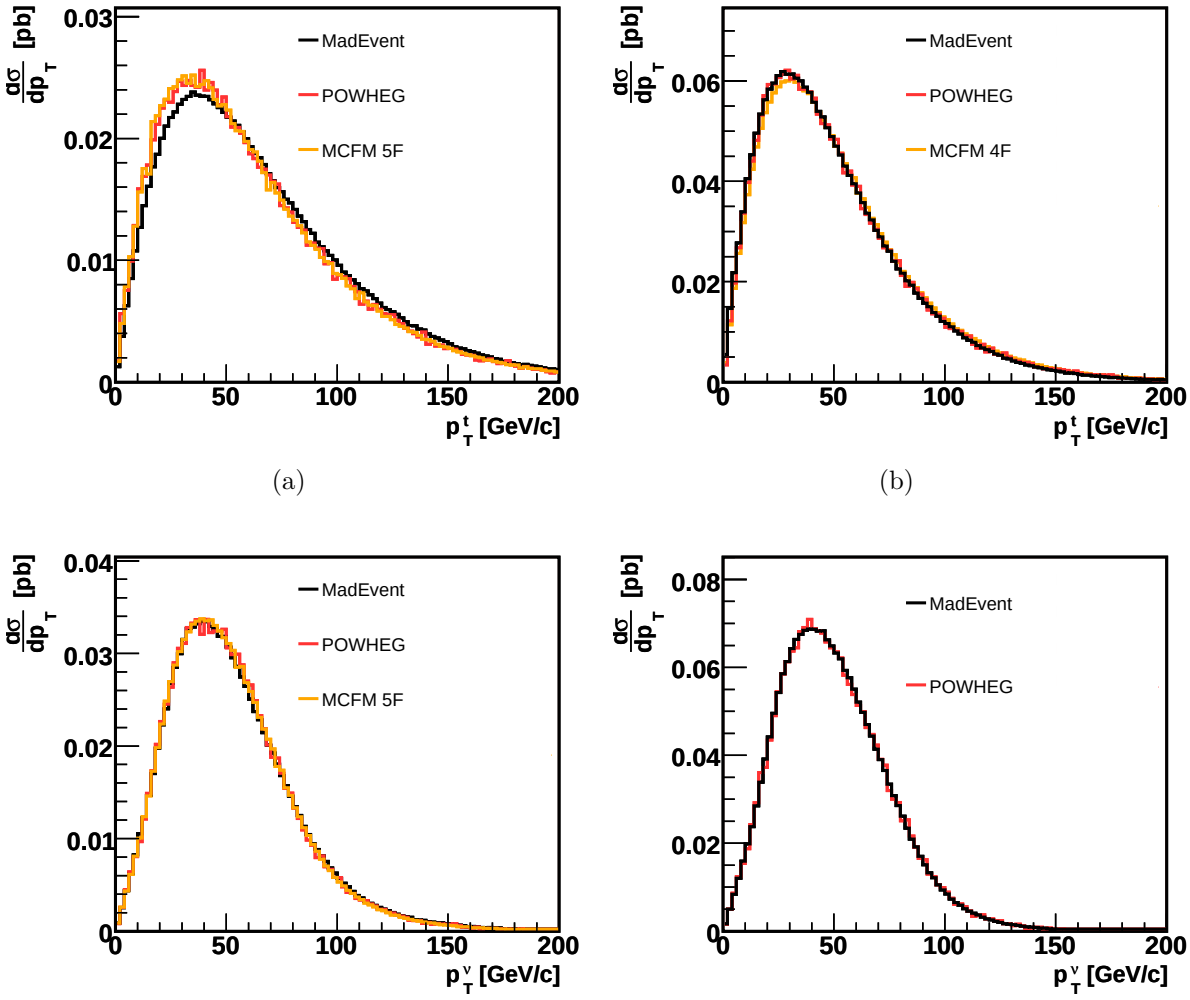


Figure 8.3.: Differential single top production cross-section for the s- (a,c) and t-channel (b,d) production modes as a function of the transverse momentum of the top quark (a,b) or the neutrino from the W decay (c,d). The black, red and yellow lines are obtained using MadEvent, Powheg and MCFM respectively. Source: [166]

POWHEG Samples

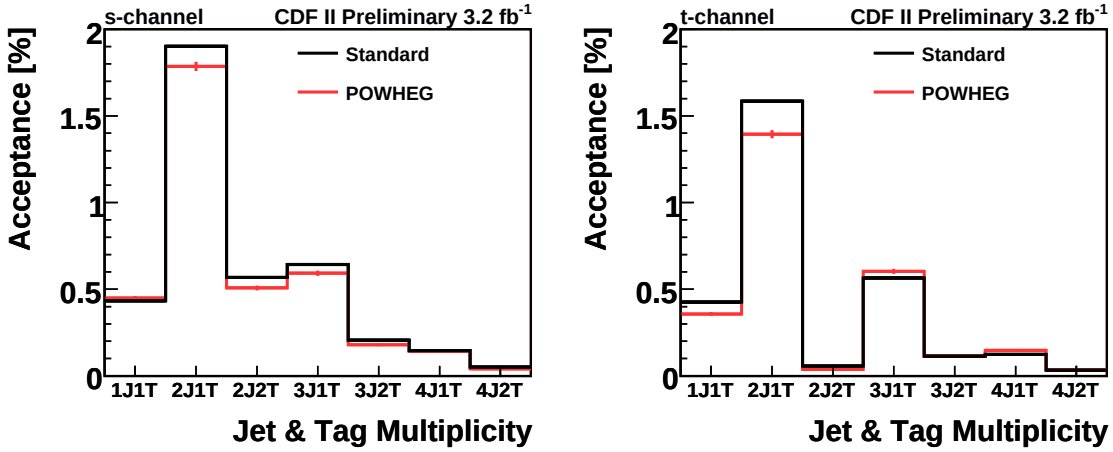
To address the complexity associated with the use of the MADEVENT samples, we use the POWHEG [139] method to simulate single top production at NLO. Figure 8.3 shows the differential single top production cross-section for the s- and t-channel production modes as a function of the transverse momentum of the top quark and the neutrino from the W decay using the event selection from the lepton+jets single top analysis [167]. We compare the predictions between MADEVENT, POWHEG and MCFM [168]; the latter two have excellent agreement, as is the case for additional checks [166]. The small discrepancies with respect to MADEVENT are due to the matching procedure. Figure 8.4 shows the acceptance to s- and t-channel single top production in the lepton+jets selection. We observe that POWHEG populates the higher jet multiplicities more in the t-channel. This is because POWHEG includes more radiation compared to LO MADEVENT through additional NLO diagrams.

These checks confirm that the POWHEG samples are in good agreement with MADEVENT, but provide a better simulation by including NLO effects. In the $\not\!\!E_T$+b-jets signature, we compared MADEVENT to POWHEGwith the same conclusion.

8.6.3 Top Pair

Top pair production (cf. figure 8.5) yields a measurable contribution to the background in the signal region. Due to its large mass and the semi-leptonic decay of the top, these events are energetic, bear large $\not\!\!E_T$ and high jet multiplicity. The $t\bar{t}$ events were generated using the PYTHIA generator assuming a top quark mass of $172.5\,\mathrm{GeV/c^2}$, and normalized to

$$\sigma_{t\bar{t}} = 7.04^{+0.24}_{-0.36}\ (\text{scale}) \pm 0.14\ (\text{PDF}) \pm 0.30\ (\text{mass}) \tag{8.3}$$



(a) (b)

Figure 8.4.: Acceptance to single top for the *s*- (a) and *t*-channel (b) for various *b*-tag and jet multiplicities. The black (red) line corresponds to POWHEG(MADEVENT). The event selection corresponds to that of the lepton+jets single top analysis [167]. Source: [166]

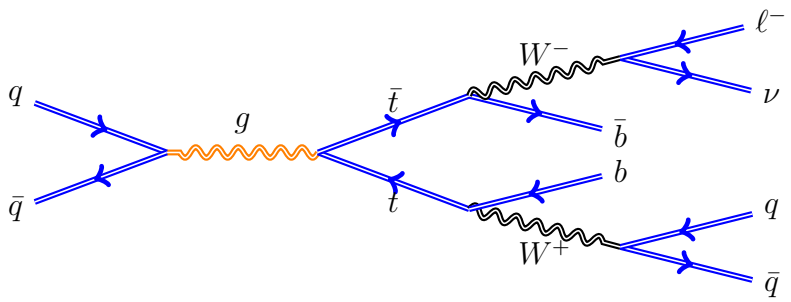


Figure 8.5.: Leading order Feynman diagram for top pair production. The pictured mode produces about 85% of the Tevatron top pairs; the similar $gg \to t\bar{t}$ graph describing the production of the remainder.

following the calculation of Moch and Uwer [169] using the MSTW2008NNLO PDF set [45] and assuming a top quark mass $m_t = 173.0 \pm 1.2$ GeV/c^2 [170]. Other calculations of $\sigma_{t\bar{t}}$ are provide similar results [171, 172].

8.6.4 Dibosons: $WW/WZ/ZZ$

The diboson simulations (cf. figure 8.6), also obtained using PYTHIA, have all-inclusive boson decays. MCFM [143] is used to compute the NLO cross sections for

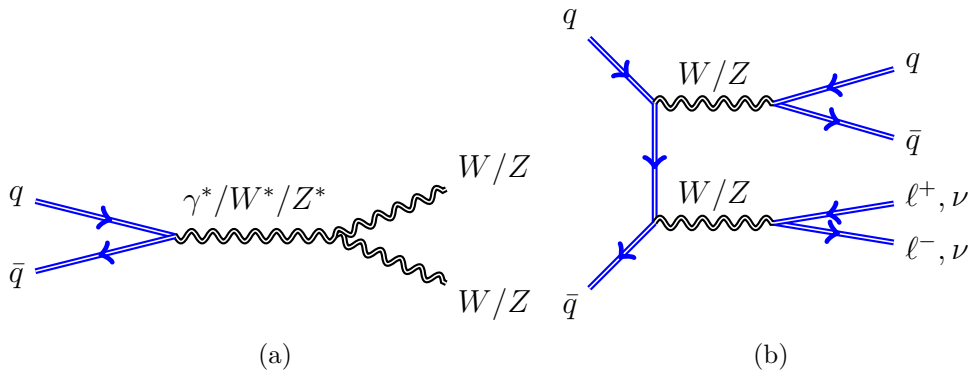


Figure 8.6.: Feynman diagrams for diboson production (WW, WZ, and ZZ).

WW, WZ, and ZZ production. Using a scale choice $\mu_0 = M_V^2 + p_T^2(V)$ and the MSTW2008 PDF set [45], the cross section for inclusive W^+W^- production is

$$\sigma_{W^+W^-} = 11.34^{+0.56}_{-0.49}\ (\text{scale})\ ^{+0.35}_{-0.28}\ (\text{PDF})\ \text{pb} \tag{8.4}$$

and the cross section for inclusive $W^{\pm}Z$ production is

$$\sigma_{W^{\pm}Z} = 3.22^{+0.20}_{-0.17}\ (\text{scale})\ ^{+0.11}_{-0.08}\ (\text{PDF})\ \text{pb} \tag{8.5}$$

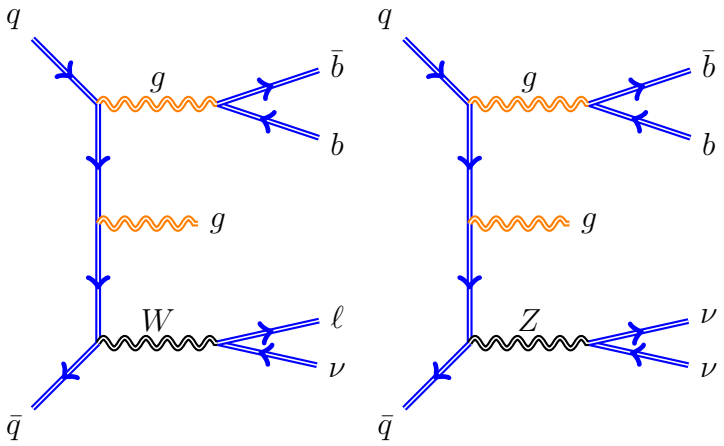
For the Z, leptonic decays are used in the definition, with both γ and Z exchange. The cross section quoted above involves the requirement $75 \leq m_{\ell^+\ell^-} \leq 105$ GeV for the leptons from the neutral current exchange. The same dilepton invariant mass requirement is applied to both sets of leptons in determining the ZZ cross section which is

$$\sigma_{ZZ} = 1.20^{+0.05}_{-0.04}\ (\text{scale})\ ^{+0.04}_{-0.03}\ (\text{PDF})\ \text{pb}. \tag{8.6}$$

For the diboson cross section calculations, $|\eta_\ell| < 5$ for all calculations. Loosening this requirement to include all leptons leads to $\sim$+0.4% change in the predictions. Lowering the factorization and renormalization scales by a factor of two increases the cross section, and raising the scales by a factor of two decreases the cross section. The PDF uncertainty has the same fractional impact on the predicted cross section independent of the scale choice. All PDF uncertainties are computed as the quadrature sum of the twenty 68% C.L. eigenvectors provided with MSTW2008 (MSTW2008NLO68CLs).

Some considerations regarding the normalization of ZZ

We calculate the ZZ cross-section using MCFM and use it to normalize our PYTHIA sample. There are some subtleties in doing this correctly, because of the Z/γ^* interference. The couplings of γ^* to fermions are very different than those of the Z, which means that, when γ^* is relevant, the branching fraction of the process is affected. According to the PYTHIA manual:



(a) $W + b\bar{b} + jets$ (b) $Z + b\bar{b} + jets$

Figure 8.7.: Feynman diagrams for the production of (a) a W or (b) Z boson in association with jets.

> Wherever Z_0 is written, it is understood that γ^* and γ^*/Z_0 interference should be included as well (with possibilities to switch off either, if so desired). In practice, this means that fermion pairs produced from γ^*/Z_0 decay will have invariant masses as small as the program cutoff, and not regulated by the large Z mass [111].

Pythia has a complete leading order model for all the four-fermion final states, while MCFM simulates at next to leading order only the final states that are likely to be measured ($\ell\ell\ell\ell$, $\ell\ell\nu\nu$, $\ell\ell b\bar{b}$, $\nu\nu b\bar{b}$), so our Pythia sample contains final states for which MCFM does not even calculate a cross-section. For this reason, the cross-section must be corrected by a factor which is approximately 3, and was derived from comparing the yields inside and outside of the $75 \leq m_{\ell^+\ell^-} \leq 105$ GeV resonance.

8.6.5 W/Z+jets

We use Alpgen + Pythia to model the production of a W or Z boson in association with jets (figure 8.7) with the corresponding parton shower. The cross-sections for the many samples used are summarized in table 8.3. This model provides

Table 8.3: The production cross sections for the production of a W or Z boson in association with jets assumed in this dissertation.

Process	ALPGEN cross-section [pb][1]				
	0p	1p	2p	3p	≥4p
$W \to \ell\nu$	1800.00	225.00	35.30	5.59	1.03
$W \to \ell\nu + b\bar{b}$	2.98	0.89	———	0.29	———
$W \to \ell\nu + c\bar{c}$	5.00	1.79	———	0.63	———
$W \to \ell\nu + c$	17.10	3.39	0.51	—— 0.08 ——	
$Z \to ee(\mu\mu)$	158.00	21.60	3.47	0.55	0.10
$Z \to ee(\mu\mu) + b\bar{b}$	0.51	0.13	———	0.04	———
$Z \to ee(\mu\mu) + c\bar{c}$	1.08	0.33	———	0.11	———
$Z \to \tau\tau$	158.00	21.60	———	4.14	———
$Z \to \tau\tau + b\bar{b}$	———	———	0.62	———	———
$Z \to \tau\tau + c\bar{c}$	———	———	1.28	———	———
$Z \to \nu\nu$	1150.13	155.34	25.06	3.95	0.71
$Z \to \nu\nu + b\bar{b}$	2.85	0.71	———	0.18	———
$Z \to \nu\nu + c\bar{c}$	7.94	2.43	———	0.78	———

[1] The cross-sections shown do not reflect the 1.3 ± 0.4 k-Factor derived from $\sigma(MCFM)/\sigma(Alpgen)$

statistics for a good description of the tails of the distributions[5]. The cross-sections from table 8.3 are scaled up by a k-Factor of 1.3 to account for the NLO effects. The ratio of light- to heavy-flavor events is assumed to be that provided by the ALPGEN cross-section[6]. We assign a 30% uncertainty on the W/Z+jets production cross-section. We now discuss the matching procedure.

Matching the samples for W/Z+jets production

When modeling the production of a real W or Z boson in association with quarks or gluons, several aspects have to be considered. For instance, quarks and gluons can

[5]Iterations of this analysis prior to this work used LO PYTHIA predictions, normalized as necessary. The move to ALPGEN is motivated by the fact that these all-inclusive samples lacked a proper description of the tails, especially as the integrated luminosity kept increasing; for the bulk of the distributions, the two categories of samples performed similarly.

[6]Several dedicated studies were performed, showing no significant departure from the ratio predicted by ALPGEN.

both hadronize into a jet, and, additionally, a gluon can split into a pair of heavy flavor quarks.

The samples are generated with a number of light flavor partons p in the matrix element. We use seven different sample categories for each lepton family. Three categories model the Z+jets contribution: $Z + Xp$, $Z + b\bar{b} + Xp$ and $Z + c\bar{c} + Xp$. Four categories model the W+jets contribution: $W + Xp$, $W + b\bar{b} + Xp$, $W + c\bar{c} + Xp$, $W + c + Xp$. Together, all these categories cover events with final state parton multiplicities from zero to four.

Because we generate each final state explicitly, we ought to remove some events to avoid double counting. For example, it is possible for an additional jet produced by PYTHIA in the parton shower to occupy the same phase space as an event from a higher parton multiplicity sample.

We perform our matching based on reconstructed quantities. We explicitly reject events where the heavy flavor quarks from the matrix element fall in the same jet (leaving the generation of such events to the parton shower). We also remove heavy flavor from the parton shower when only one of the two heavy quarks is inside the jet cone. In both cases, we use a cone size of 0.4. Additionally, we discard events from the $W/Z + Xp$ samples with a (massless) b/c quark from the matrix element.

In summary, the idea is to remove charm and bottom flavor from the light flavor samples and bottom flavor from the charm samples (or the second charm in the $W + c + Xp$ case).

8.6.6 Luminosity-driven corrections

After the simulated events are generated, the detector simulation code is used to model "real-world" conditions. This process depends strongly on the detector conditions (as implemented in the simulator). In particular, the vast majority of MC samples were generated with a low-luminosity profile. Additional samples with a medium and high luminosity profile were generated, but this effort was not pursued

to account for the high luminosity regimes of the last 2.5 years of data-taking. The consequence of this is an lack of high-activity events in our simulations. We attempt to correct for this effect by re-weighing our distributions according to the number of primary vertices in the event. We use events with taggable[7] jets from a region enriched in $Z \to \ell\ell$ events – two same-sign leptons with $75 < m_{\ell\ell} < 105$ GeV/c^2 – to correct for the luminosity profile. This sample was chosen because it is exempt of multi-jet background, for which we don't have a *pre-tagged* model (more information below). To avoid any unnecessary bias, we do not propagate the rate difference between the MC and the data when re-weighing. Moreover, we maintain the relative fraction of background processes given by the MC. In Figure 8.8, we show the effect of the re-weighing procedure.

8.7 QCD Multijet Production Modeling

The biggest background in the $\not\!\!E_T$+jets signature is by far multijet QCD production (MJ, figure 8.9). QCD jet production has a large cross-section ($\sim \mu$b), which is about 9 orders of magnitude greater than the signals typically looked for in this signature before requiring the presence of b-jets. Although these processes generally do not produce neutrinos, mis-measured jets do cause imbalance in the total transverse energy. Therefore, the QCD events can pass the basic selection cuts if one of the jets is mis-tagged as a b-jet. Furthermore, QCD b-quark pair production yields taggable jets and if one B meson originating from the b-quark undergoes a semi-leptonic decay, it could also produce $\not\!\!E_T$. The main sources in the $\not\!\!E_T$+b-jets final state are the QCD production of heavy-quark pairs ($b\bar{b}$ and $c\bar{c}$) and false tags, i.e. mis-tags (sections 6.1.2 and 6.2.2) from light-quark jets.

Because of the high production rate, a model of the QCD jet production with enough statistics is needed in order to describe the process adequately. Achieving this

[7]The adjective taggable refers to jets that have certain quality requirements – which depend on the b-tagging algorithm – to allow the jet to be b-tagged. By extension, a taggable event is an event with at least one taggable jet.

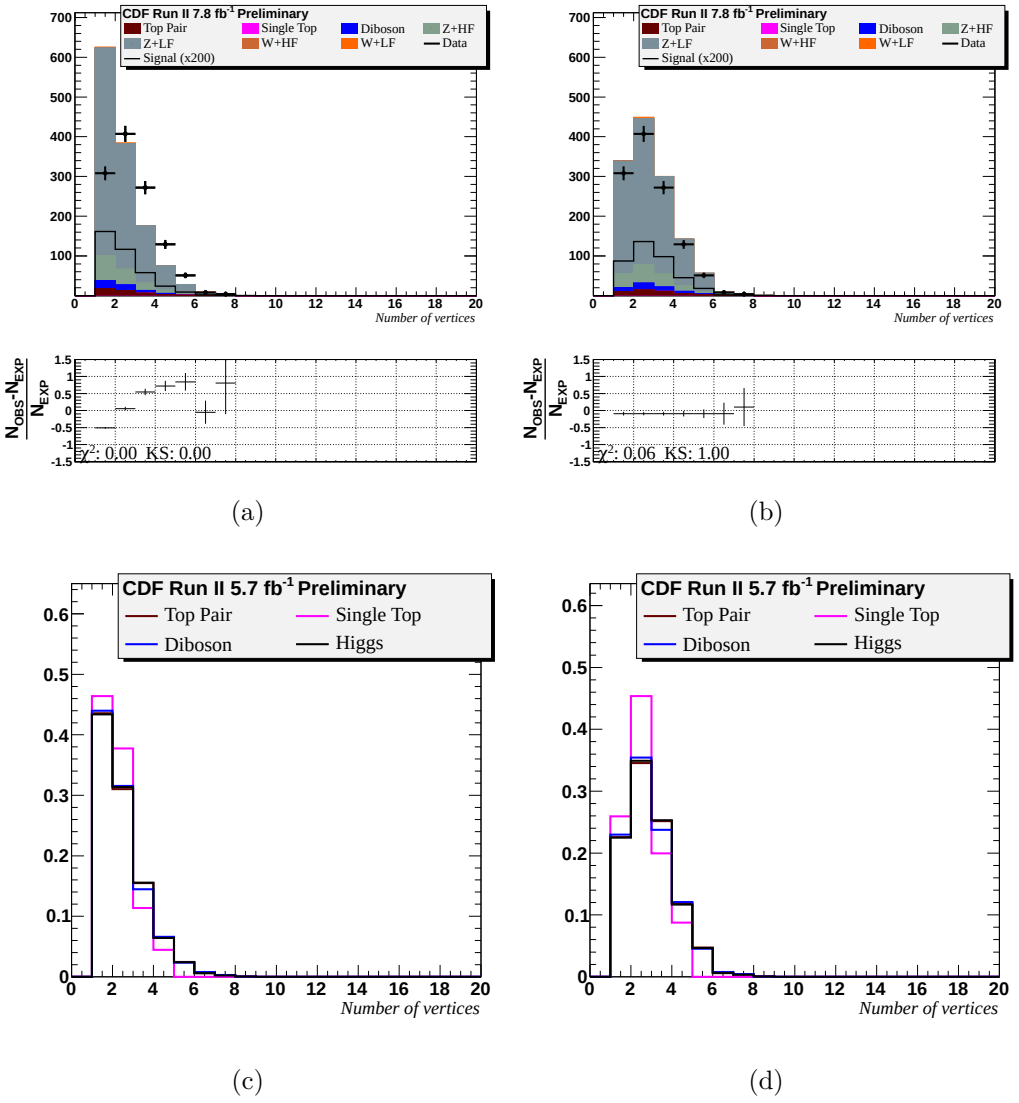
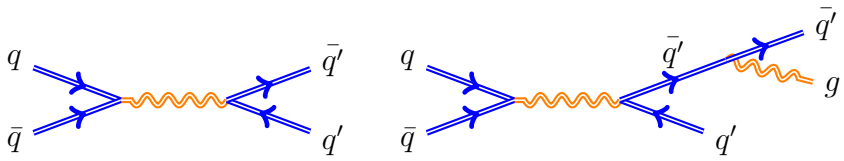


Figure 8.8.: Distribution of the number of vertices in a $Z \to \ell\ell$ enriched region before (a) and after (b) re-weighing. Due to limited statistics, all events with $N_{vtx} \geq 7$ are corrected by the same factor. (c) and (d) show the effect of this re-weighing on several Monte Carlo distributions. In (a) and (b), HF and LF refer to the *tree-level* processes in ALPGEN.



(a) (b)

Figure 8.9.: Feynman diagrams for (a) di-jet and (b) tri-jet QCD production.

with traditional Monte Carlo simulation is extremely computing intensive[8]. Moreover, the systematic uncertainties associated to the MC and detector simulation of QCD jet production are high. For these reasons, it is important to estimate the QCD background without using MC simulation, thus solely from data.

8.7.1 A Data-driven Model for QCD Multijet Production

Because of its complexity and its large cross-section, QCD multijet production cannot be easily modeled using MC events. The non-negligible high-order effects in QCD are very difficult to include in simulations, and yield to large systematic uncertainties. Additionally, the generation of large amounts of events is time consuming. For these reasons, we resort to data to estimate this background contribution.

To model the multi-jet QCD background (QCD MJ), we define a four-dimensional *Tag-Rate-Matrix* (TRM) to determine the probability for an event to be tagged. We use the following four variables: $\not\!\!p_T$ (transverse momentum flow imbalance), H_T (scalar sum of jet energies), $\sum {p_T}^{j_1}_{chgd}/{p_T}^{j_1}$, and $\sum {p_T}^{j_2}_{chgd}/{p_T}^{j_2}$ (ratio of SecVtx Pass1 tracks over jet energy). The matrix is split into bins according to table 8.4.

Table 8.4: Binning of the *Tag-Rate-Matrix* variables.

Variable	Units	Binning										
$\not\!\!p_T$	GeV/c	45	80	100	120	150	200	1000				
H_T	GeV	0	8	15	25	35	45	60	1000			
$\frac{\sum {p_T}^{j_i}_{chgd}}{{p_T}^{j_i}}$		0	0.025	0.05	0.1	0.15	0.2	0.3	0.4	0.6	0.8	10

We define one TRM for each of our tagging categories: 1S, SS or SJ. To model QCD MJ in the relaxed kinematic regime, the TRM has been derived from events with:

- 2 or 3 jets, each with $|\eta| < 2.0$ and at least one being central ($|\eta| < 0.9$);
- $35 < \not\!\!E_T < 70$ GeV, $E_T^{j_1} > 25$GeV, and $E_T^{j_2} > 20$ GeV;

[8]In addition, the vast majority of (simulated) MJ events must be rejected in order to be sensitive to processes with much smaller cross-sections.

- $\Delta R(\vec{j_1}, \vec{j_2}) > 0.8$, and $\Delta\phi(\not\!\!E_T, j_2) < 0.4$;

This region is dominated by QCD, with minimal ($< 1\%$) contamination by other processes. The main characteristic of the selected events is that they are mis-measured. Indeed, mis-measured di-jet events would in the vast majority have their $\not\!\!E_T$ aligned to the second (mis-measured) jet, as shown in figure 11.2. The region $\not\!\!E_T > 70$ GeV has lower statistics but is still QCD-dominated, and is therefore used to cross-check the QCD MJ modeling.

This *Tag-Rate-Matrix* model is improved with respect to [114,161] ; it is now *event*-based rather than *jet*-based. We found that this approach yields a better description of the kinematic quantities studied because there are no complex correlations between the jets to take into account[9]. This model provides a better modeling of the leading jet-E_T's even though these do not appear the matrix. The inability to perfectly model the transverse energy of the jets lies into the way we selected our QCD-rich training sample. Indeed, we have explicitly selected events with a mis-measured second jet for this purpose. Unfortunately, there is no better alternative to this choice. We have tried to use pre-selection[10] events with $\Delta\phi(\not\!\!E_T, \not\!\!p_T) > \pi/2$ without clear improvement in modeling, and with much less statistics[11].

8.7.2 The components of our data-driven model

The TRM model allows the prediction of the *probability* of either of the two leading jets[12], or both, to be in any of the tagging categories being used (cf. section 11.2.5).

[9]When deriving and using a jet-based TRM, the jets must be assumed to be independent. This assumption causes slight mis-modeling for events where two jets are b-tagged, due to the fact that the two jets in those events are correlated. These correlations are difficult to take into account.
[10]The pre-selection consists of events passing the selection cuts described in section 11.2. These remove events with mis-measured jets or with identified leptons. The region in which the limit on the Higgs production is set consists of pre-selection events passing an artificial neural network selection (cf. section 11.2.6).
[11]The choice of $\Delta\phi(\not\!\!E_T, \not\!\!p_T) > \pi/2$ is justified by the fact that $\Delta\phi(\not\!\!E_T, \not\!\!p_T)$ is symmetric around $\pi/2$ for QCD dijet production (cf. section 11.2.4).
[12]An attempt was made to include the third jet, but this didn't yield viable, because of the curse of dimensionality (i.e. the need for an extra dimension in the matrix, reducing the population in each bin) combined with the reduced statistics (QCD productions yields mainly di-jet events).

However, because this model is purely data-driven, it also accounts for tagged jets coming from other processes, among which W/Z+jets and $t\bar{t}$ production. To avoid double counting of events already accounted for by the relevant Monte Carlo models, we subtract the contribution of those processes, which we refer to as *non-QCD* backgrounds.

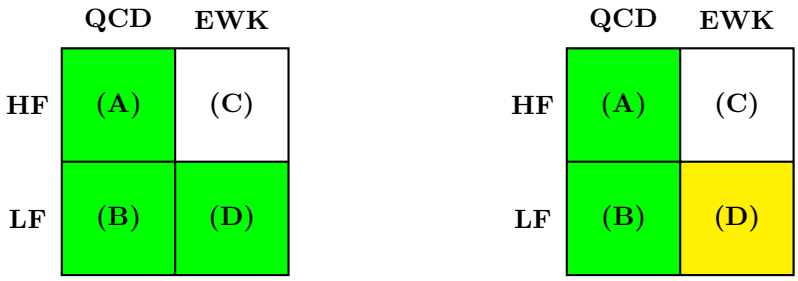


Figure 8.10.: Components of multijet sample: (A) QCD heavy flavor, (B) QCD light flavor, (C) electroweak heavy flavor, and (D) electroweak light flavor production. (a) Data-driven model for all components but heavy flavor electroweak; (b) Data-driven model for QCD production, with electroweak light flavor from mis-tag matrices (cf. sections 6.1.2 and 6.2.2).

In the first step of our data-driven method, we apply the TRM weights to each data event with taggable jets and obtain an initial model for mutlijet in each tagging category which includes the following contributions:

1. **QCD HF:** these are QCD events with jets from heavy flavor partons;
2. **QCD LF mis-tags:** these are QCD events with jets from light flavor partons that have been incorrectly tagged as coming from heavy flavor ones;
3. **Non-QCD HF:** these are events from known SM electroweak processes[13] with jets from heavy flavor partons;

[13]We consider the contributions from the following sources: $t\bar{t}$ W/Z+jets, diboson and single top.

4. **Non-QCD mis-tags:** these are events from electroweak processes with jets from light flavor partons that have been incorrectly tagged. This category is also referred to as EWK LF mis-tags.

We strive to model all QCD contributions from data. The non-QCD component is obtained from Monte Carlo simulations. We subtract the latter to avoid double counting, as described in the next section. It is worth noting that the QCD HF is dominant throughout our analysis regions but that the *non-QCD mis-tags* becomes more important as we approach the signal region and in our electroweak control regions.

8.7.3 Non-QCD background subtraction

We start by applying the TRM weight to each taggable event (in MC and in data). We then subtract all the weighted MC from the weighted data. The Higgs signal is not subtracted since it is an hypothetical component of the SM.

After this subtraction, we obtain a model of the QCD LF and HF components (A and B in figure 8.10b). We include EWK LF mis-tags (i.e. non-QCD mis-tags) applying the mis-tag matrices (cf. sections 6.1.2 and 6.2.2) to the EWK LF MC (D component in figure 8.10b). This prescription differs from [114, 161] in which only EWK HF MC was subtracted to yield a single model for QCD and EWK LF mis-tags (cf. figure 8.10a).

The new modeling technique separates two very different components: QCD multijet production and EWK LF mis-tags. This allows for a significant improvement of the background model, as discussed in section 8.7.6.

8.7.4 Jet E_T re-weighing

The data-driven model provides an excellent agreement with data for angular variables but suffers from issues in the case of to energy-related kinematic variables such as H_T, the scalar sum of the jet energies, and $M(j_1, j_2)$, the invariant mass

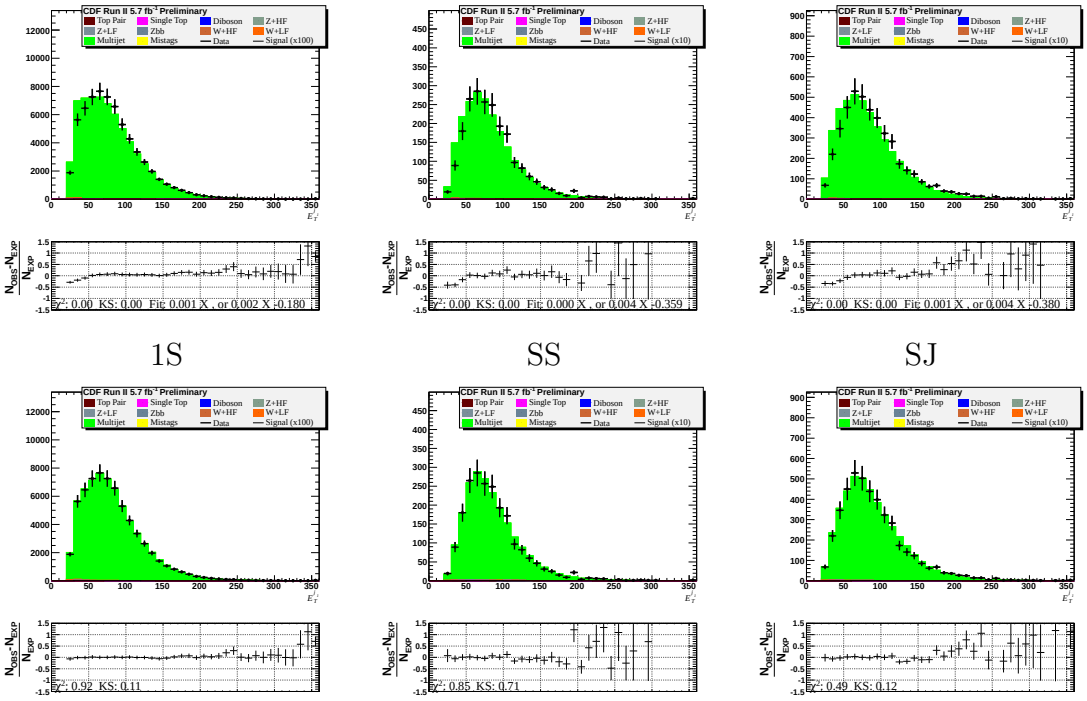
of the two leading jets. This is due to the inherent nature of the sample used to derive the TRM. Since we have no other alternative to derive a high statistics QCD sample, we must incorporate information from a sample in which the second jet is not mis-measured (i.e. aligned to the $\not\!\!E_T$).

For this purpose, we compare the shape of the two-dimensional distribution of ($E_T^{j_1}$, $E_T^{j_2}$) for the data-driven model to the same distributions for the data in a region where QCD is the dominant background (less than 1% contamination by other processes, but with much less statistics than the region in which the TRM was derived). This region is defined requiring $NN_{QCD} < 0.1$, where NN_{QCD} is the output of an artificial neural network trained to separate the QCD background from the Higgs signal (cf. section 11.2.6). Multijet events mostly have low NN_{QCD} output values, while the other SM processes have higher output values. We take the ratio in each bin (table 8.5) and re-weigh the *shape* of the QCD MJ distribution after the non-QCD background subtraction; this procedure does not affect the normalization of the QCD MJ (done after these weights are applied, and described below) nor any other background.

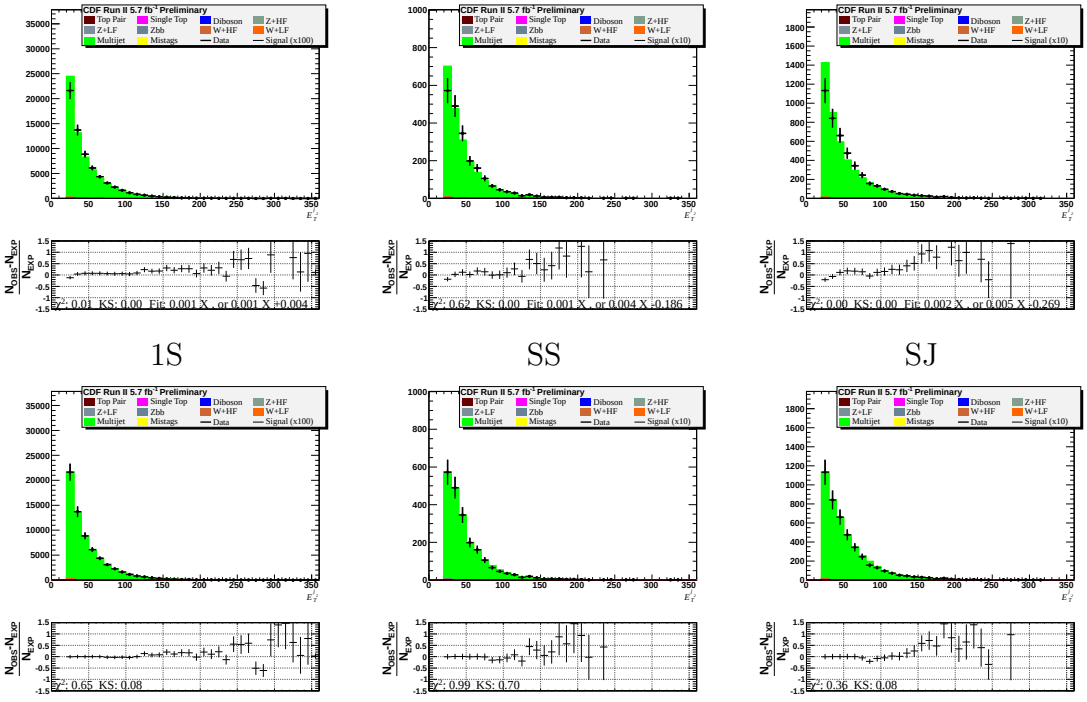
Table 8.5: Binning of the Jet E_T re-weighing matrix.

Variable	Units	Binning								
$E_T^{j_1}$	GeV	25	35	45	55	65	75	85	95	1000
$E_T^{j_2}$	GeV	20	30	40	50	60	70	1000		

We derive the weights in a region dominated by QCD, namely the lower end of an artificial neural network, NN_{QCD}, presented in section 11.2.6, trained to separate the QCD MJ background from the signal. Figure 8.11 shows the effect of this re-weighing in pre-selection events with $NN_{QCD} < 0.1$. Figure 8.12 shows the effect in the overall pre-selection on H_T and $M(j_1, j_2)$. The improvement is clearly visible, as the slope of the ratio plot ($N_{obs}/N_{exp} - 1$) is horizontal after re-weighing the events.

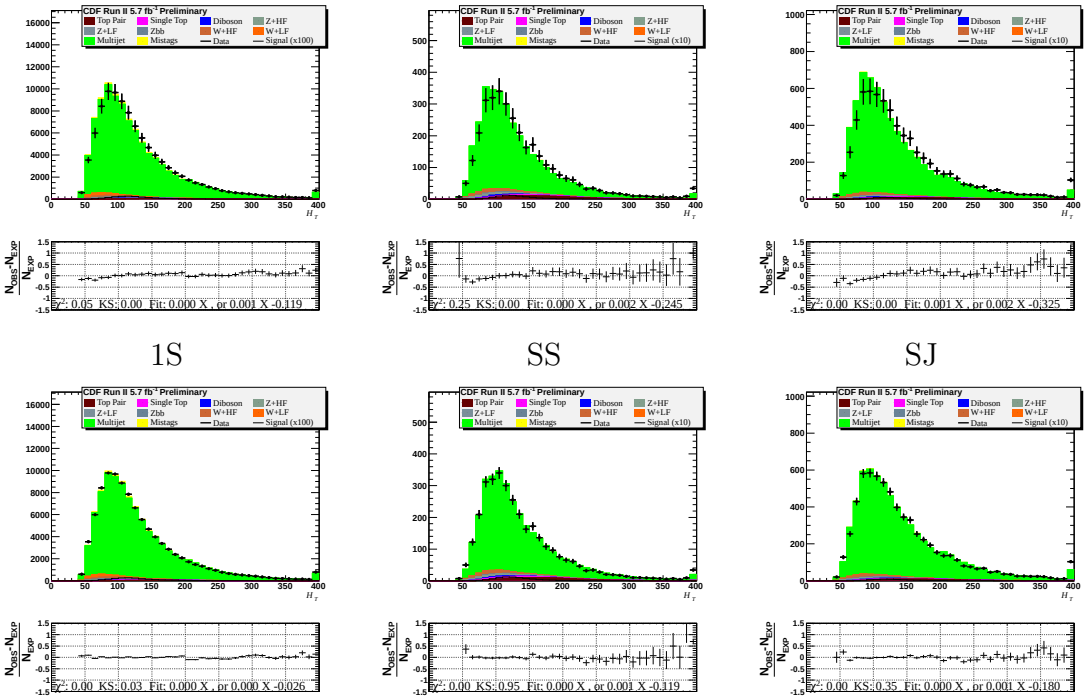


(a) Effect on E_T^{j1}: before (top) and after (bottom) re-weighing the events.

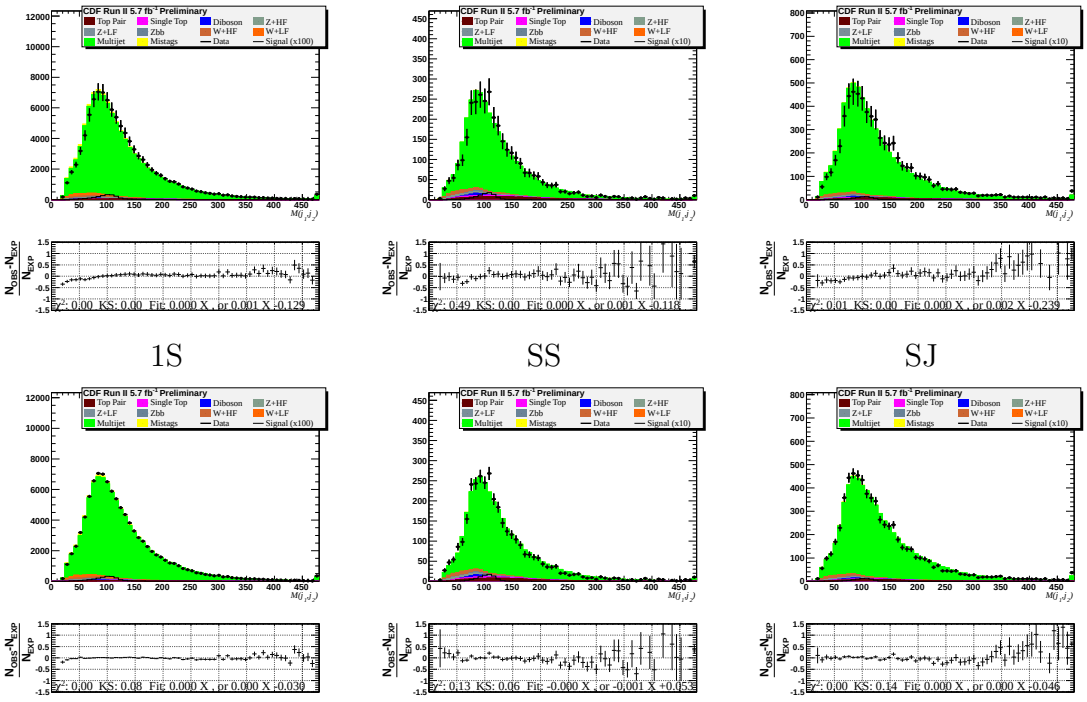


(b) Effect on E_T^{j2}: before (top) and after (bottom) re-weighing the events.

Figure 8.11.: Jet E_T-dependent re-weighing in the low NN_{QCD} region: effect on (a) E_T^{j1} and (b) E_T^{j2}. In these plots, HF and LF refer to the *tree-level* processes in ALPGEN; both can yield heavy flavor quarks through parton showering. Mistags refers to EWK LF mis-tags (cf. section 8.7.2).

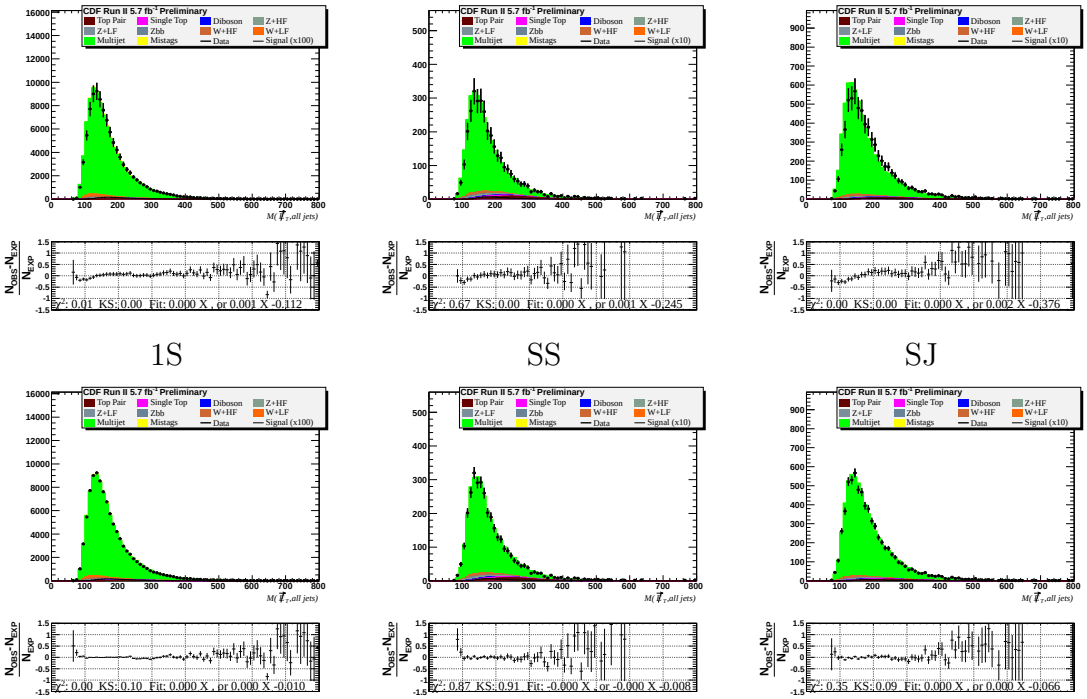


(a) Effect on H_T: before (top) and after (bottom) re-weighing the events.

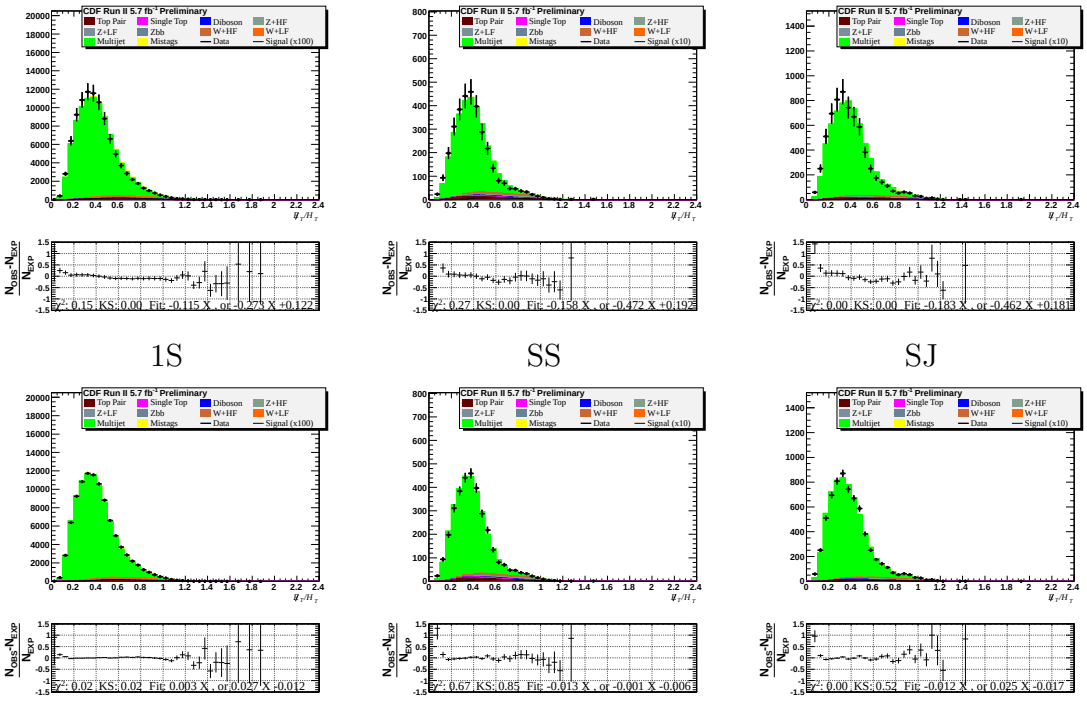


(b) Effect on $M(j_1, j_2)$: before (top) and after (bottom) re-weighing the events.

Figure 8.12.: Effect of the jet E_T-dependent re-weighing in the pre-selection on (b) H_T and (b) $M(j_1, j_2)$. In these plots, HF and LF refer to the *tree-level* processes in ALPGEN; both can yield heavy flavor quarks through parton showering. Mistags refers to EWK LF mis-tags (cf. section 8.7.2).



(a) Effect on $M(\not\!\!E_T, \text{jets})$: before (top) and after (bottom) re-weighing the events.



(b) Effect on $\not\!\!E_T$: before (top) and after (bottom) re-weighing the events.

Figure 8.13.: Effect of the jet E_T-dependent re-weighing in the pre-selection on (a) $M(\not\!\!E_T, \text{jets})$ and (b) $\not\!\!E_T/H_T$. In these plots, HF and LF refer to the *tree-level* processes in ALPGEN; both can yield heavy flavor quarks through parton showering. Mistags refers to EWK LF mis-tags (cf. section 8.7.2).

Table 8.6: QCD MJ normalization scale factors in different control regions. These are compatible with one.

CDF Run II 7.8 fb^{-1} Preliminary

Control region	1S			SS			SJ		
TRM	1.005	±	0.003	1.025	±	0.009	1.031	±	0.006
QCD	0.971	±	0.013	1.085	±	0.046	1.06	±	0.035
EWK[1]	0.719	±	0.193	0.88	±	0.818	0.932	±	0.289
PreSel	1.008	±	0.014	0.966	±	0.032	0.978	±	0.018
$NN_{QCD} < 0.1$[2]	1.001	±	0.005	0.986	±	0.021	0.991	±	0.014
$0.1 < NN_{QCD} \leq 0.45$	1.034	±	0.023	0.962	±	0.052	0.977	±	0.034
$(NN_{QCD} > 0.45)$ (SR)	1.001	±	0.005	0.986	±	0.021	0.991	±	0.014
$NN_{QCD} > 0.7$	0.998	±	0.229	1.242	±	0.484	1.081	±	0.224

[1] The QCD MJ component is small in this region, and thus has a high uncertainty.
[2] NN_{QCD} is the output of an artificial neural network trained to separate the QCD background from the Higgs signal (cf. section 11.2.6).

8.7.5 Multijet Background Normalization

In order to estimate the backgrounds originating from QCD multijet production, we use the *Tag-Rate-Matrix* method described in section 8.7.1. This method provides an excellent model to describe the shapes of the background. Nonetheless the normalization of the background must be determined in a kinematic region closer to the final signal region than the region the TRM was derived in.

In order to constrain the expected rates of these backgrounds in the signal region (cf. section 14.2), we utilize the lower region of the NN_{QCD} output ($NN_{QCD} < 0.1$). This region is rich in QCD events and has sufficient statistics.

Once we are confident that the shapes are well reproduced by the matrix, we extract the normalization factor used in the final measurement by scaling the multijet prediction to data minus MC background minus mis-tags. The scale factor is close to and compatible with 1.0. Table 8.6 shows the scale factors for each event category (cf. section 11.3).

We propagate the uncertainty on the scale factor using the following equations

$$\mathrm{SF_{MJ}} = \frac{\mathrm{Data} - \mathrm{Expected\ Electroweak\ SM\ (EWK)}}{MJ} \tag{8.7}$$

$$\Delta\mathrm{SF_{MJ}} = \sqrt{\left(\frac{\Delta\mathrm{EWK}}{\mathrm{Data\ -\ EWK}}\right)^2 + \left(\frac{\Delta\mathrm{MJ}}{MJ}\right)^2} \tag{8.8}$$

where EWK is the predicted amount of electroweak events (from all processes), and is obtained from MC simulation, MJ is the QCD MJ prediction[14], ΔMJ is the statistical uncertainty on the MJ prediction and ΔEWK is the sum[15] of the statistical and systematic (rate only) uncertainty on its components.

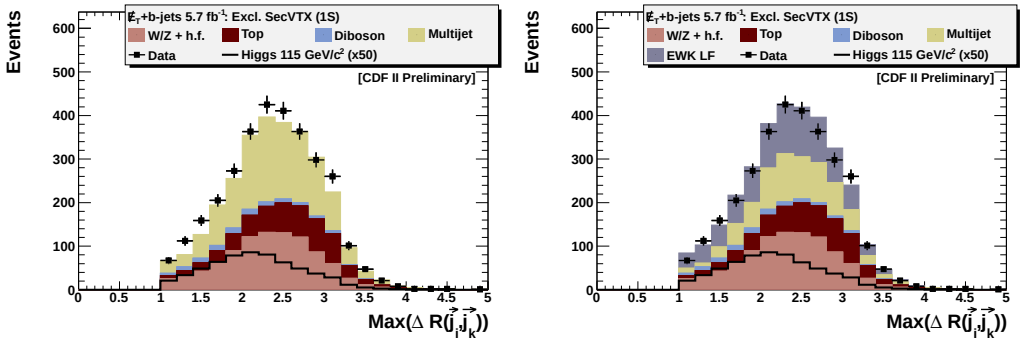
8.7.6 Improvement in background modeling

After many efforts aimed at understanding our backgrounds, and improving both our MC-driven and data-driven models, we appreciate the full extent of the modeling improvements. Figure 8.14 shows the improved modeling in the 5.7fb^{-1} signal region[16], i.e. with non relaxed cuts, for events with one SecVtx-tagged jet. The summer 2010 model follows [114, 161] and uses a data-driven model for QCD LF, QCD HF, and EWK LF (cf. section 8.7.2), while the summer 2011 data-driven model only models the QCD LF and HF, the EWK LF now being accounted for using mis-tag matrices (cf. sections 6.1.2 and 6.2.2). In this new model, the normalization scale factor derived in the previous section is independent of the EWK LF, whose relative contribution depends on the kinematic sample and is different in the low and high ends of the NN_{QCD} output.

[14]Which processes are subtracted from the raw TRM prediction depends on the model, as described in section 8.7.3.

[15]The sum is performed linearly or in quadrature depending on the nature (correlated or uncorrelated) of the source of uncertainty.

[16]The Alpgen model for W/Z+jets was introduced for the 5.7 fb^{-1} analysis, and yielded a significant improvement over the 2.1 fb^{-1} Pythia modeling [114, 161]. The relaxed kinematic requirements and the new data-driven model were finalized after the 5.7fb^{-1} analysis, which was became the benchmark for quantifying the improvement.



(a) Summer 2010 (b) Summer 2011

Figure 8.14.: Improvement in background modeling since the summer 2010 analysis. The comparison is done using 5.7fb^{-1} of data, and shows events with one SecVtx-tagged jet (1S) in the signal region. (a) Data-driven modeling for QCD LF mis-tags, QCD HF and EWK LF mis-tags (cf. figure 8.10a), and (b) data-driven modeling for QCD LF mis-tags and QCD HF, with EWK LF mis-tags derived from MC weighted by mis-tag matrices.

The improvement is clear, with the new modeling is in excellent agreement with the data. In the process, thanks to the increased statistics, we have been also able to reduce the systematic uncertainties on the simulation.

For simplicity, we merge the EWK LF and QCD (HF and LF) distributions into the *Multijet* background, keeping in mind that it also contains EWK LF mis-tags.

8.8 Validation of the Simulation

Now that we have described the techniques used to model the processes yielding events in the $\not\!\!E_T$+b-jets signature, let us reflect on what we have done. In simple terms, we have transposed our knowledge about the standard model into observables which can be compared against the experiment. How this comparison is done will be described in chapter 12.

Before that, it is important to understand the *limitations* of our models, and know how to behave accordingly. As asserted by the authors of Pythia,

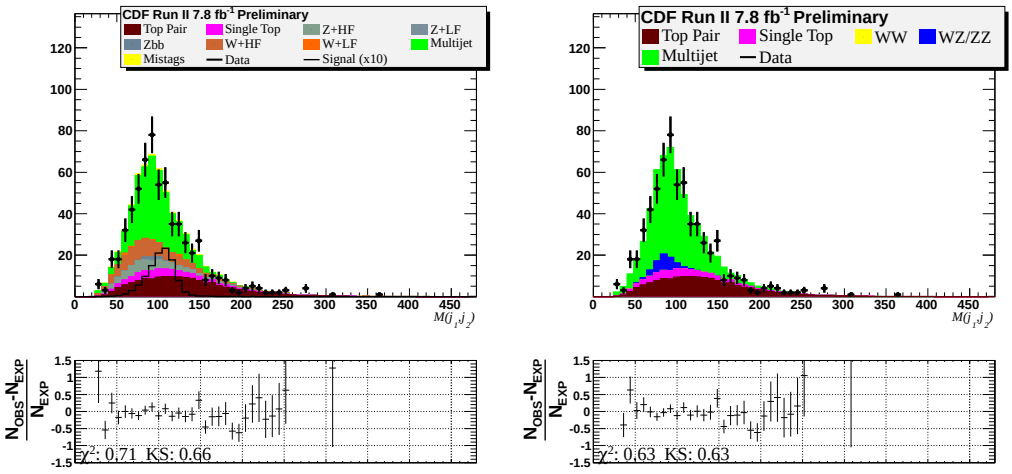
> (...) an event generator cannot be thought of as all-powerful, or able to give intelligent answers to ill-posed questions; sound judgement and some understanding of a generator are necessary prerequisites for successful use [111].

Indeed, one mustn't blindly accept any model at face value. Instead, models must be confronted to data in control samples which are known to be sensitive to the known phenomena which are being modeled, but are not expected to be sensitive to the physics phenomenon sought for. We already presented two of these regions, out of a total of six which are used in the analyses presented in this dissertation. The details regarding the control regions are presented in section 11.3.

8.9 Alternative Data-driven Model for Multijet Production

In the previous section, we presented a model for QCD multi-jet production which explicitly subtracts the contribution from W+jets and Z+jets processes, which are simulated using MC. We also investigated the possibility of using a single data-driven model for these three contributions, and the light flavor mis-tags, i.e. the model obtained from applying our TRM to a sample of taggable events. In this scenario, only single top, diboson, and top pair production (for which the MC simulation entails less systematic uncertainty) are subtracted.

The advantage of this technique is obvious: MC simulations have systematic uncertainties associated with them, of the order of 30% for W/Z+jets. However, we must find a way to properly estimate the systematic uncertainties of this data-driven technique; this is not easy. The question of the normalization of the data-driven estimate must also be addressed, especially since the largest source of background is now the data. Although very promising, this technique needs a careful estimation of the systematic uncertainty associated to it before it is used as the main model for multi-jet background in this signature. In particular, the estimation of the systematic uncertainty derived from the TRM would be too small to account for the variations



(a) (b)

Figure 8.15.: Comparison for events with two SecVtx-tagged jets of the main background modeling scheme of this analysis (cf. section 8.7.1), in which only the estimation of QCD is data-driven, with the alternative model from section 8.9, in which QCD, W/Z+jets, and mis-tags are modeled using one template. In (a), HF and LF refer to the *tree-level* processes in Alpgen; both can yield heavy flavor quarks through parton showering. Mistags refers to EWK LF mis-tags (cf. section 8.7.2).

in this complex, multi-compontent, background, as the relative contribution from QCD and W/Z+jets varies from sample to sample. Other analyses have used this technique [173], but not in the context of searches for small signals.

We have checked the performance of this technique in a sample with identified leptons (where W+jets and Z+jets are large contributors) and found good agreement. Figure 8.15 compares the modeling of section 8.7.1 with that of this section for events in our analysis region (cf. chapter 14) having two jets tagged by the SecVtx algorithm; we observe good agreement.

8.10 Summary

In this chapter, we discussed the processes contributing the the $\not\!\!E_T$+b-jets signature. We described the MC simulations used to estimate the production of electroweak

processes and the data-driven technique used to predict QCD multi-jet production. The latter technique is the state-of-the-art at the Tevatron.

9. NEURAL NETWORKS FOR PATTERN RECOGNITION

In the last years, there have been significant advancements in multivariate analysis (MVA) techniques. Decision trees (DT) and artificial neural networks (ANN) have allowed for such significant improvements in analysis sensitivity that there are more and more analyses using them. In this thesis, two key components of the analysis are handled using ANNs: the parameterization of the trigger efficiency, and the isolation and rejection of multi-jet QCD (MJ) events. Additionally, the final discriminants used to set a limit on the Higgs production cross-section or measure the single top or diboson production cross-sections are all the output of a neural network.

This chapter provides a more formal discussion on multivariate techniques – ANNs in particular[1] – with the goal to provide a motivation for their use, and in particular shed light on what is usually considered a *black-box* by many users (data analysts). In this discussion, we mostly follow [174].

9.1 Statistical Pattern Recognition

Data analysis in collider particle physics attempts to provide insight on the existence (or in-existence) of the physical process being investigated given a set of simulated data (cf. section 8.5) reflecting the current knowledge about a model (often the standard model, SM) and a dataset obtained from actual measurements in the detector. Using some kind of likelihood profile (cf. chapter 12), the analyst determines the compatibility level between the simulation and the observation. For some types of signal, this is relatively straightforward to achieve from a distribution of a physical

[1]We also investigated the use of boosted decision trees (BDT), which are closer to a traditional cut-based approach. However, we choose neural networks for their mathematical formalism and their relationship to probability density estimation.

property; often one scans a reconstructed invariant mass spectrum, which exhibits a peak that is characteristic of the physical process considered.

But there are cases where the mass spectrum is not sufficient by itself to achieve sufficient discrimination between the signal and the background processes. This is especially true when looking for small signals, i.e. with expected yields of the order of the uncertainty on the background, or smaller. These require combining information from several kinematic distributions in order to improve on the discrimination. The term *statistical pattern recognition* is well crafted to summarize this process.

9.2 Regression and Classification

Pattern recognition tasks are classified in two categories: regression and classification. The first consists in mapping a set of input variables $\boldsymbol{x} = (x_1, \ldots, x_N)$ to a continuous *regression function* $f(\boldsymbol{x})$ giving the average value at $\boldsymbol{x}$ of the distribution of the random variable g we seek to estimate (which is assumed to be correlated to $\boldsymbol{x}$). The second groups problems which map $\boldsymbol{x}$ to the output of *discriminant functions* $\boldsymbol{y} = (y_1, \ldots, y_c)$ representing the class labels, e.g. signal and background. Contrary to the regression case, the output variables take *discrete* values.

Both categories are particular cases of *function approximation*. In the case of regression problems, we approximate the regression function; for classification problems, we approximate the posterior probabilities of class membership $y_k = p(C_k|\vec{x})$ where C_k is the kth class and $\boldsymbol{x}$ the input vector. Therefore, many of the key issues related to pattern recognition are common to both regression and classification problems.

9.3 The Multilayer Perceptron

Parametric methods are the most straightforward approach to probability density estimation. These represent the probability density $p(\boldsymbol{x})$ in terms of a specific func-

tional form with a number of adjustable parameters. The latter are optimized by a fit to the data using *maximum likelihood* or *Bayesian inference*.

The maximum likelihood methods seeks to determine the most likely value for the parameter vector $\boldsymbol{\theta}$ given the observed data. Bayesian inference operates on a different basis. Before the data is observed, the parameters θ_i are described by a *prior* probability density, which reflects the prior belief of the value a given parameter should take. The prior probability density is often very broad, to reflect the fact that we have little idea of what values the parameters should take. Once the data is observed, Baye's theorem [175] allows determining the *posterior* probability density, which is narrower than it's prior counterpart. This process is known as *Bayesian learning*.

The idea behind artificial neural networks is to avoid to perform probability density estimations, and instead postulate specific parameterized functional forms for the discriminant functions and use the training data set to determine appropriate values for the parameters by minimizing an *error function* (cf. section 9.5).

We now proceed with the description of the key concepts behind artificial neural networks.

9.3.1 Single-layer Networks

Let us consider a two-category classification[2] problem. We seek to identify a discriminant function $y(\boldsymbol{x})$ such that $\boldsymbol{x}$ belongs to class $(C)_1$ if $y(\boldsymbol{x}) > 0$ and to $\mathcal{C}_2$ if $y(\boldsymbol{x}) < 0$.

The simplest choice of discriminant function consists of a linear combination of the d input variables, $\boldsymbol{x}$:

$$y(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + w_0 \tag{9.1}$$

[2]Event selection and separation are classification problems. We will address a regression problem when we describe the trigger efficiency parameterization in chapter 10.

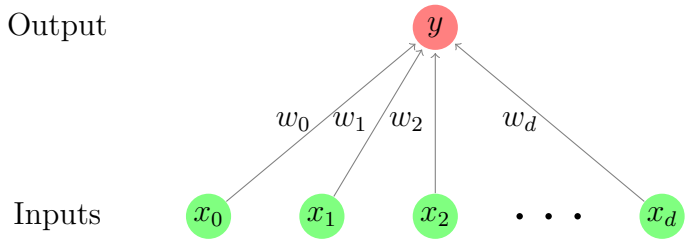


Figure 9.1.: Graphical representation of a linear discriminant function. Each component of the diagram corresponds to one term of (9.1). The bias w_0 can be considered as a weight parameter from an extra input whose value is permanently set to +1.

where the *bias*[3] w_0 and the *weight vector* $\boldsymbol{w}$ represent the parameters of the model. The geometrical interpretation of (9.1) is straightforward: the decision boundary $y(\boldsymbol{x}) = 0$ corresponds to a $(d-1)$-dimensional hyperplane in the input vector space. Since for any two points on the boundary, $y(\boldsymbol{x_A}) = y(\boldsymbol{x_B}) = 0$, we have that $\boldsymbol{w}^T(\boldsymbol{x_B} - \boldsymbol{x_A}) = 0$, and hence $\boldsymbol{w}$ is normal to any vector lying on this hyperplane; thus w determines the orientation of the decision boundary. The normal distance from the origin to the hyperplane, l, is given by

$$l = \frac{\boldsymbol{w}^T\boldsymbol{x}}{\|\boldsymbol{w}\|} = -\frac{w_0}{\|\boldsymbol{w}\|} \tag{9.2}$$

where $\boldsymbol{x}$ is any point on the hyperplane. In other words, the bias w_0 determines the position of the decision boundary.

It is convenient to re-write this expression by extending the input and weight vectors to $d+1$ dimensions:

$$y(\boldsymbol{x}) = \boldsymbol{w'}^T\boldsymbol{x'} \tag{9.3}$$

where $\boldsymbol{x'} = (1, \boldsymbol{x})$ and $\boldsymbol{w'} = (w_0, \boldsymbol{w})$. We shall only use this notation henceforth, assuming the presence of an additional input unit whose value is always 1.

Figure 9.1 shows a graphical representation of a linear discriminant function. Such a linear discriminant function can easily be generalized to N output nodes.

[3]In the literature, $-w_0$ is sometimes referred to as a *threshold*.

9.3.2 Activation Functions

It is possible to generalize this discriminant by applying a non-linear transformation expressed via an *activation function* g, so that

$$y = g(\boldsymbol{w}^T\boldsymbol{x} + w_0) \tag{9.4}$$

where g is usually chosen to be monotonic. A typical activation function is the *logistic sigmoid* s given by

$$g(a) = s(a) \equiv \frac{1}{1+\exp(-a)}. \tag{9.5}$$

This choice is motivated by the fact that using the sigmoid activation function allows the outputs of the discriminant to be interpreted as posterior probabilities. Indeed, assuming a two-class problem with the following class-conditional densities

$$p(\boldsymbol{x}|\mathcal{C}_k) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_k})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_k})\right) \tag{9.6}$$

where we assume for simplicity that both classes have equal covariance matrix $\boldsymbol{\Sigma}$, and using Bayes' theorem, we can express the posterior probability of membership of class $\mathcal{C}_1$

$$p(\mathcal{C}_1|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\mathcal{C}_1)P(\mathcal{C}_1)}{p(\boldsymbol{x}|\mathcal{C}_1)P(\mathcal{C}_1) + p(\boldsymbol{x}|\mathcal{C}_2)P(\mathcal{C}_2)} \tag{9.7}$$

which can be rewritten as

$$p(\mathcal{C}_1|\boldsymbol{x}) = \frac{1}{1+\exp(-a)} = s(a) \tag{9.8}$$

where

$$a = \ln\frac{p(\boldsymbol{x}|\mathcal{C}_1)P(\mathcal{C}_1)}{p(\boldsymbol{x}|\mathcal{C}_2)P(\mathcal{C}_2)}. \tag{9.9}$$

Substituting the class-conditional densities (9.6) in (9.9), we obtain, using matrix calculus and the fact that the covariance matrix is symmetric,

$$a = \boldsymbol{w}^T \boldsymbol{x} + w_0 \tag{9.10}$$

where

$$\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \tag{9.11}$$

$$w_0 = -\frac{1}{2}\boldsymbol{\mu_1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu_1} + \frac{1}{2}\boldsymbol{\mu_2}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu_2} + \ln \frac{P(\mathcal{C}_1)}{P(\mathcal{C}_2)}, \tag{9.12}$$

showing the probabilistic interpretation of the discriminant (9.4).

An alternative choice for the activation function is the hyperbolic tangent

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}, \tag{9.13}$$

which provides a similar scaling to the sigmoid function but with a range from -1 to 1, instead of from 0 to 1. For this reason, it is often used in place of the sigmoid function. It should be noted that these two activation functions differ only through a linear transformation, for an activation function $\tilde{g}(b) = \tanh(b)$ is equivalent to an activation $g(a) = s(a)$ if we apply the tranformation $b = a/2$ to the input and $\tilde{g} = 2g - 1$ to the output. Therefore, a network using tanh activation functions behaves as a network using logistic sigmoid activation functions, although the weights and biases have different values. Empirically, the hyperbolic tangent activation function yields faster convergence of the training algorithms [174].

The Heaviside step function

$$g(a) = \begin{cases} 0 & \text{when } a < 0 \\ -1 & \text{when } a \geq 0 \end{cases} \tag{9.14}$$

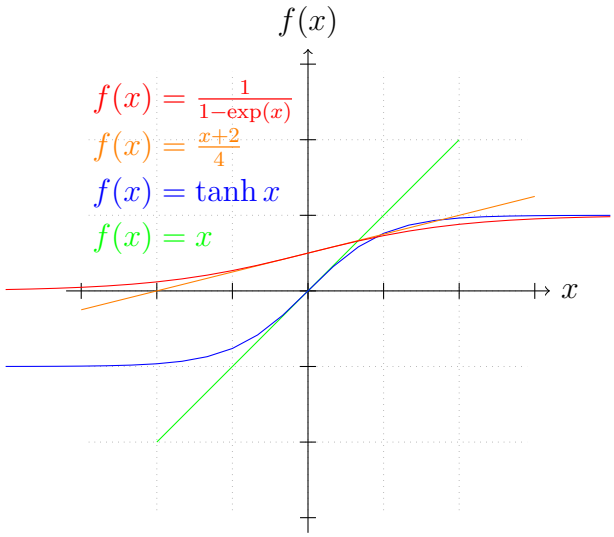


Figure 9.2.: Logistic sigmoid and hyperbolic tangent activation functions together with their linear approximations around the origin.

could be another alternative, which however has the disadvantage to have a ill-behaved derivative, and therefore is not suited for our purpose[4]. Around the origin, all sigmoid functions, including (9.5) and (9.13), behave like a linear transformation, the simplest case of activation function; this is shown in figure 9.2, together with the logistic sigmoid function and the hyperbolic tangent.

One of the main advantages of sigmoid functions is that the determination of their derivatives at any given point only involves simple arithmetic operations once the sigmoid function is known at that given point, for

$$\tanh'(x) = 1 - \tanh^2(x) \tag{9.15}$$

and

$$s'(x) = s(x)\,[1 - s(x)]\,. \tag{9.16}$$

[4]We shall see that the derivative of the activation function has a role in the determination of the weights.

9.3.3 Multilayer Perceptron

The linear discriminant function discussed in the previous section is a network with a single layer of weights. More general mappings of the inputs to the output(s) are allowed by considering successive transformations corresponding to networks having several layers of weights. In fact, networks with two layers of weights can approximate any continuous functional mapping. To ensure that the network outputs can be calculated as explicit functions of the inputs and the weights, we must restrict ourselves to *feed-forward* networks, i.e. without feedback loops. Such networks are referred to as *multilayer perceptrons* in the literature. Their mathematical form is readily obtained from the above, and yields

$$y_k = \tilde{g}\left(\sum_{j=0}^{M} w_{kj}^{(2)} \cdot z_j\right) = \tilde{g}\left(\sum_{j=0}^{M} w_{kj}^{(2)} \cdot g\left(\sum_{i=0}^{d} w_{ji}^{(1)} x_i\right)\right),\ k = 1 \ldots N \qquad (9.17)$$

where we have again incorporated two bias nodes $x_0 = 1$ and $z_0 = 1$, as shown in figure 9.3. The activation function $\tilde{g}$ must not necessarily be the same as g. The network described in figure 9.3 is a multilayer perceptron with d input nodes, M nodes in the hidden layer, and N output nodes.

9.3.4 Choice of Network Topology

The two-layer network topology shown in figure 9.3 can approximate any functional continuous mapping up to arbitrary precision provided that the number M of hidden nodes is sufficiently large. A corollary in the context of classification problems is that two-layered networks with sigmoidal activation functions can approximate any decision boundary with arbitrary accuracy [174].

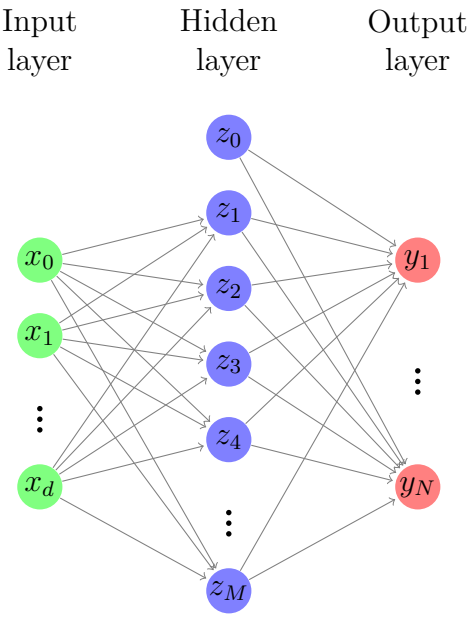


Figure 9.3.: Topology of a multilayer perceptron with d input nodes, M nodes in the hidden layer, and N output nodes.

We present here a simple proof of the concept. According to Fourier analysis, it is possible to approximate any continuous function with arbitrary accuracy by a Fourier decomposition. The case of two input variables, x_1 and x_2, yields

$$y(x_1, x_2) = \sum_n \sum_m A_n m \cos(m x_1) \cos(n x_2) \tag{9.18}$$

where A_{nm} are constants. Using the trigonometric identity

$$\cos a + \cos b = \frac{1}{2} \left(\cos(a+b) + \cos(a-b)\right) \tag{9.19}$$

and since any function can be approximated by a piecewise constant function, i.e. a sum of Heaviside step functions[5], of the form

$$f(z) = f_0 + \sum_{i=0}^{N} \{f_{i+1} - f_i\} H(z - z_i), \tag{9.20}$$

it is clear that $y(x_1, x_2)$ can be expressed by a two-layer network with N threshold hidden units and linear output units. This result can be expanded to an arbitrary number of input variables. However, such an expansion does not approximate the *derivatives* of the function, since the expansion (9.20) does have zero derivative everywhere except at discrete points where it is undefined.

Two-layer networks with sigmoidal hidden units, however, can simultaneously approximate both a function and its derivatives up to arbitrary precision, provided that there are enough hidden nodes [176, 177]. But because we cannot easily implement hidden layers with a large number of nodes, we resort to a network topology with two hidden layers (three layers in total) in this dissertation; this configuration provides enough flexibility while being affordable to train.

[5] Arbitrary accuracy is obtained by reducing the step size, thus increasing the terms in the sum.

9.4 High-order Generalization

While it might be tempting to generalize networks to higher order, e.g., to

$$y(\boldsymbol{x}) = \boldsymbol{w^1}^T\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{W^2}\boldsymbol{x} + \sum_{n=3}^{N}\left(\sum_{i_1=0}^{d}\dots\sum_{i_n=0}^{d} W^n(i_1,...,i_n)\cdot x(i_1)\dots x(i_n)\right) \quad (9.21)$$

where $\boldsymbol{w^1}$ is a one-dimensional weight vector and $\boldsymbol{W^2}$ a two-dimensional weight matrix, and W^n a n-dimensional weight matrix, this turns out to be computationally prohibitive, for the number of weights increases rapidly with the order of the network [178]. However, such networks can be made invariant to various transformations of the inputs by imposing an appropriate set of constraints to the weights. For instance, a third-order unit can be invariant to translations, rotations and scalings of the input patterns drawn from pixels in a two-dimensional usage. These type of networks are far beyond the scope if this chapter, as they are not necessary for our purpose.

9.5 Network training

The central goal of network training is not memorizing the training set, but to model the *underlying generator* of the set. Indeed, we want the network to provide the best possible prediction for the output vector $\boldsymbol{t}$ when it is presented with a new, never used before, value of the input vector $\boldsymbol{x}$.

The most general (and complete) description of the generator of a dataset is the probability density $p(\boldsymbol{x},\boldsymbol{t})$ in the input-target space. We can decompose that probability density into the product of the conditional density of the target data, conditioned on the input data, and the unconditional density of the input data,

$$p(\boldsymbol{x},\boldsymbol{t}) = p(\boldsymbol{t}|\boldsymbol{x})\cdot p(\boldsymbol{x}), \quad (9.22)$$

where $p(\boldsymbol{t}|\boldsymbol{x})$ represents the probability density of $\boldsymbol{t}$ given that $\boldsymbol{x}$ takes a particular value. This is the term that needs to be modeled for the purpose of making predictions. The unconditional density of $\boldsymbol{x}$

$$p(\boldsymbol{x}) = \int p(\boldsymbol{t}, \boldsymbol{x}) d\boldsymbol{t} \tag{9.23}$$

is of more relative interest, since it plays a role limited the choice of the training method.

The training involves an iterative procedure for minimization of an error function, making adjustments to the weights at each step. Each step is characterized by two stages. The first is the evaluation of the derivatives of the error function with respect to the weights; the back-propagation technique (section 9.5.1) provides a computationally efficient method for achieving this. During the second stage, the derivatives are used to compute the adjustments to be made to the weights.

Networks with differentiable activation functions allow for a powerful, and computationally efficient, method, *error back-propagation* (section 9.5.1) for finding the derivatives of an error function with respect to the weights in the network.

9.5.1 Error back-propagation

Now we have established the capabilities of neural networks, we consider how such networks can learn a suitable mapping from a given dataset. The learning process depends on a *error function*, assumed to be suitably defined for our problem. We discuss error functions in more detail in section 9.6.

The main idea is that, for networks with differentiable activation functions, the output units are differentiable functions of both the input variables, the weights and the biases. Hence, an error function that is a differentiable function of the outputs is a differentiable function of the weights. Thus, we can evaluate the derivatives of the error with respect to the weights, with the intent to find weight values which minimize the error function. For this purpose, one can use simple gradient descent or more

powerful optimization methods (section 9.7). The process of evaluating the derivates of the error function is called *back-propagation*, for it corresponds to a propagation of errors backwards through the network. Assuming a general network with arbitrary feed-forward topology, arbitrary differentiable non-linear activation functions, and an arbitrary differentiable error function E, we proceed to the derivation of the back-propagation algorithm.

In the general feed-forward network considered here, each unit computes a weighted sum of its inputs

$$a_j = \sum_i w_{ji} z_i \tag{9.24}$$

where z_i is the activation of a unit, or input, which sends a connection to the output j, and w_{ji} is the weight associated to that connection. This sum is transformed by a non-linear activation function g to yield the activation $z_j = g(a_j)$ of unit j. We consider separate error functions for each pattern, which can be summed to obtain the total error

$$E = \sum_n E^n(y_1, \ldots, y_c) \tag{9.25}$$

where we assume n patterns and c output nodes. For simplicity, we now consider one pattern at the time, and omit the superscript; the generalization only involves summing the error derivatives for each pattern.

We suppose that the activations of all the hidden and output units have been calculated from the input vector (this process is called *forward propagation*), and evaluate the derivative of the error (for a given pattern) with respect to a weight w_{ji}

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial a_j}\frac{\partial a_j}{\partial w_{ji}} = \delta_j \frac{\partial a_j}{\partial w_{ji}} = \delta_j z_i \tag{9.26}$$

where we apply the chain rule for derivatives, introduce the *error* δ_j, and make use of (9.24). The evaluation of δ_k for the output units is straightforward:

$$\delta_k \equiv \frac{\partial E}{\partial a_k} = g'(a_k)\frac{\partial E}{\partial y_k}. \tag{9.27}$$

For the hidden units, δ_j is given by

$$\delta_j \equiv \frac{\partial E}{\partial a_j} = \sum_k \frac{\partial E}{\partial a_k}\frac{\partial a_k}{\partial a_j} = g'(a_j) \cdot \sum_k w_{kj}\delta_k \tag{9.28}$$

where the sum runs over all units k to which unit j sends connections. This equation is the *back-propation* formula, and allows to obtain the value of all δ's starting from the δ's of the output units, which are known.

The back-propagation procedure can be summarized as follows:

1. Apply an input vector $\boldsymbol{x^n}$ and forward-propagate through the network to find the activations of all the hidden and output units;
2. Evaluate δ_k for all output units;
3. Back-proagate the δ's to obtain one for each hidden unit;
4. Evaluate the required derivatives.

With the knowledge of all the partial derivatives of the error function, one can proceed with its minimization, in order to obtain the most adequate weights for the network.

9.5.2 Practical Aspects of Network Training

As in every iterative process, some choice must be made of when to stop the training process, in order to avoid infinite loops, or even divergence. Some possible choices are to

Stop after a fixed number of iterations. This is a very common approach to network training. The problem here is to know in advance how many iterations are needed; but preliminary test can help alleviate this.

Stop when the error function falls below some specified value. While intuitively making sense, the specified value may never be reached. This criterion should be doubled by some limit on the CPU (central processing unit) time.

Stop when a relative change in the error function falls below some specified value. Although making sense for most cases, this may lead to a premature termination of the training if the error functions decreases slowly during some part of the training (e.g., near a saddle point).

Stop training when the error measured using an independent validation set starts to increase. This strategy is often used to avoid *over-training*, i.e. the fine-tuning of the weights to the particularities (features) of the training sample, which is not present in other, independent, samples.

Stop when a pre-determined amount of CPU time has been used. Besides the difficulty of determining what constitutes an appropriate time, there is the possibility that over-training has occurred, or that the algorithm spend much time in a relatively flat region.

As one can see, there is no choice proving clear advantages with respect to the others, and therefore one uses a combination of criteria in practice. We choose to stop the training after a fixed number of iterations (often 1000), and check that the error measured using an independent validation set has not diverged from that measured on the training sample; we check that both have reached a plateau.

9.6 Error function

When training any kind of multivariate analysis tool, it is crucial to monitor the performance of the tool. In most cases, the important criterion is the departure from the true target value for a given input vector[6]. This departure can be determined in many ways, depending on the particular application; it is provided by an *error function*.

[6]More complex criteria, such as the sensitivity of a given output towards a given process, can be considered, provided that there is a prescription to adapt the weights given the departure from the criteria, and that this prescription converges.

Most error functions are motivated from the principle of maximum likelihood. For a set of training data $\{\boldsymbol{x^n}, \boldsymbol{t^n}\}$, the likelihood can be expressed as

$$\mathcal{L} = \prod_n p(\boldsymbol{x^n}, \boldsymbol{t^n}) = \prod_n p(\boldsymbol{t^n}|\boldsymbol{x^n}) \cdot p(\boldsymbol{x^n}) \tag{9.29}$$

where each data point $(\boldsymbol{x^n}, \boldsymbol{t^n})$ is assumed to be drawn independently, allowing the probabilities to be multiplied. We wish to maximize this likelihood or, equivalently, minimize its negative logarithm. We therefore minimize the error function

$$E = -\ln \mathcal{L} = -\sum_n \ln p(\boldsymbol{t^n}|\boldsymbol{x^n}) - \sum_n \ln p(\boldsymbol{x^n}) \tag{9.30}$$

where the second term does not depend on the network parameters and represents an additive constant which can be dropped from the error function.

From this, different assumptions about the form of the conditional distribution $p(\boldsymbol{t}, \boldsymbol{x})$ lead to different choices of error function. In regression problems, the targets $\boldsymbol{t}$ consist of continuous quantities whose values we try to predict, while in classification problems, they represent labels defining class membership.

9.6.1 Error Surface

The ultimate goal of network training, as discussed in section 9.5, is to find a weight vector w which minimizes an error function $E\boldsymbol{w}$. In this context, a simple geometrical picture of the error minimization process is very helpful, and allows to connect the results discussed below with intuition and known concepts of vector calculus.

We can view $E(\boldsymbol{w})$ as an *error surface* above the weight space. In the most general case, where the error function is a highly non-linear function of the weights, the error surface may have many[7] minima satisfying $\nabla E = 0$, where ∇E is the gradient of E in weight space. At least on of these *local minima* yields a value for E smaller than

[7]This case refers also to the case of one single minimum.

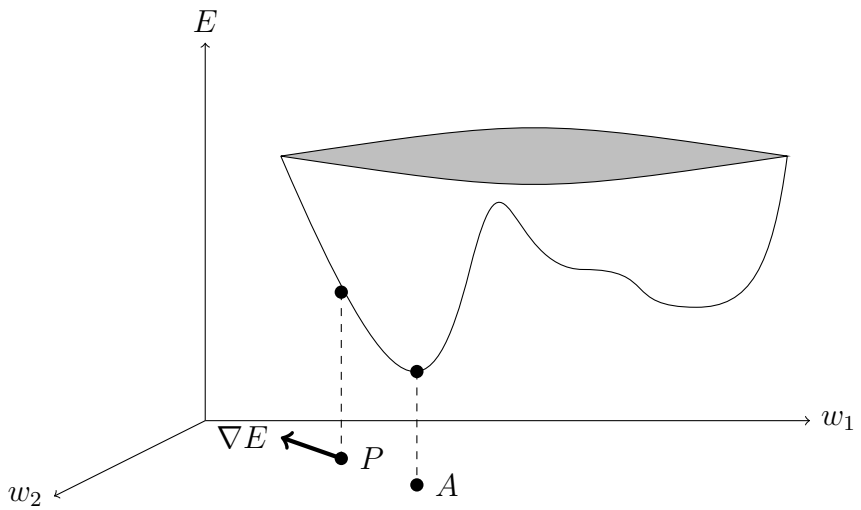


Figure 9.4.: Geometrical interpretation of $E(\boldsymbol{w})$. Point A is the global minimum. At any point P, the local gradient of the error function is given by ∇E. Inspiration source: [174]

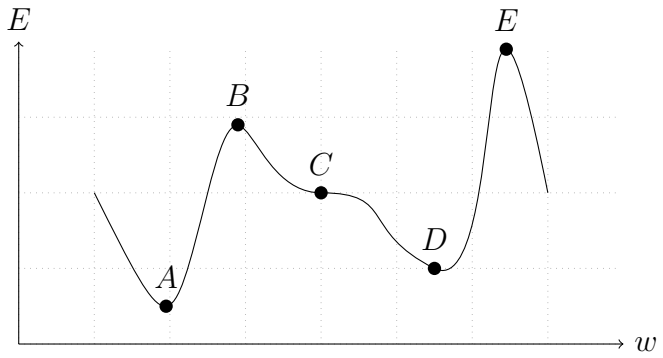


Figure 9.5.: Schematic error function of a single weight w, showing five stationary points, where $dE/dx = 0$). Point A is the global minimum, point B is a local maximum, point C is a saddle point, point D is a local minimum, and point E is the global maximum. Inspiration source: [174]

the others, and is called the *global minimum*. It is worth noting that $\nabla E = 0$ also characterizes other types of*stationary points* such as maxima and saddle points (see figure 9.5).

Because of the non-linearity of the error function, it is in general not possible to find closed-form solutions for the global minimum (or any minimum). Instead, we have to resort to algorithms which search through weight space in successive steps of the form

$$\boldsymbol{w}^{\nu+1} = \boldsymbol{w}^{\nu} + \Delta\boldsymbol{w}^{\nu} \tag{9.31}$$

where ν labels the iteration step. The different training algorithms involve different choices for the weight increment $\Delta\boldsymbol{w}^{\nu}$.

9.6.2 Sum-of-squares error

In this section, we derive the sum-of-squares errors from the principle of maximum likelihood. Let us consider the case of c independent[8] target variables t_k where $k = 1, \ldots, c$. Then we can write

$$p(\boldsymbol{t}|\boldsymbol{x}) = \prod_{k=1}^{c} p(t_k|\boldsymbol{x}). \tag{9.32}$$

Let us assume that the target variable t_k is given by some deterministic function of $\boldsymbol{x}$ with added Gaussian noise, so that

$$t_k = h_k(\boldsymbol{x}) + \varepsilon_k, \tag{9.33}$$

where ε_k is distributed according to a normal distribution with zero mean and standard deviation σ independent of $\boldsymbol{x}$ or k, i.e.

$$p(\varepsilon_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\varepsilon_k^2}{2\sigma^2}). \tag{9.34}$$

We seek to model the functions $h_k(\boldsymbol{x})$ by a neural network with outputs $y_k(\boldsymbol{x}; \boldsymbol{w})$, where $\boldsymbol{w}$ is the set of weight parameters governing the network mapping. We can now express the probability distribution of target variables as

$$p(t_k|\boldsymbol{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{[y_k(\boldsymbol{x}; \boldsymbol{w}) - t_k]^2}{s\sigma^2}\right) \tag{9.35}$$

where we have replaced $h_k(\boldsymbol{x})$ by the network model $y_k(\boldsymbol{x}; \boldsymbol{w})$. This leads to the following expression for the error function:

$$E = \frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{k=1}^{c} [y_k(\boldsymbol{x^n}; \boldsymbol{w}) - t_k^n]^2 + Nc \ln \sigma + \frac{Nc}{2} \ln(2\pi). \tag{9.36}$$

[8]Independence is here taken to mean that the distributions of the different target variables are independent.

Noting that, for our purpose of error minimization, the second and third terms, as well as the overall factor of σ^{-2} in the first term, can be omitted, we obtain the sum-of-squares error function

$$E = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{c}[y_k(\boldsymbol{x^n};\boldsymbol{w}) - t_k^n]^2 = \frac{1}{2}\sum_{n=1}^{N}||\boldsymbol{y}(\boldsymbol{x^n};\boldsymbol{w}) - \boldsymbol{t^n}||^2. \tag{9.37}$$

While we have assumed Gaussian distributed target data in (9.34) for the derivation of the sum-of-squares error from the principle of maximum likelihood, this error function does not require the target data to have such a distribution. However, it is worth noting that a sum-of-squares error function does not allow to distinguish between the true distribution and any other distribution with the same mean and variance. But this is of no consequence to our purpose.

The sum-of-weights error is very appropriate for regression problems, where the goal is to model the conditional distribution of the output variable. Classification problems have another goal: to model the posterior probabilities of class membership, also conditioned on the input variables. For these problems, it is also possible to use a sum-of-square error, but there are other functions, which are more appropriate. However, since there are currently no practical implementations available to the HEP community other than for sum-of-square errors, we do not discuss these further. The interested reader is invited to consult reference [174].

9.7 Parameter Optimization Algorithms

The problem of minimizing continuous differentiable functions of many variables has been widely studied, for it has many applications in a broad range of fields. For this reason, many of the conventional approaches to this problem are applicable to the training of neural networks. In this section, we review the optimization algorithm used while training the networks used in this dissertation, namely gradient descent, and briefly discuss a few other, more advanced, algorithms.

In order mimick biological networks or to parallelize network training, it is tempting to compute updates to a given weight only using information locally available. But we shall not concern ourselves with this consideration, so that the algorithms presented below require a complete knowledge of the state of the network (i.e. the value of every node, weight, and bias).

9.7.1 Gradient Descent

Gradient descent is the simplest training algorithm, and consist in the intuitive solution, which is to start with some initial guess $\boldsymbol{w}^0$ (often chosen at random) and iteratively update the weight vector, moving a short distance in the direction of the greatest rate of decrease of the error, i.e. in the direction of the negative gradient, so that

$$\Delta \boldsymbol{w}^\nu = -\eta \nabla E \Big|_{\boldsymbol{w}^\nu} \tag{9.38}$$

where η is the *learning rate*. This is why this algorithm is sometimes referred to as *steepest descent*. An alternative, pattern-based, gradient descent algorithm consists in the minimization of the error associated with one pattern at the time, and taking the gradient of E^n instead of that of E.

The choice of learning rate parameter η is critical to ensure convergence of the algorithm. Chosen too small, it increases the convergence time (sometimes drastically); if too big, it could lead to the weight vector passing over a narrow global minimum. Additionally, it is possible for the weight vector to be stuck around a local minimum, without ever getting the opportunity to jump to another region where there might be a global minimum if the learning rate is too small. It is therefore recommended to restart the training procedure with a different initial weight vector, and compare the resulting outputs.

The convergence of the gradient descent is also hindered in special cases in which the curvature of E varies significantly with the direction, so that at most points, the

local gradient doesn't point towards the minimum. The gradient descent procedure is very inefficient in these cases, although it is not prevented from converging.

Because of its simplicity, it is very easy to implement, and available with all known neural network implementations. We now proceed with a brief description of other techniques for the sake of more completeness.

9.7.2 Advanced Optimization Algorithms

Advances over the gradient descent algorithm include heuristic modifications to improve performance, conjugate gradient methods, and quasi-Newton methods, to name three general categories. Several of these advanced algorithms require evaluating higher-order derivatives. For this purpose, the Jacobian $\mathcal{J} = (\partial y_k / \partial \delta x_i)$ and the Hessian $\mathcal{H} = (\partial^2 E / \partial w_{ji} \partial w_{lk})$, and its eigenvalues in particular, are very useful. Several techniques have been developed to minimize the computation steps needed to compute these, often based on the form in which these matrices are to be used in calculations.

Improved gradient descent

For some category of problems, the convergence is poor, e.g., because the smallest eigenvalue of the Hessian is much smaller than its largest eigenvalue. In these cases, one can provide alternatives for the distance in weight space between each iteration. A simple example is the addition of a *momentum* term, which adds inertia to the motion through weight space, and can accelerate the convergence. The modified gradient descent in this case is

$$\Delta \boldsymbol{w}^{\nu} = -\eta \nabla E \Big|_{\boldsymbol{w}^{\nu}} + \mu \Delta \boldsymbol{w}^{\nu-1} \tag{9.39}$$

where μ is the momentum parameter. In the approximation that the gradient is unchanging, the addition of the momentum term corresponds to an increase of the

effective learning rate, from η to $\eta/(1-\mu)$. The momentum term can lead to faster convergence, without causing divergent oscillations.

Another alternative is the *enhanced gradient descent*, in which the learning rate is adjusted according the the following rule

$$\eta_{\text{new}} = \begin{cases} r\eta_{\text{old}} & \text{if } \Delta E < 0 \\ s\eta_{\text{old}} & \text{if } \Delta E > 0 \end{cases} \tag{9.40}$$

where r and s are chosen to be respectively slightly larger and significantly lower than unity. The rationale behind this choice is that the algorithm can converge faster when it is progressing towards lower values of E, yet allowing to revert quickly when it overshoots the minimum.

Line search

The error minimization algorithms presented in this chapter involve a sequence of steps through weight space. We can separate these steps in two parts. The first consists in deciding the *direction* in which to move, and the second in determining the *step size*, i.e. how far to move in that direction. In the case of gradient descent, the direction is given by the local negative gradient, and the step size is determined by the learning rate.

A better procedure would be to move along the chosen direction and find the point at which the error is minimized. Such procedure is referred to as *line search*, and is the basis for several algorithms. The idea is to update the current weight vector $\boldsymbol{w}^\nu$ following

$$\boldsymbol{w}^{\nu+1} = \boldsymbol{w}^\nu + \lambda^\nu \cdot \boldsymbol{d}^\nu \tag{9.41}$$

where the parameter λ^ν minimizes the error

$$E(\lambda) = E(\boldsymbol{w}^\nu + \lambda \cdot \boldsymbol{d}^\nu). \tag{9.42}$$

This procedure is shown in figure 9.6 using parabolic interpolation. The computational cost for minimizing the error function along a given direction depends on the accuracy required. A trade-off is required between this cost and that of additional line search iterations (i.e. choosing a direction and minimizing the error along it).

Now that we have determined a procedure to find the minimum of the error along a given line, we proceed with the description of two classes of methods used to choose the search direction.

Conjugate gradient

The conjugate gradient technique minimizes each component of the vector independently. At each step of the algorithm, the component of the gradient parallel to the previous search direction is unaltered (to lowers order). Assuming we have just minimized the error along the direction $\boldsymbol{d}^{\nu}$ starting from $\boldsymbol{w}^{\nu}$, we require

$$\boldsymbol{g}(\boldsymbol{w}^{\nu+1})^{T}\boldsymbol{d}^{\nu} = \mathbf{0} \tag{9.43}$$

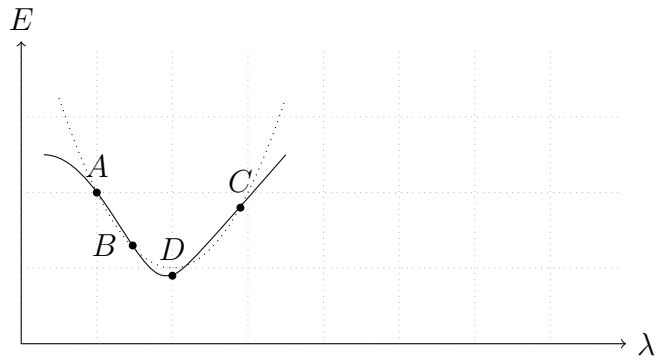


Figure 9.6.: Parabolic interpolation used for line search minimization. The solid curve represents the error as a function of the distance λ along the search direction. A parabola (dotted curve) is fitted to the points A, B, and C, chosen so that $A < B < C$ and $E(A) > E(B)$ and $E(C) > E(B)$. The location of the minimum of the parabola at D gives an approximation of the minimum of $E(\lambda)$. This process can be repeated using the points B, C, and D, improving the estimation of the minimum at the cost of increased computation time. Inspiration source: [174]

where $\boldsymbol{g} \equiv \nabla E$. The next search direction $\boldsymbol{d^{\nu+1}}$ is determined such that

$$\boldsymbol{g(w^{\nu+1} + \lambda d^{\nu+1})^T d^{\nu} = 0.} \tag{9.44}$$

To first order, this requires that

$$\boldsymbol{d^{\nu+1T} H d^{\nu} = 0} \tag{9.45}$$

where $\boldsymbol{H}$ is the Hessian matrix evaluated at $\boldsymbol{w}^{\nu+1}$.

The repetitive evaluation of the Hessian matrix can be computationally prohibitive, especially for non-quadratic error functions, for which $\boldsymbol{H}$ is not constant. Several algorithms have been developed to speed up this evaluation; these are based on the observation that we can determine the search direction without the explicit knowledge of the Hessian itself [174].

Because we do not make use of the conjugate gradient method in this dissertation due to the lack of suitable code libraries, we do not develop it further, and redirect the reader to the literature [174, 179].

Newton's method and quasi-Newton methods

Contrary to the conjugate gradient method, which makes implicit use of second-order information about the error surface (represented by the Hessian matrix), Newton's method makes explicit use of $\boldsymbol{H}$. Using quadratic approximation, the location of the minimum (more generally a stationary point) can be directly obtained from the error function. At any point $\boldsymbol{w}$, the gradient is given by

$$\boldsymbol{g} = \nabla E = \boldsymbol{H}(\boldsymbol{w} - \boldsymbol{w}) \tag{9.46}$$

where the vector $\boldsymbol{w}^*$ corresponds to the minimum of the error function and satisfies

$$\boldsymbol{w}^* = \boldsymbol{w} - \boldsymbol{H}^{-1}\boldsymbol{g}. \tag{9.47}$$

The direction $-\boldsymbol{H}^{-1}\boldsymbol{g}$ is known as the *Newton direction*. Unlike the local gradient vector, the Newton direction for a quadratic error surface points directly at the minimum of the error function.

The main difficulties related to such approach at computational. Indeed, the exact evaluation of the Hessian matrix for non-linear networks is computationally demanding, and is computationally prohibitive if done at each iteration. Additionally, the inversion of $\boldsymbol{H}$ is also very demanding. Finally, this algorithm can move towards a maximum or saddle point when the Hessian is not positive-definite, and thus the error is not guaranteed to be reduced at each iteration. It is also worth noting that the step size might be too large to take us outside the validity of the quadratic approximation, which leads to an unstable algorithm.

Quasi-Newton methods overcome the computational issues related to the Newton method by building an approximation to the inverse Hessian over a number of steps. These involve generating a sequence of matrices $\boldsymbol{G}^{\nu}$ which represent increasingly accurate approximations of $\boldsymbol{H}^{-1}$ using only information on the first derivatives of the error function. At each step, care is taken to ensure that $\boldsymbol{G}^{\nu}$ remains positive-definite. A commonly used update formula for $\boldsymbol{G}^{\nu}$ is the Broyden-Fletcher-Goldfarb-Shannon (BFGS) procedure [180–183]. This method usually requires a smaller number of iterations. However, because the computing time for one iteration is proportional to the squared number of synapses, large networks are particularly penalized [184]. We therefore did not retain the technique for this dissertation.

9.8 Network Ensembles

Given the random choice of the initial weight vector, one rightfully can ask whether different starting points converge to the same minimum of the error function. Other analysts [185] therefore use *network ensembles*, averaging out the effects of individual networks. The rationale is that by using averages, the fluctuations in the observed limits are reduced.

We do not follow this procedure for the reason that wether the output of the network is a true (global) minimum of the error function or not is of little practical interest to us, since our main goal is to *compare* the distribution of the network output in data an in the simulation[9]. For this reason, we believe that any network output is a valid one, provided that it is not prone to over-training and that it provides a good agreement between the data and the simulation in the control regions (section 11.3).

While we do check the output of several different trainings, we do only retain the first, in the case of similar outputs; otherwise we re-train the network, modifying its topology or the parameters of the training. Our motivation for claiming that this is enough for our purpose is that the uncertainty related to the network output, which is kept under visual control, is small compared to the uncertainties on the background. Moreover, any discrepancy in the network output would propagate to each process in the same way, changing very little to the likelihood function.

9.9 Learning and Generaliation: Bias vs. Variance

As mentioned before in this chapter, the goal of network training is not to learn an exact representation of the training data, but rather to build a statistical model of the process which generates the data. This concern can be simply viewed in figure **??**, where we see that a high-order polynomial fitting the noisy data exactly provides poor generalization, and that this is the same for a network with too few terms, i.e. little flexibility.

This highlights that the complexity (number of coefficients in the polynomial) must be optimized in order to achieve the best generalization. This is better view in light of the concept of *bias-variance trade-off*, in which the generalization error is the sum of the *bias* squared plus the *variance*. A model which is too simple and inflexible will have large bias, while one with too much flexibility will have too much vari-

[9]Although we clearly favor solutions that are true minima of the error function, the point is here that this is not our primary goal.

ance. The optimal is the compromise somewhere in between. Several regularization techniques exist to help the network training converge towards this compromise [174].

9.10 NeuroEvolution of Augmenting Topologies

In the past sections of this chapter, we discussed fixed-size multilayer perceptrons, and how to optimally determine the weights required to approximate a given functional form. We presented several algorithms to achieve this goal.

An interesting concept in machine learning[10] is that of *genetic algorithms*, a form of *evolutionary algorithms*, which mimic the process of natural evolution in order to find optimal answers to given problems.

Evolutionary algorithms generate solutions to optimization problems using: (a) inheritance (crossover); (b) mutation, and (c) selection. In practice, candidate solutions to the optimization problem are individuals in a population. Through the process of reproduction, two of these individuals are merged to create one (or more) new individuals, which, just like human individuals, inherit properties from their parents through crossover operations. Additionally, these individuals can mutate in order to reach another potentially interesting solution. Each individual is assigned a *fitness* which quantifies how well the solution is fit for the problem.

The individuals from each generation that have the (usually[11]) highest fitness are allowed to live and produce offsprings. The training continues for multiple generations until some stopping criterion is met (similar to those discussed in section 9.5.2). Although this algorithm could in principle be used to modify only the weights of the network population, it is more powerful when allowed do alter the *structure* of the

[10]Machine learning is a discipline that is concerned with the development of algorithms allowing a machine to learn via inductive reference, and therefore much relevant to the field of multivariate analysis.

[11]Having only the individuals most fit to a given problem survive requires to have a perfect algorithm to determine that no part of the individual's genetic code could be beneficial to the problem. Since it is not the case, it is best to allow some less performing individuals to live, for they might produce strong offsprings by the appropriate mating.

network, e.g., by adding new (hidden) nodes or by creating new connections between them[12].

In order for these algorithms to be successful, a balance must be obtained between the fitness of the solutions to a given problem and their diversity. Also, the species must allow to evolve while preserving innovations, in order to efficiently develop complex topologies that are good candidate solutions from simple initial structures. As with every powerful tool, great care must be taken to ask the proper question, for a fitness function not suited to the problem could lead to a huge waste of computing time; worse, it can lead to a poor solution.

As expected, these techniques have much benefited from the advent of modern computing, and the access to powerful computers. In 2002, the NeuroEvolution of Augmenting Topologies (NEAT) [186] was developed. Unfortunately, we were unable to find an implementation suitable for our purpose.We mention, however, that the combination of the individual analysis contributing the the observation of electroweak single top production [161,167] (using a five input and one output node) was obtained using this technique.

9.11 Neural Network Implementation

We conclude this chapter by a description of the implementation used to train the networks used in this dissertation, namely that of the TMVA [184, 187] package. We employ version 4.1.0 of the package, part of v5.28 of the ROOT framework [97].

We require a supervised learning method, since the desired target output (t) is known for every input vector $\boldsymbol{x}$, and use the *back-propagation* algorithm (section 9.5.1) for adjusting the network weights and optimize the classification performance. We use the gradient descent algorithm (section 9.7.1) with a learning rate of 0.02 to update the weight vector at each step.

[12]However, in order to benefit from back-propagation (cf. section 9.5.1), it is best to maintain the feed-forward property of the network, and not create feedback loops, which require iterative solvers, and are time-consuming.

We use the following network topology, consisting in three layers of adaptive weights:

$$y_{NN} = \sum_{k=1}^{N} \tilde{z}_k \cdot \tilde{w}_k \tag{9.48}$$

$$= \sum_{k=1}^{N} \tanh\left(\sum_{j=1}^{N+7} \bar{z}_j \cdot \bar{w}_{jk}\right) \cdot \tilde{w}_k \tag{9.49}$$

$$= \sum_{k=1}^{N} \tanh\left(\sum_{j=1}^{N+7} \tanh\left(\sum_{i=1}^{N} x_i \cdot w_{ij}\right) \cdot \bar{w}_{jk}\right) \cdot \tilde{w}_k \tag{9.50}$$

where we used hyperbolic tangent activation functions.

We use 1000 training cycles (epochs), and check the similarity of the outputs using different initial values for the initial weight vector. We also monitor evolution of the error when obtained from an independent *test* sample compared to that of the error obtained from the *training* sample in order to avoid *over-training*. We do not however use ensemble networks, for they do not significantly alter the outcome of the classification, especially because our final goal is to apply (compare) the distribution of the *same* variable (the network output) in data and simulation (cf. section 9.8).

9.12 Summary

In this chapter, we have presented the key concepts behind artificial neural networks. We have shown that suitable topologies can approximate any continuous function by minimizing an error function. We discussed algorithm to perform such minimization, in particular the techniques of back-propagation and gradient descent, which are used in this dissertation. We also discussed the limitations of these techniques and Introduced advanced training methods.

With these tools in hand, we can now apprehend the complex task of deriving a complex parameterization of the trigger efficiency curve in a multidimensional space.

10. THE $\not\!\!E_T$ TRIGGERS AND THE TRIGGER EFFICIENCY

One of the main challenges at hadron colliders is to design an efficient data acquisition system. The Tevatron is no exception. The vast majority of collisions don't involve a large momentum transfer, and are mostly un-interesting in our search for rare standard model (SM) processes. Indeed, the cross-section for inelastic scattering is about 60 mb, while the top pair production has a cross-section which is of the order of 7 μb, which is ten orders of magnitude smaller. Moreover, it is impractical to store every single collision taking place at the center of the CDF detector: the CDF data acquisition system can operate at a maximum rate of 18 MB/s. Given an average size of 170 kB per event, this corresponds to a maximum rate of 100Hz. Since the collision rate at the Tevatron is at about 1.7 MHz, we are forced to reject 99.9% of the events.

Searches for the Higgs boson and physics beyond the SM are among the main goals of the Tevatron physics program. These processes have small production cross-sections σ. The number of events produced for a given reaction is given by $N = \sigma\mathcal{L}$. A large amount of data are therefore required to search for these processes. The instantaneous luminosity $\mathcal{L}$ in the Tevatron accelerator was increased several times – and is now routinely running well above the design goal (see figure 4.4) – to maximize the production rate of these processes. Similarly, the trigger rate R grows as follows with the luminosity:

$$R = \alpha + \beta\mathcal{L} + \gamma\mathcal{L}^2 + \delta\mathcal{L}^3, \tag{10.1}$$

where the higher orders in luminosity are caused by overlapping physical objects from multiple interactions (which happen in high luminosity regimes), and by luminosity-dependent fake objects. For high purity triggers, such as those collecting events with

high-p_T muons, usually have $\gamma \sim \delta \sim 0$. $\not\!\!E_T$ triggers, on the contrary have large high-order terms, due to the large fake (instrumental) $\not\!\!E_T$ in high occupancy events. Additionally, low-energy jets produced in events with multiple interactions could be merged and be reconstructed as a single, high-E_T jet, which is then accepted by a trigger requiring the presence of high-p_T jets.

The CDF trigger system, designed to maintain high efficiency for interesting physical signatures while rejecting the majority of un-interesting events, is composed of three levels, which are described in section 4.3.6. This multi-level architecture is chosen to select events using increasingly more complete information at each level, thereby reducing the dead-time (as the bulk of obviously un-interesting events is quickly rejected, without requiring the additional computation of event properties).

Triggers are always designed with specific physics goals in mind. At CDF, as at other discovery machines, the goal was to accommodate a broad physics program, while maintaining high efficiency. We now introduce several triggers designed for searches in signatures with a large energy imbalance, and high-p_T jets ($\not\!\!E_T$+jets).

As described in section 4.3.6, the CDF trigger system selects events at a rate of about 100 Hz[1]. This is implemented using a three-level trigger system, with decreasing rates of 25 kHz (L1), 350 Hz (L2) and 100 Hz.

10.1 The $\not\!\!E_T$ Triggers

During Run II, the $\not\!\!E_T$ triggers have evolved with time, following changes in luminosity conditions, and in detector upgrades. Four main trigger paths are used at CDF to collect events with large $\not\!\!E_T$: MET+BJET, MET45, MET35+JETS, and MET+DIJET;

MET+BJET This trigger collects events with $\not\!\!E_T > 20$ GeV/c^2 and two displaced tracks. It is conceptually the best trigger, since it provides a high purity sample. During operations, however, it reached a very high rate, which was too high in

[1]The CDF data acquisition system can operate at a maximum rate of 18 MB/s. Given an average size of 170 kB per event, this corresponds to a maximum rate of 100Hz.

fact to cope with the increasing instantaneous luminosity of the Tevatron. As a consequence, it is pre-scaled[2] since 2004. Therefore, it not very efficient and lacks sensitivity.

MET45 This is the inclusive $\not\!\!E_T$ trigger. It is very simple to understand, since it requires no additional objects. Although the trigger was originally designed to select events with $\not\!\!E_T > 45$ GeV/c^2 at L3, this requirement was softened to 40 GeV/c^2 at the end of 2007. Unfortunately, the $\not\!\!E_T$ trigger threshold is too high to be efficient for analyses dealing with standard model processes but it can be used as an auxiliary trigger, increasing the acceptance. This trigger path is mostly used for searches for physics beyond the standard model (BSM), such as supersymmetry (SUSY) or dark matter (DM), where large $\not\!\!E_T$ is expected to be produced.

A bug in the trigger code temporarily set the cut on $\not\!\!E_T$ to a larger value (~ 120). The effect on the overall dataset is negligible, and is accounted for by the parameterization of the trigger effiency (section 10.3).

MET35+JETS This trigger path was designed explicitly for looking at final states with $\not\!\!E_T$ and at least two high-p_T jets. Because of the two-jet requirement at Level 2, it allows selecting events with $\not\!\!E_T > 35$ GeV/c^2, thus relaxing the requirements of the MET45 path.

There are actually several implementations of this trigger. The first implementation requires events to have $\not\!\!E_T > 35$ GeV/c^2 at L3, and two clusters with $E_T > 10$ GeV/c^2 at L2. It first started collecting events on July, 22 2002. On March, 27 2005, an additional requirement was added: that one of the clusters be central, i.e. with $|\eta| < 1.1$. Due to the difficulty to operate in a high luminosity region, it was decided that, starting from September 2006, this trigger was only to be enabled when the instantaneous luminosity was below

[2]A pre-scaled trigger operating with a rate P_R records only $1/P_R$ events.

190×10^{30} cm^{-2}s^{-1}, hence not recording many precious collisions[3]. Because the performance improvement on the accelerator side, it was decided to operate the trigger with a dynamic prescale[4] (DPS) as of April 2007, to allow recording collisions when in the high luminosity part of the collider store. In this dissertation, this trigger path is only used for the first 2.2 fb^{-1} of data, although it continued to collected data after that; it was replaced by the next path for the data collected after December 2007.

MET+DIJET The MET+DIJET path, was introduced to overcome the problems of the MET35+JETS at high luminosity. It was adopted in Run IIb to collect events with $\not\!\!E_T$ and at least two jets [188]. The main idea behind this trigger is to apply a tight cut on the E_T of the leading jet by requiring the presence of a trigger tower with high energy at L1 (while allowing for looser requirements at L2).

This path has performed well since its implementation, and is the best trigger for the topology we are interested in. Its main limitation is that it was not present before the L2-CAL upgrade (cf. section 4.3.6). Therefore, it needs to be used in combination with the MET+JETS path. This causes some difficulty when estimating the efficiency of this trigger path; but these can be overcome, as will be detailed in section 10.3.

The events analyzed in this dissertation were collected by a logical combination (OR operation) of three triggers: MET45, MET35+JETS, and MET+DIJET, the latter two contributing the most. We refer to this combination as the *All MET* trigger[5]. A more detailed account on the $\not\!\!E_T$ trigger paths, together with those used for trigger studies, in given in appendix A.

[3]Since September 2007, the peak instantaneous luminosity has been most of the time well above 200×10^{30} cm^{-2}s^{-1}, as shown in figure 4.4.

[4]A dynamic prescale consists of a feedback loop that adjusts the prescale dynamically according to the total trigger rate.

[5]Although there are many more triggers that could be used to collect $\not\!\!E_T$ events, we decide to not include them in order to keep things under control, especially since these additional triggers do not contribute much to the combined path.

All the trigger paths presented above suffer from a hardware problem that creates a small inefficiency in events for which the L1 $\not\!\!E_T$ is mis-calculated because of a saturation in the energy of the towers. This inefficiency, however, has been recovered using alternative trigger paths (cf. section 10.2).

The combination of multiple triggers is the main motivation for the investigation of a multi-path efficiency, leading to the use of a NN approach that handles any correlation in the inefficiencies between the multiple paths.

10.2 Event Reconstruction in the Trigger System

We now proceed with a description of the decision-taking process at the end of which an event is selected or rejected by the trigger.

Level 1 The L1 trigger system, based on custom electronics for fast decision taking, reconstructs primitives of basic physical objects, which can be triggered on. The calorimeter towers (section 4.3.2) are mapped to a 24×24 array of trigger towers in $\eta - \phi$[6]. Several steps are taken to simplify and speed up the process:: (a) the calorimeter energy information is in an eight bit representation[7]; (b) the event primary vertex is assumed to be located at $z = 0$ when computing transverse projections, and (c) the $\not\!\!E_T$(L1) is obtained from a vector sum of trigger tower pairs. Due to the limited information available and the time constraint, the resulting $\not\!\!E_T$(L1) has poor resolution, and it is usually underestimated. For this reason, the threshold at L1 is chosen as low as possible.

Inefficiency at L1

The eight bit representation used for L1 processing result in a maximum scale of measured energy of $E_{\mathrm{MAX}} = 127$ GeV. Therefore, if the energy deposited in a single tower exceeds this value, only E_{MAX} is used in the computation. These events,

[6]Each (software) trigger tower spans approximately 15° in ϕ and 0.2 in η covering one or two physical towers.
[7]The eight bits are obtained from the 10-bit value by dropping the most and least significant bits.

however, should have been automatically accepted (at L1 only) in order to avoid the loss in trigger efficiency. However, due a hardware problem at L1 [148], this was not implemented until later. Fortunately we can recover these events by collecting them from the high-E_T jet sample (JET-100), as described in ref. [189], and add them to the $\not\!\!E_T$+jets sample. The JET-100 sample is a un-prescaled sample triggering on the presence of a single jet above 100 GeV. Therefore, it automatically contains all events with single towers above $E_{\rm MAX}$; we can thus check whether these should have been collected by the $\not\!\!E_T$ triggers as well.

Level 2 The L2 trigger system introduces a level of abstraction, reconstructing jets as *clusters* of energy deposits. The Run I algorithm used for this purpose seeks a *seed* tower (with an anergy above 3 GeV) and then adds adjacent *shoulder* towers (with an energy above 1 GeV) until there are no more shoulder towers. The cluster is assigned the position of the seed tower.

As the Tevatron luminosity increased, multiple interactions increased the detector occupancy, leading to the presence of many energetic shoulder towers around a seed. These energetic towers could be erroneously clustered into a single jet, or worse, could cause the erroneous merging of separate clusters into a single one, resulting in a significant efficiency loss for triggers looking for events with at least two clusters. To address this issue, the L2 calorimeter trigger was upgraded (section 4.3.6).

In this dissertation, we analyze data from both the old and the new trigger system. Combining these datasets into a single analysis has some practical consequences, which we will describe later. For instance, one must reject events collected prior to the upgrade with close-by jets in the $\eta - \phi$ plane, in order to maintain an efficient trigger.

Level 3 The L3 trigger system possesses the full information for each event, including the 10-bit representation of the energy in the trigger towers. Therefore, the reconstruction of physical objects is much improved, and so is the $\not\!\!E_T$ resolution.

10.3 Parameterization of the Trigger Efficiency

The purpose of the detector is to record collisions as they happen, at the highest possible rate. As we have seen, the trigger system is responsible for collecting (and requesting storage of) interesting events physics-wise; one algorithm per trigger path decides which events to keep. In order for our simulations to be at par with reality, we ought to include information about these algorithm. For this reason, we wish to know what is the probability for a given simulated event to be accepted by the trigger path under study. We aim to compute the efficiency for this trigger path, i.e the ratio of events that did fire the trigger to those that didn't; for this purpose, we collect data using other, orthogonal, trigger paths. We then seek to obtain a parameterization of this efficiency as a function of kinematic variables that are correlated to this path's decision.

We now proceed to the description of the parameterization of the trigger efficiency. Because this process is fairly complex, we start by describing the simple parameterization of the MET35+JETS trigger path, which was the path used in the $\not\!\!E_T$+b-jets searches using 2.1fb^{-1} of data [114, 161, 190]. We then proceed with the description of a more complex parameterization, which combines two paths that do not overlap in time (MET35+JETS and MET+DIJET). Given the tediousness of this parameterization task, and the inability to easily add additional triggers (which could overlap in time), we resort to a multivariate parameterization, and carefully validate this method.

10.4 Simple Parameterization of the Trigger Efficiency

The MET35+JETS trigger path has the following requirements, detailed for each trigger level: (a) $\not\!\!E_T > 25$ GeV at L1; (b) two clusters with $E_T > 10$ GeV at L2, and (c) $\not\!\!E_T > 35$ GeV at L3. This trigger has been changed along time in order to accommodate higher instantaneous luminosity regime while keeping the bandwidth low (cf. section 10.1).

The trigger efficiency for each of the three levels of the CDF trigger is calculated using data collected with a high p_T muon, and with inclusive jet samples, and then parametrized as a function of offline quantities. These were initially obtained using 1.0 fb^{-1} [148] and updated using 2.1 fb^{-1} of data [114, 191]. The trigger parameterization is[8]:

$$\epsilon_{L1}(\not\!\!E_T) = \frac{1}{1 + e^{\frac{31.2 - \not\!\!E_T}{6.6}}} \tag{10.2}$$

$$\epsilon_{L3}(\not\!\!E_T) = \frac{1}{1 + e^{\frac{37.0 - \not\!\!E_T}{6.6}}} \tag{10.3}$$

and the uncertainties (on the whole trigger path) are provided by:

$$\Delta\epsilon/\epsilon(\not\!\!E_T) = \begin{cases} 0.01 \times \frac{(80 - \not\!\!E_T)}{10} + 0.25 \times \left(\frac{80 - \not\!\!E_T}{50}\right)^4 & \text{if } \not\!\!E_T < 80 \text{ GeV}, \\ 0 & \text{if } \not\!\!E_T \geqslant 80 \text{ GeV}. \end{cases} \tag{10.4}$$

Figures 10.1 and 10.2 show the L1 MET25 and L3 MET35 trigger efficiencies in 2.1 fb^{-1} of data.

10.5 Fit-based Parameterization of the Trigger Efficiency

The complex trigger structure created by the logical combination of three triggers with different assumptions makes it hard to obtain an acceptable parameterization of the trigger-turn in terms of one or two variables, as is often the case for a single trigger path depending on $\not\!\!E_T$ (as shown in the previous section) and eventually the jet energies.

It is possible to obtain an approximate solution by splitting the problem in multiple pieces and making appropriate assumptions. This is the approach used in

[8]The $\not\!\!E_T$ referred to here is actually the corrected $\not\!\!E_T$, as defined in section 5.7.1.

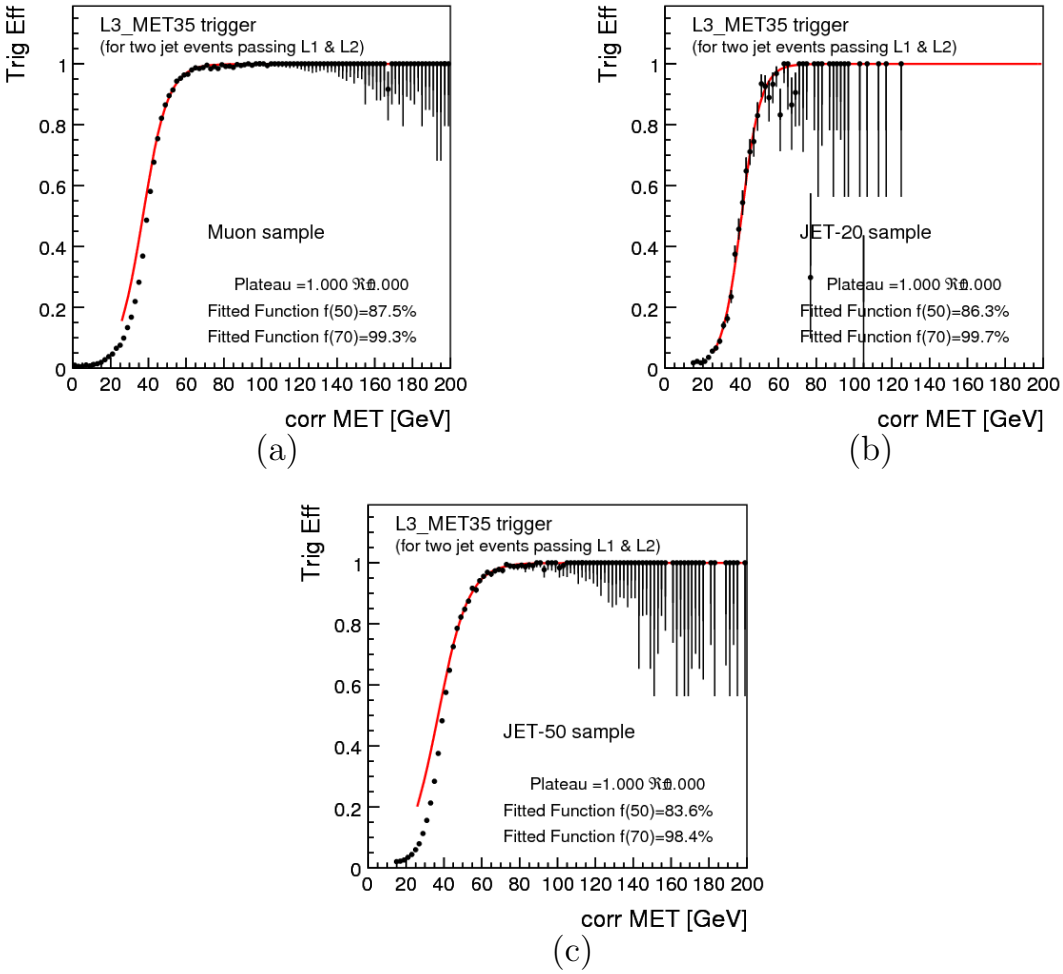


Figure 10.1.: Efficiency of the L3 MET35 trigger in 2.1 fb^{-1}of data for (a) the muon sample, (b) the JET20 sample, and (c) the JET50 sample. The parameterization of the trigger efficiency is shown in red, as a function of the corrected $\not\!\!E_T$ (see section 5.7.1).

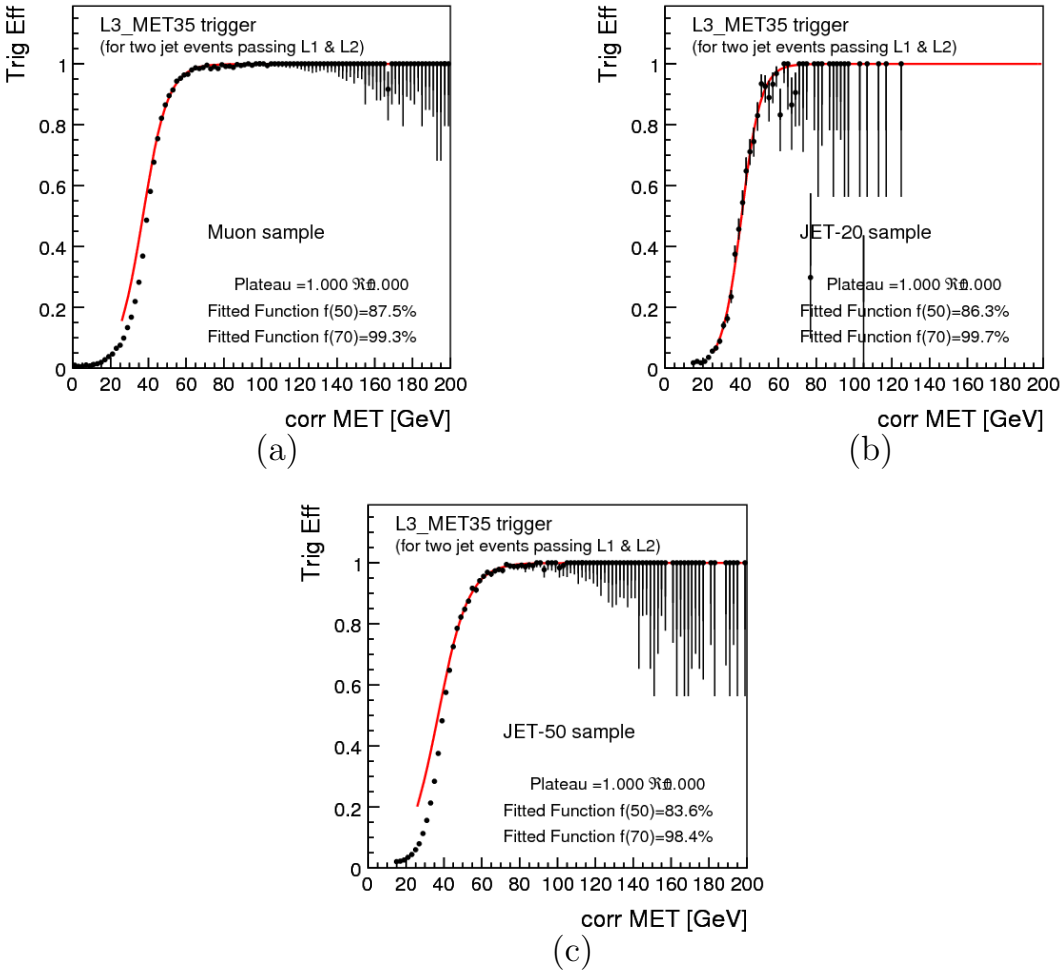


Figure 10.2.: Efficiency of the L3 MET35 trigger in 2.1 fb^{-1}of data for (a) the muon sample, (b) the JET20 sample, and (c) the JET50 sample. The parameterization of the trigger efficiency is shown in red, as a function of the corrected $\not\!\!E_T$ (see section 5.7.1).

ref. [192] to derive a simple parameterization of the combination of the MET35+JETS and the MET+DIJET trigger paths (which are orthogonal in time), referred to a MET+CJET+JET. In this approach, the parameterization is obtained by factorizing requirements in the trigger and assuming that the single-variable parameterizations obtained in each region are independent. This assumption is a good approximation when the analysis selection guarantees a high efficiency, especially since a conservative estimation of the systematic uncertainty covers possible issues related to this assumption. The parameterization uses the following variables: (a) the $\not\!\!E_T$; (b) the E_T, η, and ϕ of the jets; (c) the distance between the jets in the $\eta - \phi$ plane;, and (d) the angular separation between the $\not\!\!E_T$ and the jets in the transverse plane. Although the trigger depends explicitly only on the $\not\!\!E_T$ and the E_T of the jets, parameterizing according to other variables is important. For instance, it is useful to know whether the jet is aligned to the $\not\!\!E_T$ or not, hence the inclusion of $\Delta\phi(\not\!\!E_T, j_i)$. Besides, η dependence can arise through correlations: for example, the b-tagging requirements (cf. chapter 6) tend to a more central jet population than in the inclusive samples used to derive the parameterization. For this reason, the pseudo-rapidity of the jets is included in the parameterization.

The systematic uncertainty was estimated from the comparison of the measured efficiency obtain parameterization in two different samples covering different topologies of the final state: a high-p_T muon sample, which provides a good estimation for events with *true* $\not\!\!E_T$, and a medium-E_T jet sample, which contains events in which $\not\!\!E_T$ is produced by a mis-measurement of the jet energies; the latter also accounts for the dependence on the jet energy scale.

This method performs very well in its domain of validity, i.e when applied to events with a certain amount of $\not\!\!E_T$ and jet energies above a certain threshold, in order to insure a high-efficiency trigger operation. This, however, severely limits the possibility of relaxing the kinematic cuts and gain extra acceptance to rare standard model processes. The method is also labor intensive since we must obtain the appropriate parameterization in each individual region, as well as to keep track of each separate

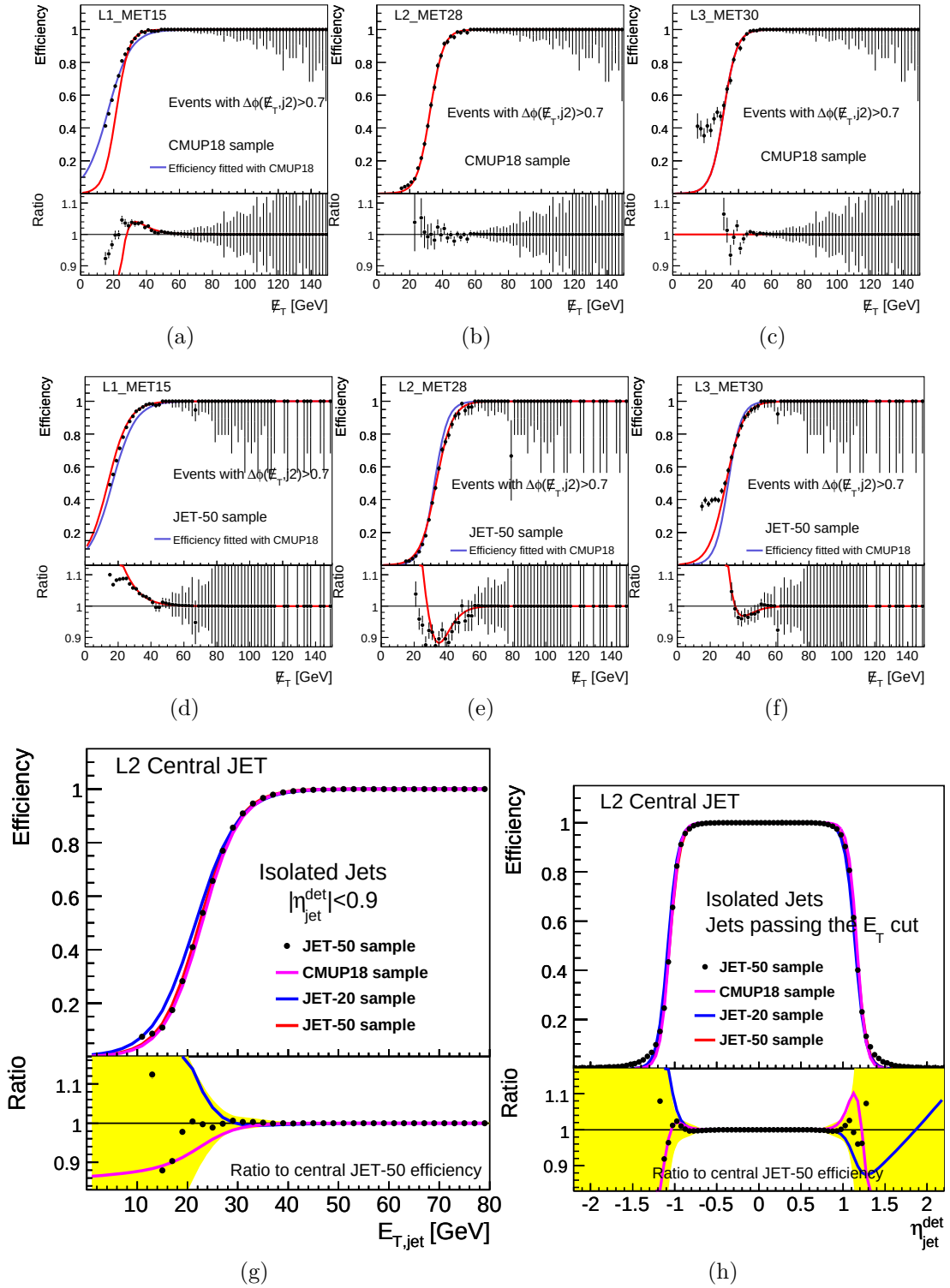


Figure 10.3.: Efficiency of the fit-based parameterization as a function of (a-f) $\not\!\!E_T$, (g) E_T^j, and (h) $\eta^{\mathrm{det},j}$. These individual parameterizations for each level are combined to provide the global efficiency.

region. Finally, adding additional triggers is very tedious, especially when two (or more) of them have strongly correlated efficiencies.

As this method was not retained as the final version for this dissertation, we do not discuss it further, and proceed to the description of a better approach.

10.6 Multivariate Parameterization of the Trigger Efficiency

The solution to the constraints (and concerns) outlined above is to perform a parameterization by training a multivariate tool, in this case an artificial neural network (NN), providing a multi-dimensional regression function (section 9.2). This allows dealing with the problem in a global and straightforward way, since the NN can in principle handle as many variables as needed. In addition, no prior assumptions have to be made regarding the domain of a given trigger (i.e. the region from which events could possibly fire a given trigger), which facilitates dealing with multiple triggers. This also means that no study is needed in order to determine the correlations between triggers, or their overlap (in event space or time). On top of all this, this approach greatly simplifies the process of parameterizing the logical combination of multiple trigger paths in terms of human intervention (which is prone to introducing mistakes): all what is required is to train a few networks[9].

Using a NN technique introduces aspects that have to be properly understood. For instance, it is not clear how to estimate the uncertainties introduced by the chosen tools. Furthermore, testing the performance of the parameterization is not straightforward, since it involves using information from a given kinematical region and propagate it to another. We must determine if a network is able to solve the problem of parameterizing a complex multidimensional function by accounting all correlations between the input variables, the dependence of this function on these variables, as well as the correlations between the different components of the output

[9]There a still other considerations such as statistics and dimensionality that require using different networks for different purposes.

(i.e. the correlation between the individual triggers being logically combined). In principle, it should be able to; we present evidence that it is the case.

Even if we derive the optimal parameterization, it is not clear whether it is applicable to the samples that do not satisfy the requirements of the training set; e.g. can we apply a parameterization derived from events with $W \to \mu\nu$ to $t\bar{t}$ or Higgs events? In other words, we ask ourselves whether the network is able to hint at the *underlying generator* (c.f. section 9.5), for in this case the network would be applicable to any sample. The task of determining this, however, is complicated by the fact that is is very difficult, if not impossible, to check whether the trained network is introducing a bias coming from the training sample (and thus failed at determining the underlying generator) when applied to the region of interest. We must investigate the link between the topology of the event and the trigger efficiency we seek to model.

We alleviate the concerns outlined above by performing systematic studies based on two independent samples, as was done for the parameterizations described in the previous section. The outcome of these studies is to provide an appropriate estimation of the systematic uncertainty related to the parameterization tool used.

10.6.1 Sample Description and Event Selection

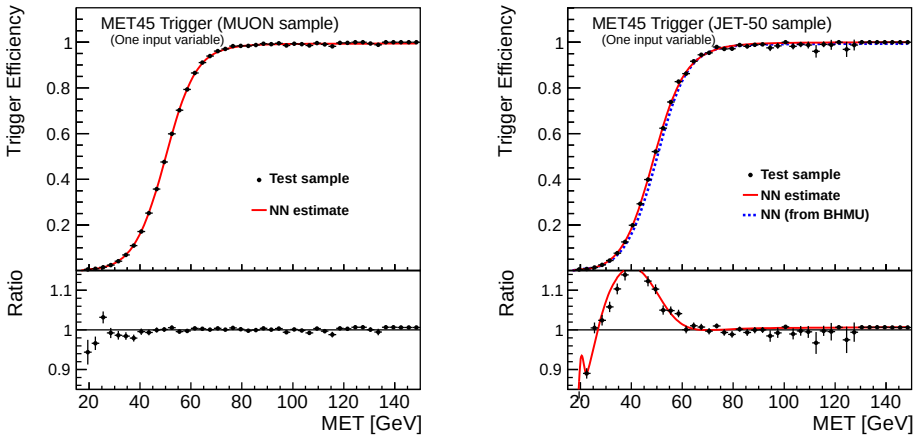
As mentioned earlier, we use two datasets: a high-p_T muon sample (MUON), and a jet sample with at least one jet with an energy above 50 GeV (JET-50). In each dataset, we make use of all the events, irrespective of the trigger that was used to collect them. For the purpose of determining the trigger parameterization for events with real $\not\!\!E_T$ produced by particles escaping the calorimeter, the MUON sample gives the best representation, and will be used to provide the central value of the parameterization. In addition, the trigger requirements we model are based on calorimeter objects, and therefore are orthogonal to those of the muon trigger, which are based on tracking and on the muon chambers; this is a clear advantage over the JET-50 sample.

We require the events to meet the clean-up cuts presented in section 8.3 [149]. Jets are reconstructed by the JetClu algorithm (*c.f.* section 5.6.1) and corrected up to L5 before applying the H1 prescription (*c.f.* section 5.6.2). Unless otherwise mentioned, we select events with at least two jets (with $E_T \geq 20$ GeV and $|\eta| \leq 2.4$) separated by $\Delta R \geq 0.8$. Events with $\not\!\!E_T \leq 20$ GeV are rejected. This selection criteria are mostly the same as the analysis in which the parameterization is to be applied, except for the cut on the $\not\!\!E_T$ (for the obvious reason of providing a smooth parameterization on the low-$\not\!\!E_T$ region of the trigger turn-on).

For the purpose of the training, we take a sub-sample of the events selected by the above cuts in order to keep the training time reasonable. During this selection, a filter was applied based on the value of the $\not\!\!E_T$ in order not to deplete the tails of the distribution. No biases are observed due to this enhancement of the main kinematic variable to be reproduced.

10.6.2 The Simple Case of MET45

We start by checking the network-driven approach using a simple case: that of the MET45 trigger path, which simply requires a cut on $\not\!\!E_T$ at the various levels of the trigger system. Figure 10.4 shows the outcome of a training using a simple network together with its application on an independent testing sample. The network has one input node ($\not\!\!E_T$), two hidden layers (the first with three nodes, the other with two), and one output node, i.e. a 1:3:2:1 configuration. For the jet-50 sample, we show (in addition to the function derived in that sample) the reference function obtained from the muon sample, which is also taken as the reference for the ratio plot. The parameterization of the trigger is well described by the network, as expected since the functional form resembles to a sigmoid. The network derived from the muon an jet-50 samples agree well at high $\not\!\!E_T$ an disagree at low $\not\!\!E_T$, as expected. Indeed, because $\not\!\!E_T$ is a complicated variable, it is expected for the efficiency to implicitly depend on other variables, albeit in limited way. This effect can clearly be seen in fig-



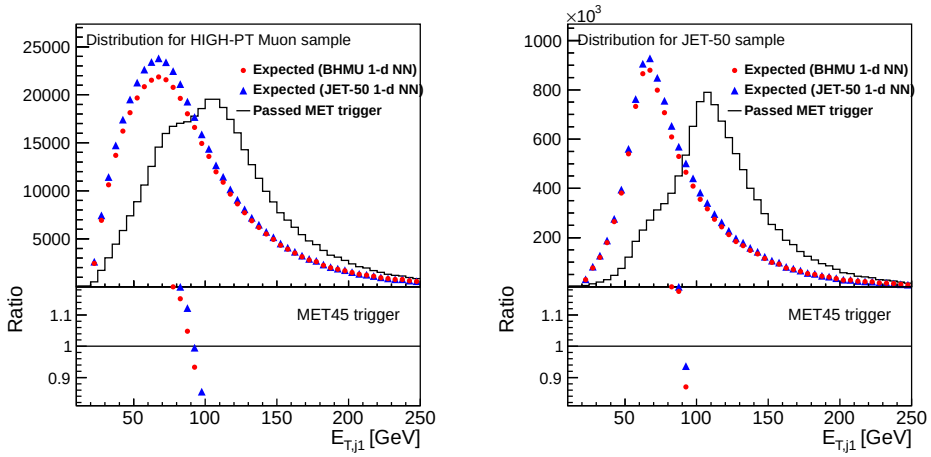
(a) (b)

Figure 10.4.: Distributions of the trigger efficiency as a function of the $\not\!\!E_T$ for events passing the MET45 trigger in the (a) MUON and (b) the JET-50 samples.The NN predictions are compared to the data (and the central, MUON, parameterization in the case of the JET-50 sample). The ratio is taken with respect to the central parameterization.

ure 10.5, where the distribution of the energy of the leading jet is not well reproduced by the parameterizations[10], while that of the $\not\!\!E_T$ is in better agreement (figure 10.6). It is worth noting that the agreement of the MUON and JET-50 parameterizations in the samples where they are derived is simply a consistency check, and that the discrepancy between both parameterizations in a given sample is only present in the region of low trigger efficiency. In the region of high efficiency, however, there is excellent agreement, and we obtain a low systematic uncertainty.

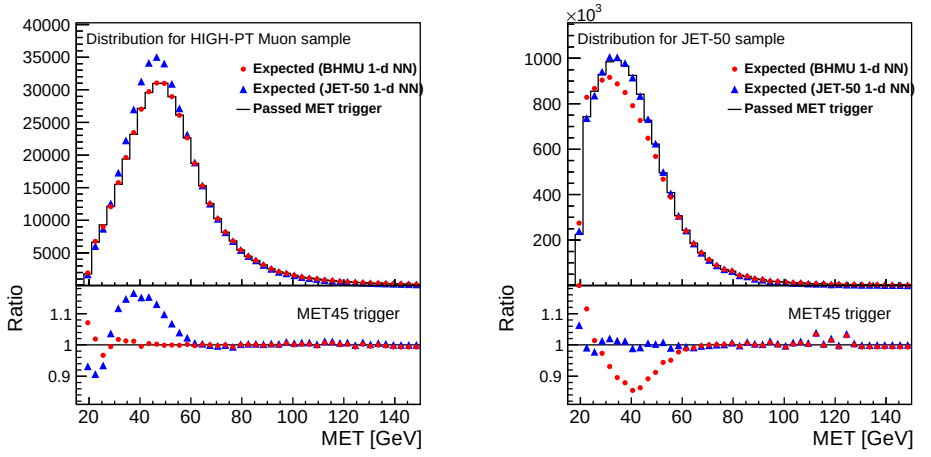
Given these considerations, we investigate the performance of a network with two input variables, $\not\!\!E_T$ and $E_T^{j_1}$, and a 2:4:3:1 configuration. Figure 10.7 shows the output of this network. The main advantage of this network is not in its modeling of the trigger efficiency as a function of the $\not\!\!E_T$, but in the fact that it is possible to view the dependence on the variable that is not plotted. Scatter plots prove to be very useful for this since they tell us that there are actually two turn-on functions (corresponding to

[10]It should be noted that the overall normalization and therefore the average efficiency is correct.



(a) (b)

Figure 10.5.: Distributions of the $\not\!\!E_T$ for events passing the MET45 trigger in the (a) MUON and (b) the JET-50 samples. Also shown are the predictions of the two NN parameterizations. The ratio is taken with respect to the data. There is a clear disagreement between the observed and expected distributions.



(a) (b)

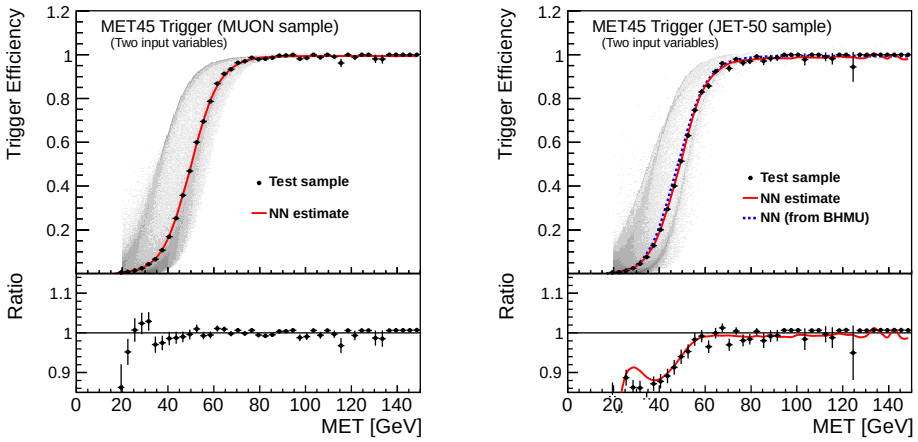
Figure 10.6.: Distributions of $E_T^{j_1}$ for events passing the MET45 trigger in the (a) MUON and (b) the JET-50 samples. Also shown are the predictions of the two NN parameterizations. The ratio is taken with respect to the data. Although the self-check yields excellent agreement with data, the cross-sample application of the parameterizations is not reproducing the distributions very well in the regions of low trigger efficiency.

the two "versions" of the MET45 trigger path described earlier). The average of these paths is closer to the high-$\not\!\!E_T$ curve, since the events with $\not\!\!E_T > 45$ GeV are accepted by both versions of the MET45 path. We confirm this interpretation by applying the MUON parameterization only to events with $E_T^{j_1} > 120$ GeV (figure 10.8), and confirming that the $\not\!\!E_T$ cut is lower.

When looking at the efficiency as a function of the second variable of the network, namely $E_T^{j_1}$, we observe a good reproduction of the average behavior the sample (cf. figure 10.9). However, the cross-sample agreement is poor, due to a known difference of behavior at low values of the $\not\!\!E_T$, confirmed by requiring $\not\!\!E_T > 50$ GeV (cf. figures 10.9c and 10.9d).

One weakness of the multivariate approach is the difficulty to compare cross-sample results, since the uncertainties in the low trigger efficiency regions are large and these same regions some times contain a large portion of the events in the samples used to derive the parameterizations (and hence have a strong role in its determination). We remind the reader that this difference in the low-efficiency region is intrinsic to the difference between the MUON and JET-50 samples, and cannot be overcome by a change in the regression technique used. Therefore, it is important to apply cuts guaranteeing a minimum efficiency, and to estimate the uncertainty by comparing the efficiencies from two different samples, as done here.

Figures 10.10 and 10.11 show the distribution of several variables for events passing the MET45 trigger compared to the two-input NN parameterizations. Besides the good agreement in the variables used for the parameterization (cf. figure 10.10), there is good agreement in the energy of the second jet $E_T^{j_2}$, a variable that is not part of the training process, suggesting that the trigger does not depend much on this variable (cf. figures 10.11a and 10.11b). The distribution of the pseudo-rapidity of the leading jet η^{j_1}, however, is not well reproduced, not only suggesting that the trigger depends on it, but also that the derivation of the parameterization should include this information. Note that this depends on the final selection cuts of the analysis, as

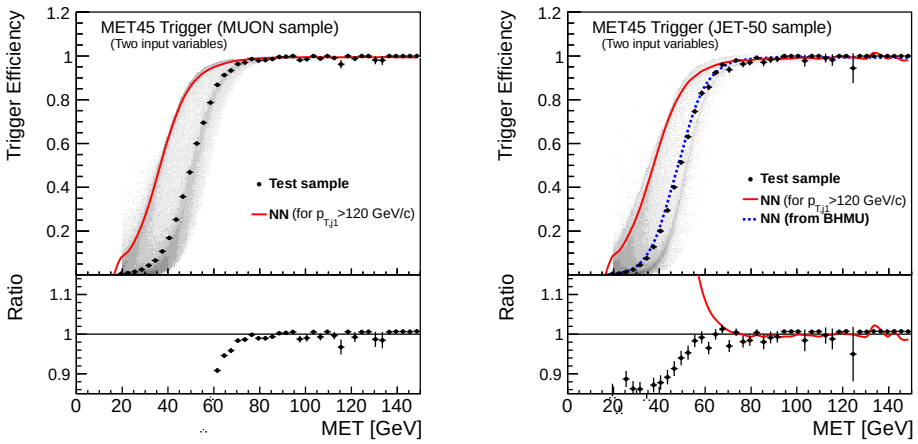


(a) (b)

Figure 10.7.: Trigger efficiency of the MET45 path as a function of the corrected $\not\!\!E_T$ as measured and parameterized in the (a) MUON and (b) JET-50 samples using a two-input NN. The scatter plot probes the dependence on $E_T^{j_1}$. The ratio is taken with respect to the MUON parameterization.

for instance the agreement in the central region $|\eta < 0.9|$ is better (although clearly not satisfactory).

Since our goal is to model a logical combination of $\not\!\!E_T$ triggers, we will not discuss the case of a single trigger further. The considerations presented in this section motivate the use of multi-dimensional parameterizations, and also confirm the applicability of the technique. However, these techniques require us to be as much as possible aware of the subtleties at stake, since we are attempting to model a region outside of the fully-efficient trigger operation. In particular, this technique also requires us setting selection criteria that guarantee a reasonable efficiency.



(a) (b)

Figure 10.8.: Trigger efficiency of the MET45 path as a function of the corrected $\not\!\!E_T$ as measured and parameterized in the (a) MUON and (b) JET-50 samples using a two-input NN. The scatter plot probes the dependence on E_T^{j1}. In each sample, the corresponding parameterization (red line) is applied to events with $\not\!\!E_T > 120$ GeV, confirming the interpretation of two turn-on functions (see text). The ratio is taken with respect to the MUON parameterization (applied to the whole $\not\!\!E_T$ range).

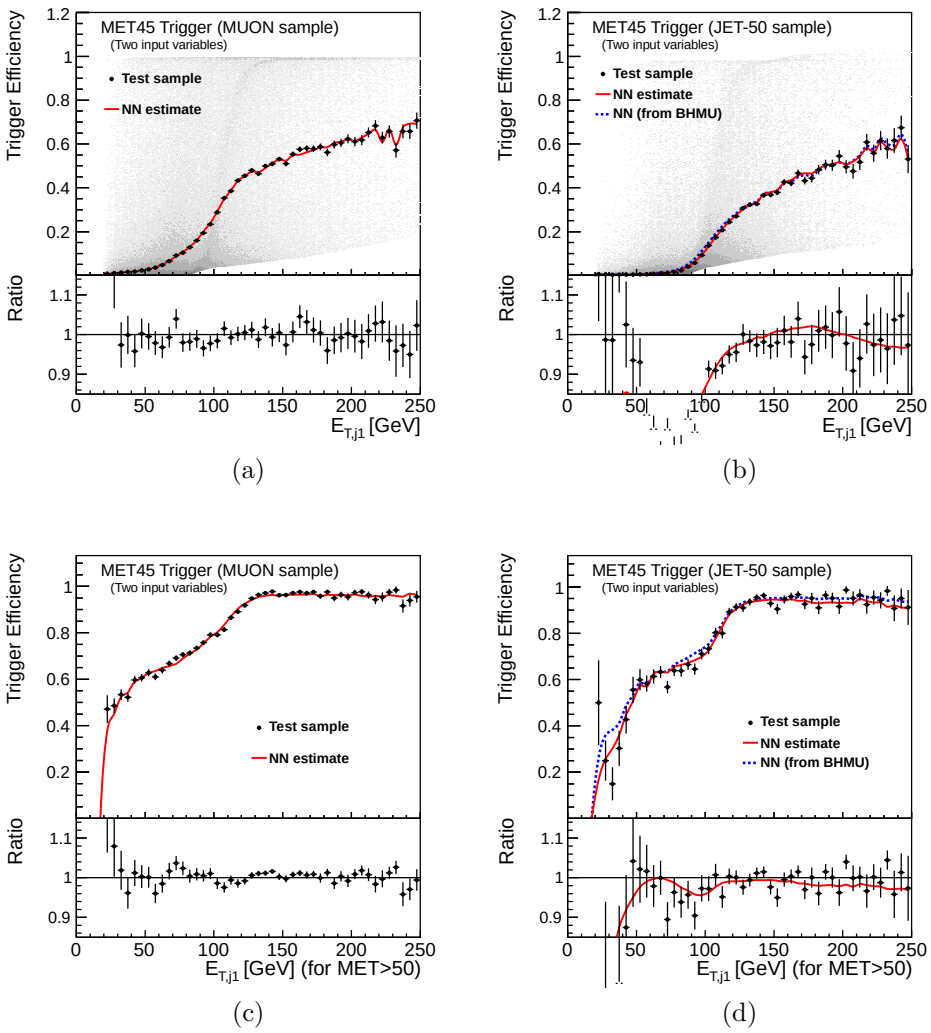


Figure 10.9.: Efficiency of the MET45 trigger path as a function of the E_T of the leading jet parameterized in the (a,b) MUON and (b,d) JET-50 samples. The parameterizations are provided by NNs with two input nodes. The agreement between the two parameterizations is much improved in events with (cd) $\not\!\!E_T > 50$ GeV, for which the trigger efficiency is larger. In (a,b), the scatter plot probes the dependence on E_T^{j1}.

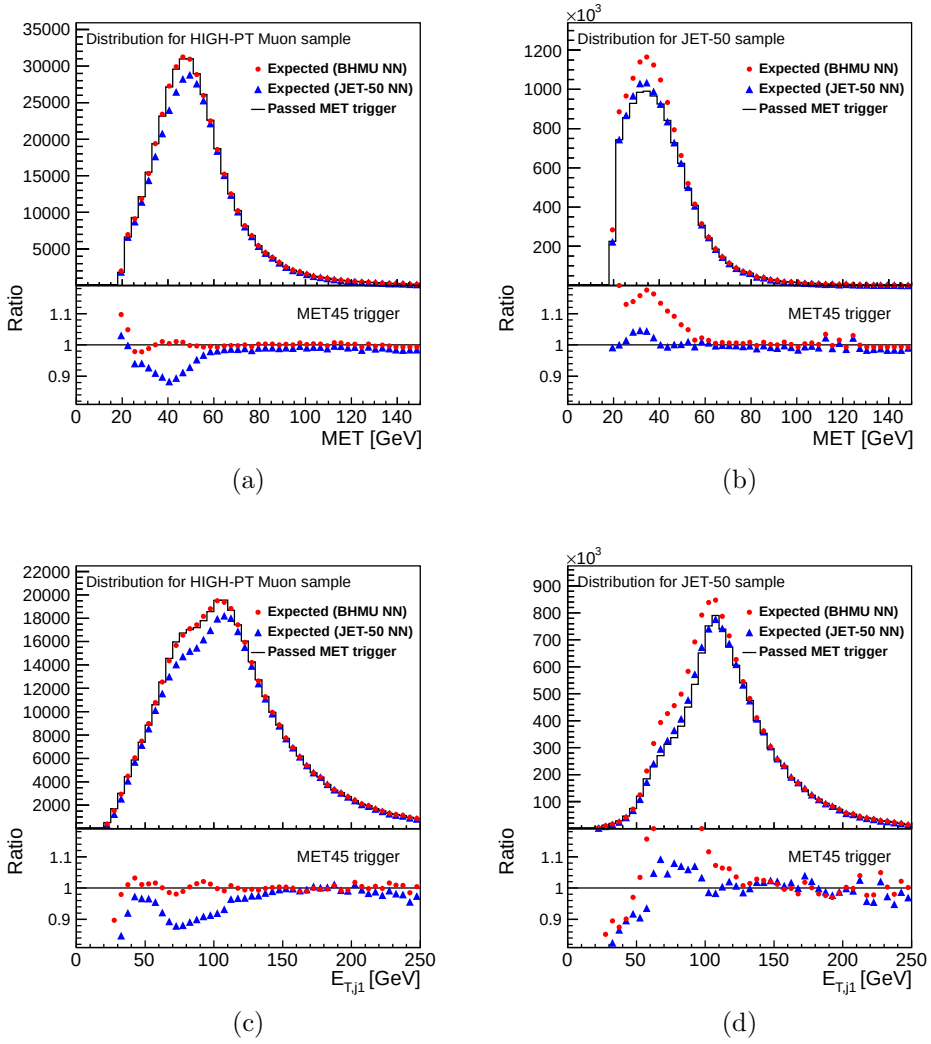


Figure 10.10.: Distribution of several kinematic variables for events passing the MET45 path in the (a,c) MUON and (b,d) JET-50 samples. The expected distributions using the two NN parameterizations are also plotted.

10.6.3 The Validation of Multivariate Tools

In order to demonstrate that the NN functions described above behave as expected, we checked that the input variables are correctly reproduced on average; the output of this check on $\not\!\!E_T$ was similar for the two different networks. We checked that the energy of the leading jet and other variables that were not used for the

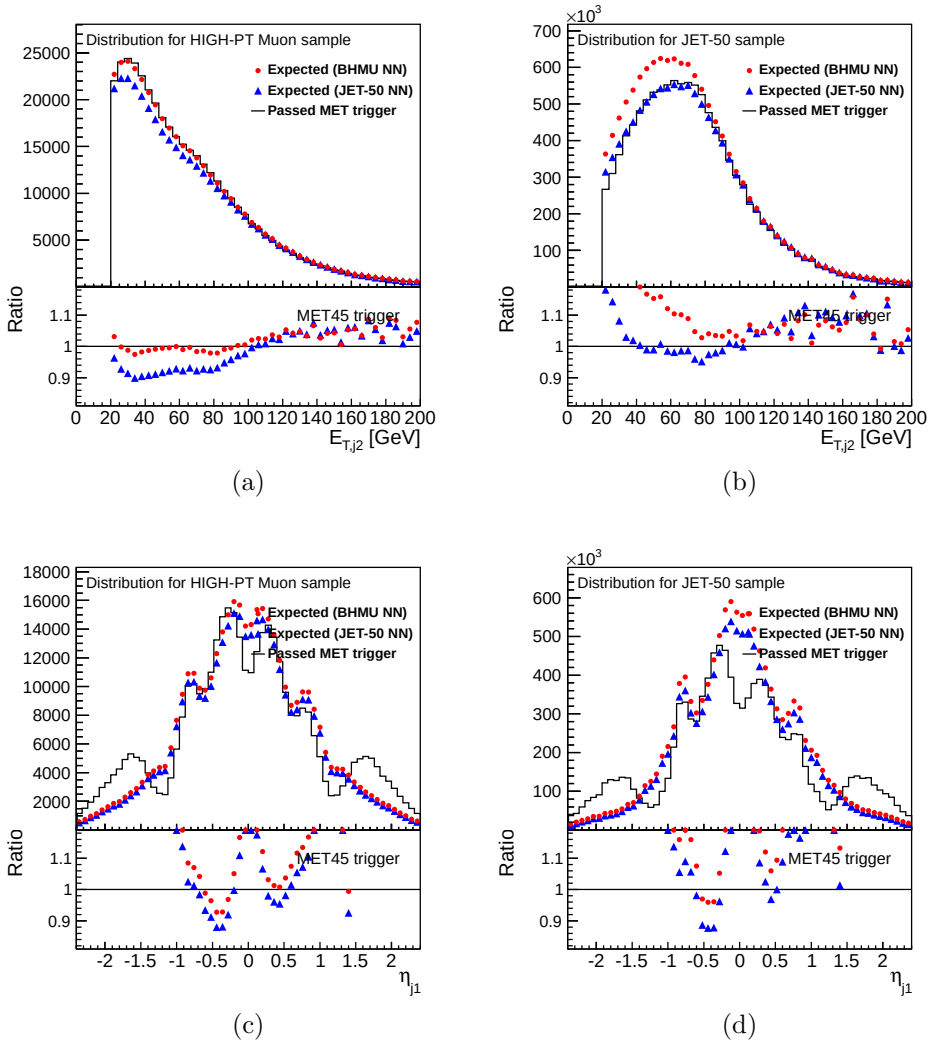
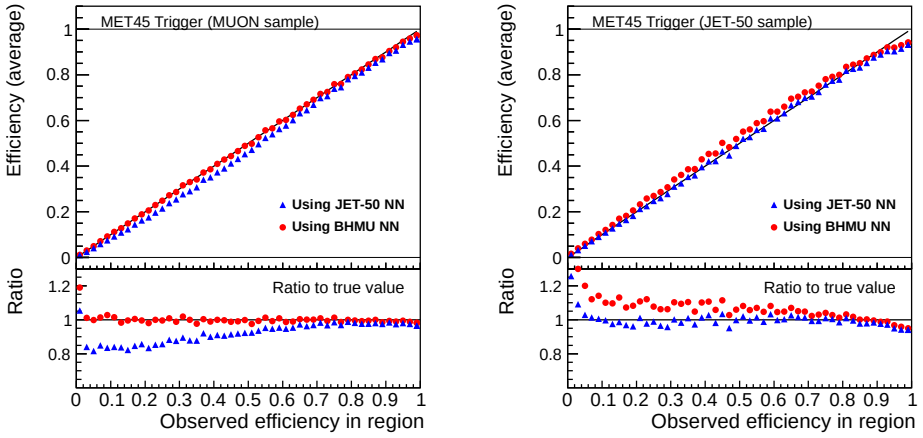
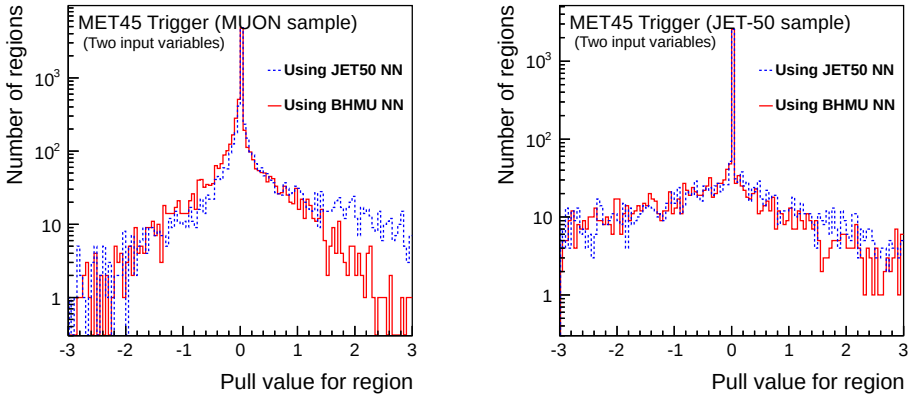


Figure 10.11.: Distribution of several kinematic variables for events passing the MET45 path in the (a,c) MUON and (b,d) JET-50 samples. The expected distributions using the two NN parameterizations are also plotted.



(a) (b)

Figure 10.12.: Expected efficiency from the two parameterizations as a function of the measured efficiency in discretized regions for the (a) MUON and (b) JET-50 samples.



(a) (b)

Figure 10.13.: Pull (difference relative to the uncertainty) of the observed efficiency of the MET45 path and the predictions from the two parameterizations in the bins filling the phase space of the two variables used to train the networks ($\not{E}_T$ and $E_T^{j_1}$); the pull is plotted in the (a) MUON and (b) JET-50 samples.

training. Clearly this check requires many plots, and this number is even bigger for parameterizations with a larger number of variables.

The perfect test consists in checking each point of the multidimensional space defined by the input variables; but it is clearly unfeasible. However, we can divide this space into a certain number of regions (bins), and checking the average behavior in each of these by comparing the measured efficiency to that predicted by the NN function. The ratio plots perform exactly that, but the concept has to be generalized to higher dimensions.

For the case presented in the previous section, we have divided the two-dimensional $(\not\!\!E_T, E_T^{j_1})$ space in 10,000 bins (100×100). Instead of using a χ^2 to quantify the agreement, we use the pulls of the predicted values (although these cannot be interpreted as pure pulls, for in one case we plot the information of a self-check). Figure 10.13 shows these distribution for the two samples. They give a clear idea on the level of agreement of the predictions with the measured efficiency in each region, which is very satisfactory in this case. The larger spread of the alternate sample can be used to account for any possible discrepancy or bias that needs to be covered by the quoted uncertainty.

It is worth noting that it is possible to add variables that were not used to train the NNs to this study, in order to check whether these should be included. In the case of a variable on which the trigger does not explicitly depend, the level at which the efficiency depends on that variable is determined by the correlation of the variable with those on which the trigger explicitly depends. Whether the variable needs to be added depends on the presence of topological effects affecting the efficiency. For instance, it might be worth knowing the location of the leading jet in $\eta - \phi$ space, because it may affect the efficiency since the $\not\!\!E_T$ resolution is different in different regions of the detector. Therefore, this variable needs to be included in the parameterization.

10.6.4 The Uncertainty on the Parameterization

Before proceeding to the more complex case that interests us, we discuss the important aspect of determining the systematic uncertainties associated to this technique, which are mainly due to effects in regions of low trigger efficiency that propagate into the networks. However, while the training sample can have a large fraction of its events in such regions, the events of interest to us are in the high and medium-high efficiency regions. We thus need ways to determine the uncertainty.

One way of achieving this is by plotting the relative difference between the average efficiencies of the two parameterizations in a given region as a function of the region, when the regions are ordered by efficiency. This is shown in figure 10.12. We observe that the agreement is good on average, and that the predictions are close to the true value of the efficiency. The discrepancy in the high efficiency regions hints to a possible dependence of the trigger on additional variables. For instance, when a jet falls into a crack, yielding high $\not\!\!E_T$ the event could be accepted by the $\not\!\!E_T$ trigger although the true $\not\!\!E_T$ is small; this dependence cannot be accounted for in the training due to the limited amount of information (input variables). The cross-sample and low efficiency issues have been discussed earlier.

10.6.5 Reproducing Known Results

Now that we have shown how a NN-based parameterization can properly describe the efficiency of a (simple) trigger path, we attempt to reproduce the fit-based parameterization of section 10.5 using a neural network. In what follows, the MET+CJET+JET path uses MET35+JETS for the first 2.2 fb^{-1} of data, and MET+DIJET thereafter (no overlap in time).

For simplicity, we compare only the two-jet parameterizations, and train a network with the same input variables as the fit-based parameterization, namely $\not\!\!E_T$, $\phi(\not\!\!E_T)$, $E_T^{j_1}$, $E_T^{j_2}$, η^{j_1}, η^{j_2}, $\Delta R(j_1, j_2)$, $\Delta\varphi(\not\!\!E_T, j_1)$, and $\Delta\varphi(\not\!\!E_T, j_2)$. The network configuration is $9:11:10:1$.

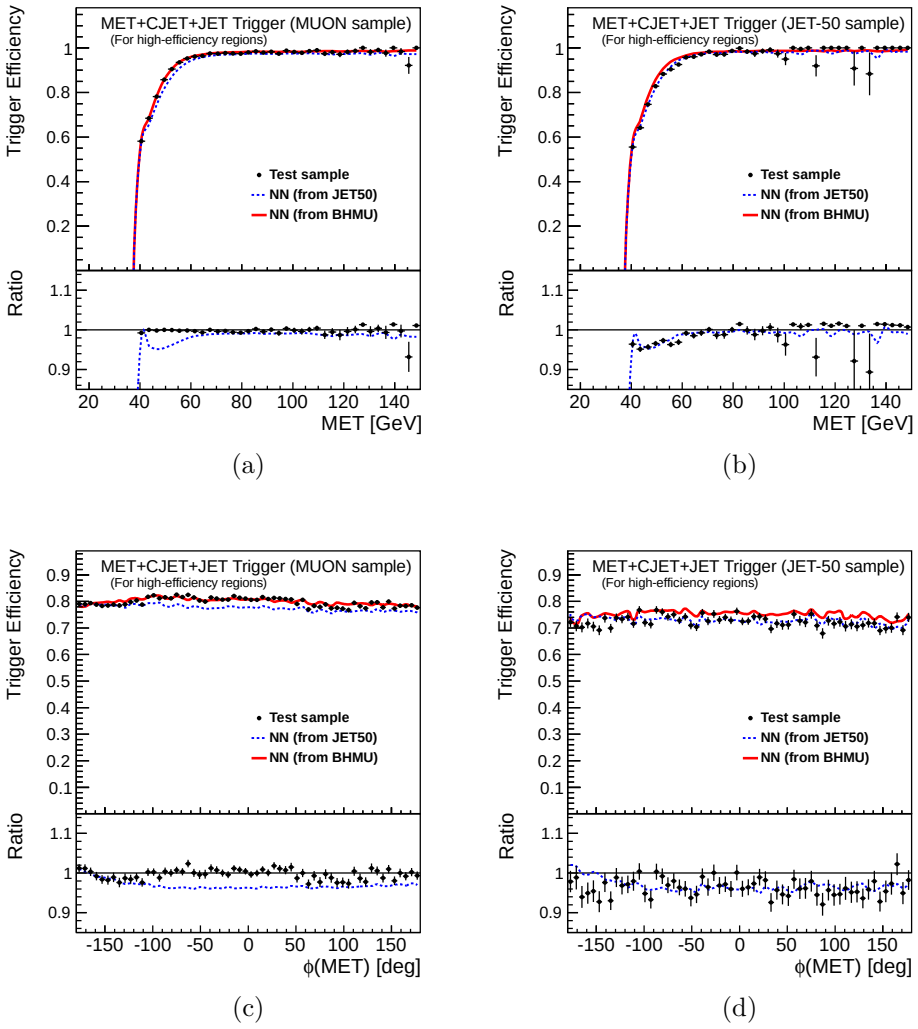


Figure 10.14.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $\not\!\!E_T$, and (c,d) $\phi(\not\!\!E_T)$.

Validation of the Network

To ensure that our parameterization is successful, we perform a validation of the output, and in a second step, described in the next section, we look at the performance of the new parameterization with respect to that of section 10.5.

The efficiency of the MET+CJET+JET path is plotted as a function the nine input variables in figures 10.14 to 10.17. Contrary to the two-variable parameter-

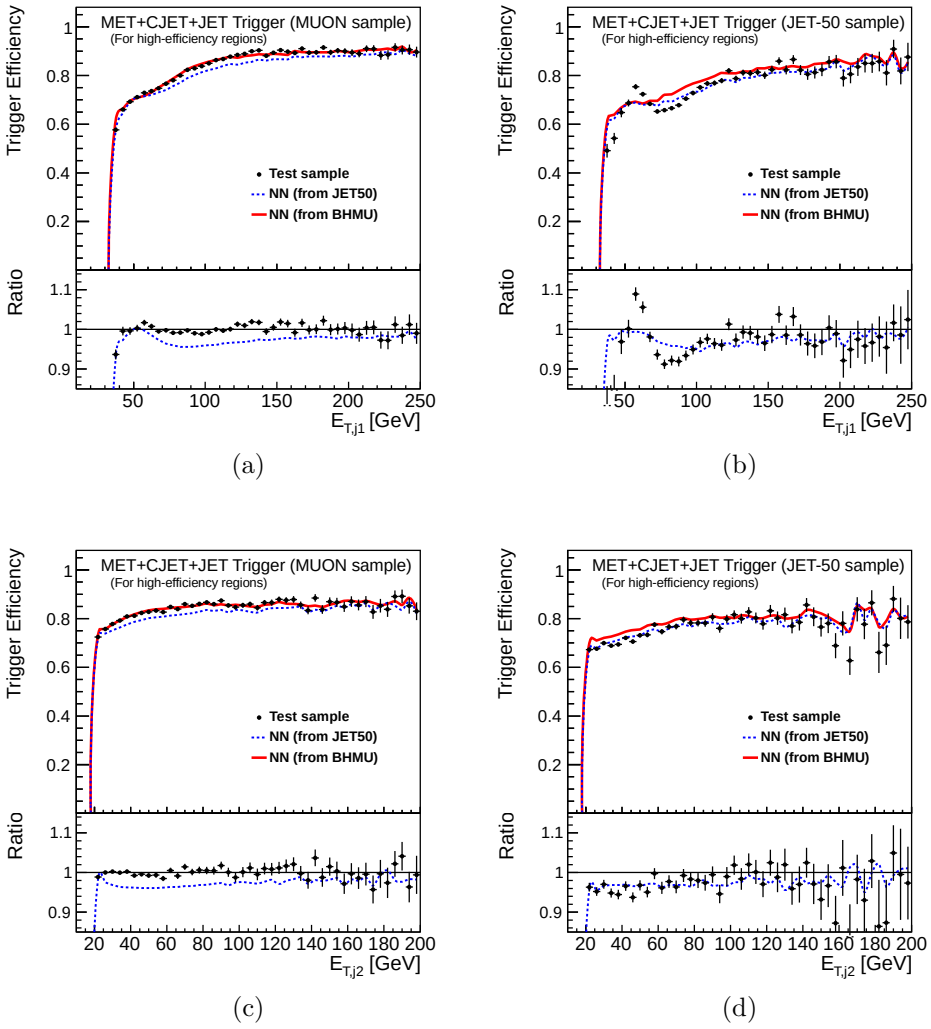


Figure 10.15.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $E_T^{j_1}$, and (c,d) $E_T^{j_2}$.

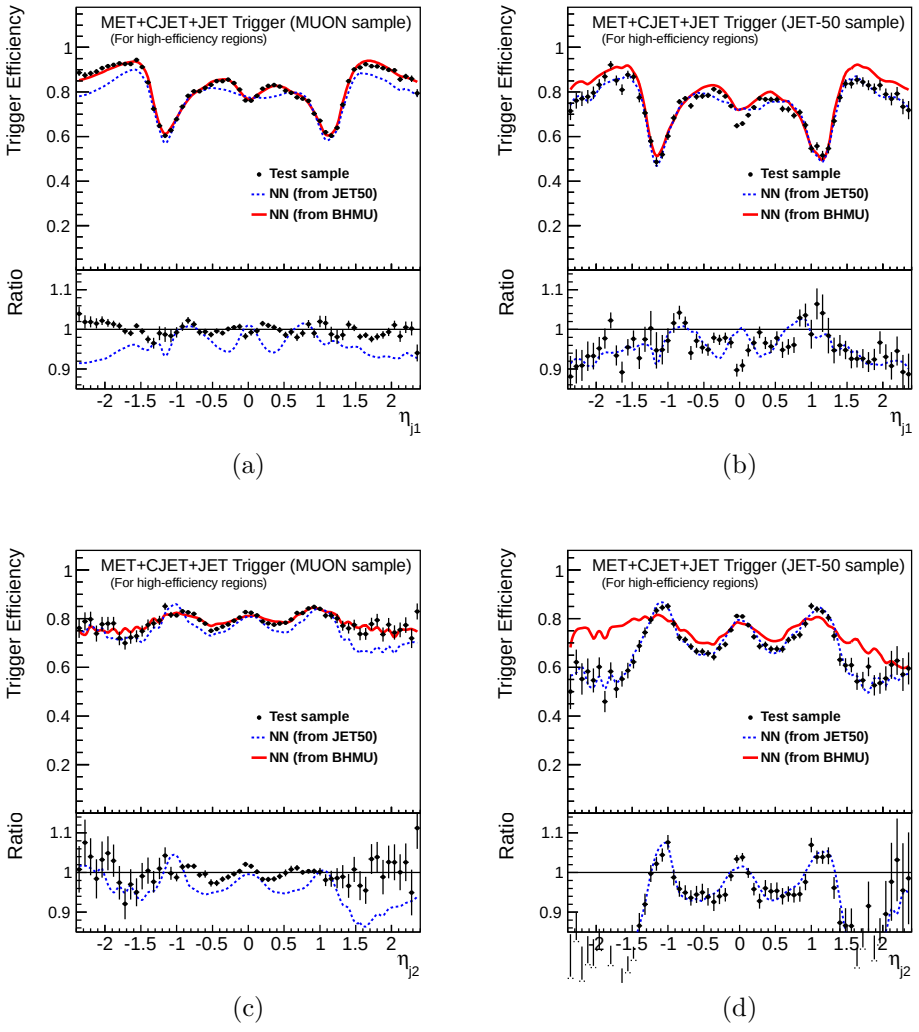


Figure 10.16.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) η^{j_1}, and (c,d) η^{j_2}.

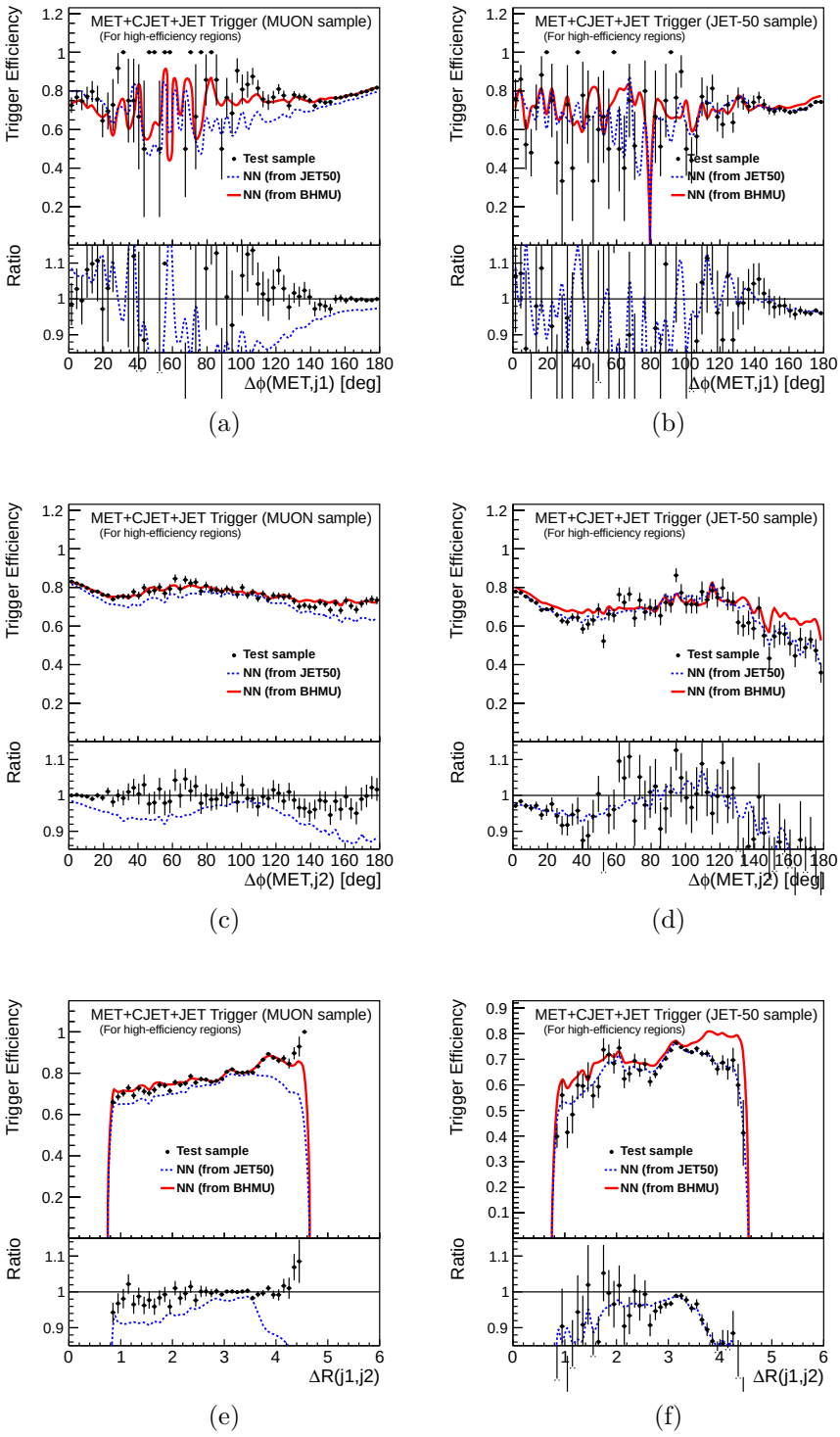


Figure 10.17.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) $\Delta\varphi(\not\!\!E_T, j_1)$; (c,d) $\Delta\varphi(\not\!\!E_T, j_2)$; and (e.f) $\Delta R(j_1, j_2)$.

ization of the MET45 path, we cannot plot the scattered distributions, since the high-dimentional nature of the problem yields random points when shown in two dimensions. We observe that the variables are well reproduced, except for some localized discrepancies here and there. For instance, some issues are present in the angular distributions, but these correspond to sparsely populated regions, and this should not affect MC samples which have higher event counts.

Differences between MUON and JET-50 are visible in the η distributions (figure 10.16) do not seem related to trigger cuts. We have not investigated the possible correlation with other variables, since our goal is not to reproduce the fit-based parameterization of section 10.5 but rather to provide a proof of concept. These discrepancies could provide an estimation of the uncertainties, and thus these effects would have been taken into account as systematic uncertainties if we were to use this parameterization.

Performance Studies

We now look at the multivariate parameterization on $ZH \to \nu\nu b\bar{b}$ events. It should be noted that we have to scale down the efficiency of this parameterization to account for the effective luminosity. This factor is required because the prescale was not taken into account in the fit-based parameterization described in section 10.5[11]. It is better not to include the prescales, since this allows to combine multiple triggers seamlessly. We have also selected the same events to apply both parameterizations to. For a fair comparison, we do not take the loose selection of chapter 11, since the fit-based parameterization was not derived with this loose selection in mind.

We thus select events satisfying the following requirements: (a) a $\not\!\!E_T > 40$ GeV; (b) exactly two jets with $E_T \geq 20$ GeV and $|\eta| < 2$, with the lead jet having $E_T > 35$ GeV and one of the two jets having $|\eta| < 0.9$; (c) $\Delta R(j_1, j_2) \geq 0.8$; (d) $\Delta\varphi(\not\!\!E_T, j_1) \geq 1.5$; (e) $\Delta\varphi(\not\!\!E_T, j_2) \geq 0.4$, and (f) the global clean-up cuts of

[11]It was, however, included in the determination of the weight to apply to each MC event.

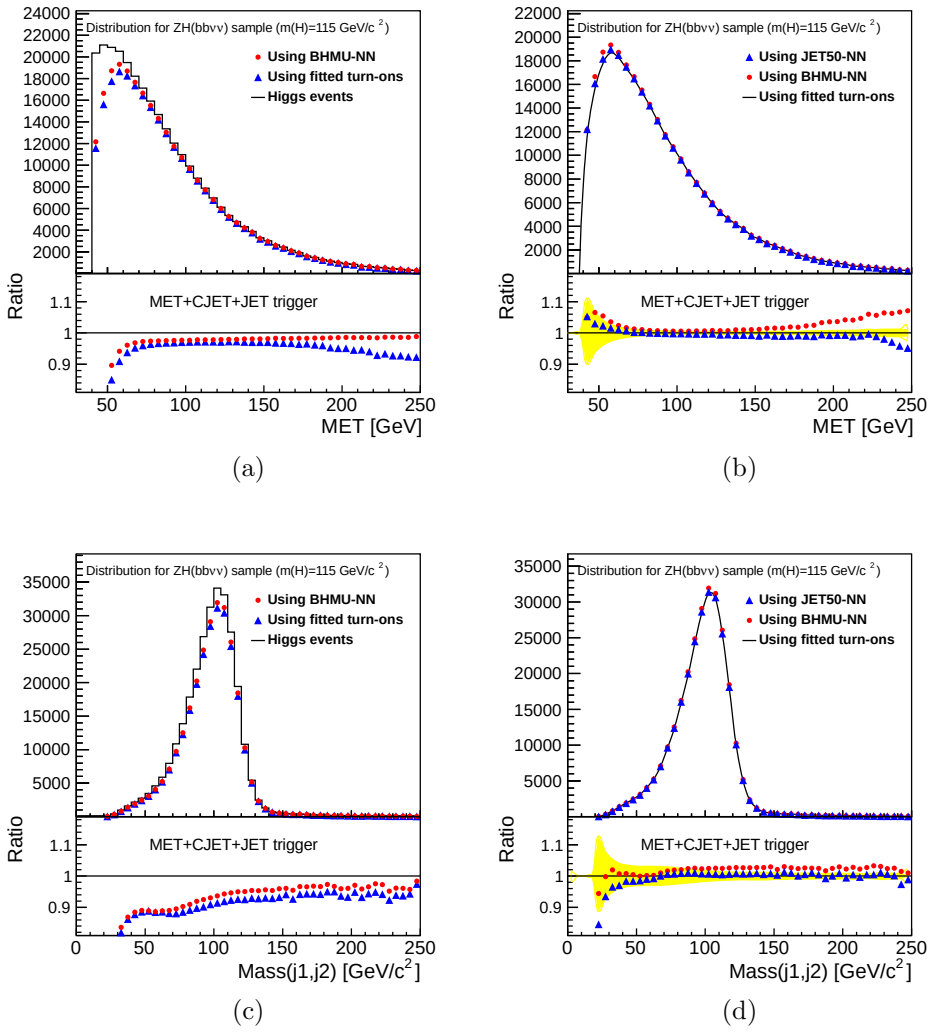


Figure 10.18.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $\not\!\!E_T$, and (c,d) $m(j_1, j_2)$.

section 8.3. In order to not introduce un-necessary complications, we only compare the parameterizations obtained from the MUON sample.

We check the performance in two ways. First, we compare the efficiency predicted by both the fit- and NN-based parameterizations when applied to a $ZH \rightarrow \nu\nu b\bar{b}$ Monte Carlo sample. Second, we compare the NN MUON and JET-50 parameterizations to the fit-based one (with its uncertainty). Figures 10.18 to 10.21 show the

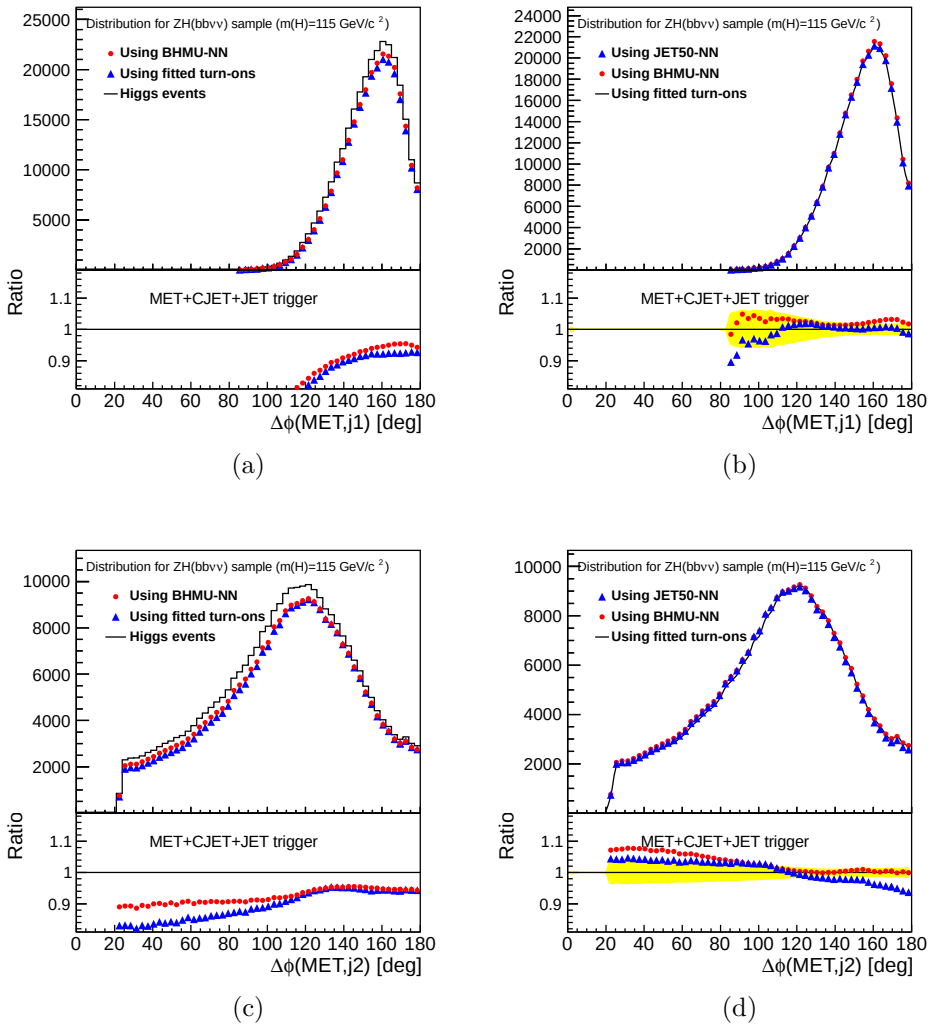
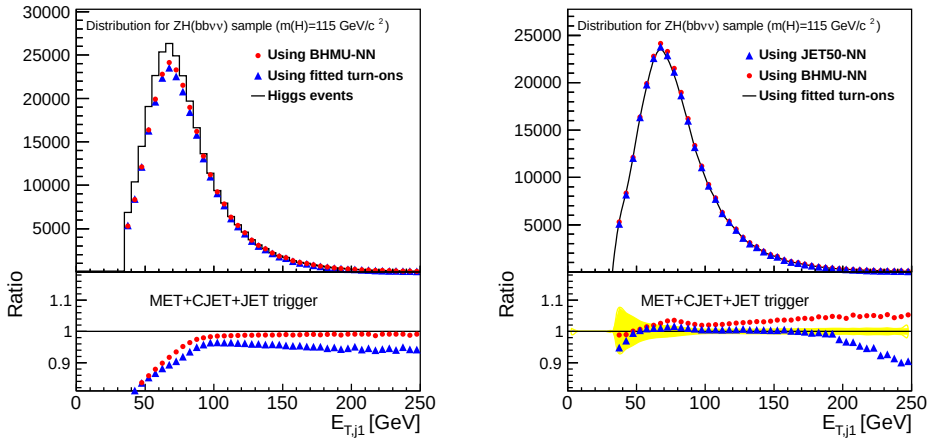
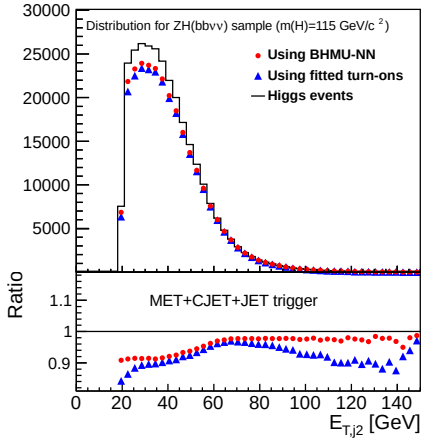


Figure 10.19.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) η^{j_1}, and (c,d) η^{j_2}.

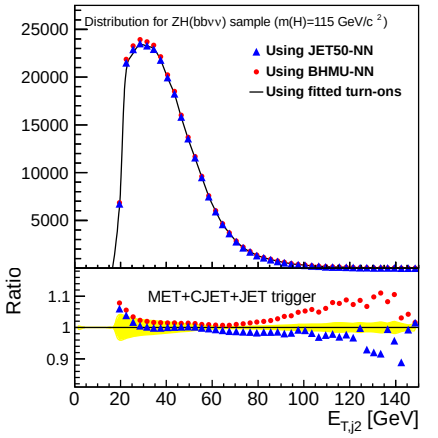


(a)

(b)



(c)



(d)

Figure 10.20.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) E_T^{j1}, and (c,d) E_T^{j2}.

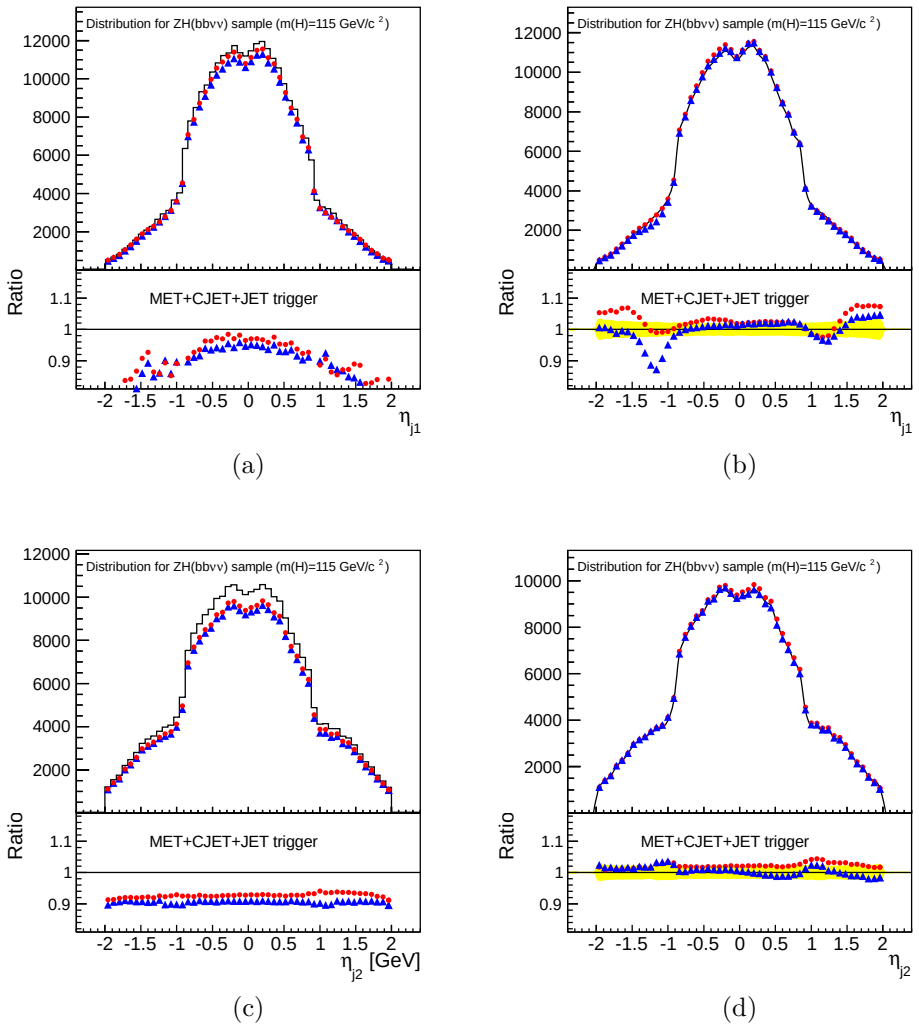


Figure 10.21.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) η^{j_1}, and (c,d) η^{j_2}.

result of these checks, which include also variables which are less well reproduced such as the energy and pseudo-rapidity of the jets.

We note that both parameterizations give compatible results. Regions where there are visible differences can be attributed to the fact that the NN parameterization is able to handle the correlations between the different variables better. For instance, this explains why the high-$\not\!\!E_T$ region exhibits a high efficiency in the new parameterization[12]. Additionally, a disagreement of $2-3\%$ could also be due to the differences in the data samples used to derive these parameterizations. Indeed, the cuts are not exactly the same, and the fit-based parameterization covers a smaller fraction of integrated luminosity with looser cuts on $\not\!\!E_T$ and tighter cuts on the E_T of the leading jet (because of the higher ratio of MET35+JETS to MET+DIJET events). These issues could explain some of the differences.

What is interesting looking at these results is that both parameterizations are in good agreement in high-efficiency regions, with a slight advantage for the NN parameterization (since it includes the correlations that were not taken into account in the fit-based approach). This fact constitutes strong evidence that the NN regression is actually performing well in an understood (and much studied) region.

Besides the comparison of the central parameterizations, it is worth investigating the evolution of the uncertainties coming from the two approaches. We expect a small benefit from the NN regression, since it handles the correlations in low-efficiency regions. From this study, we note that the NN-based parameterization has higher efficiency in the turn-on of areas (such as the low values of $\not\!\!E_T$), and that the uncertainty is reduced, since the differences between the MUON and JET-50 NN parameterizations are smaller than the uncertainty for the fit-based method. We note that the disagreements between the two NN parameterizations in the tails has only a small effect on the global uncertainty (for the cross-section is small in these regions).

[12]This was a known problem of the fit-based parameterization, which was unable to account for the correlations between multiple variables. In any case, that region was too little populated for this to be a real concern.

We expect these differences to be due to kinematic correlations in the trigger cuts (mainly the di-jet selection), although this was not investigated further.

Another cause of discrepancy is cross-sample extrapolation. Though we do not think that the differences between the MUON samples used to derive the fit- and NN-based central parameterizations and the Higgs sample are very likely have a big effect, since these differences were not observed when the parameterizations where applied to the data samples, this effect was not taken into account in the fid-based parameterization (and thus not included in the associated uncertainty), and hence the uncertainty for the NN-based parameterization should be more reliable. We believe that kinematic correlations are reduced with the inclusion of a simple trigger (MET45), which makes the NN parameterizations more stable, and simplifies the training process.

10.7 Final Neural Network Parameterization of the Efficiency

After having shown that we can reproduce known results using a new parameterization technique, we studied the regions of lower trigger efficiency and included the MET45 trigger path. This is presented in detail in section B.1. The final parameterization retained for this analysis uses the findings from this study, as well as from the studies described in the previous sections.

With respect to the selections from the previous section, we applied the following modifications: (a) we relax completely the requirements on the first jet, thus considering jets with $E_T \geq 20$ GeV and $|\eta| < 2.4$; (b) we remove the requirement of a central jet, and (c) we remove the requirement on the ΔR between the leading two jets. We use these requirements even if they are not a perfect match to those in our analysis to derive the most general parameterization.

We first train a network with a 9:15:14:1 configuration to parameterize the trigger efficiency, i.e. using the information provided by the $\not\!\!E_T$, and the two leading jets. The results of this training are presented in figures 10.22 to 10.25. We observe reasonable

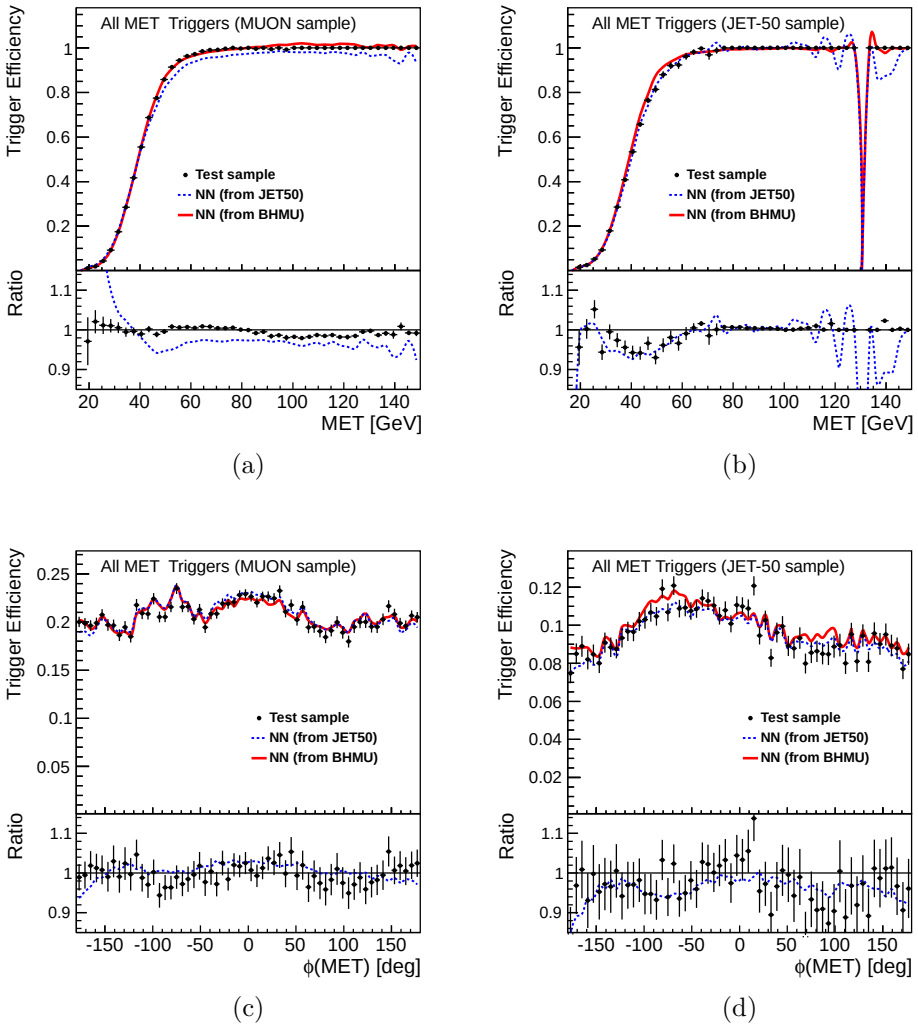


Figure 10.22.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $\not\!\!E_T$, and (c,d) $\phi(\not\!\!E_T)$.

description of the average efficiencies as a function of the input variables. Some regions are not perfect due the the limited precision, which affects some parts of phase space that are particularly hard to model. This is alleviated by using a reasonable selection. It should be noted that these regions will be assigned large systematic uncertainties due to the intrinsic differences between the MUON and JET-50 samples, for they have a lower efficiency.

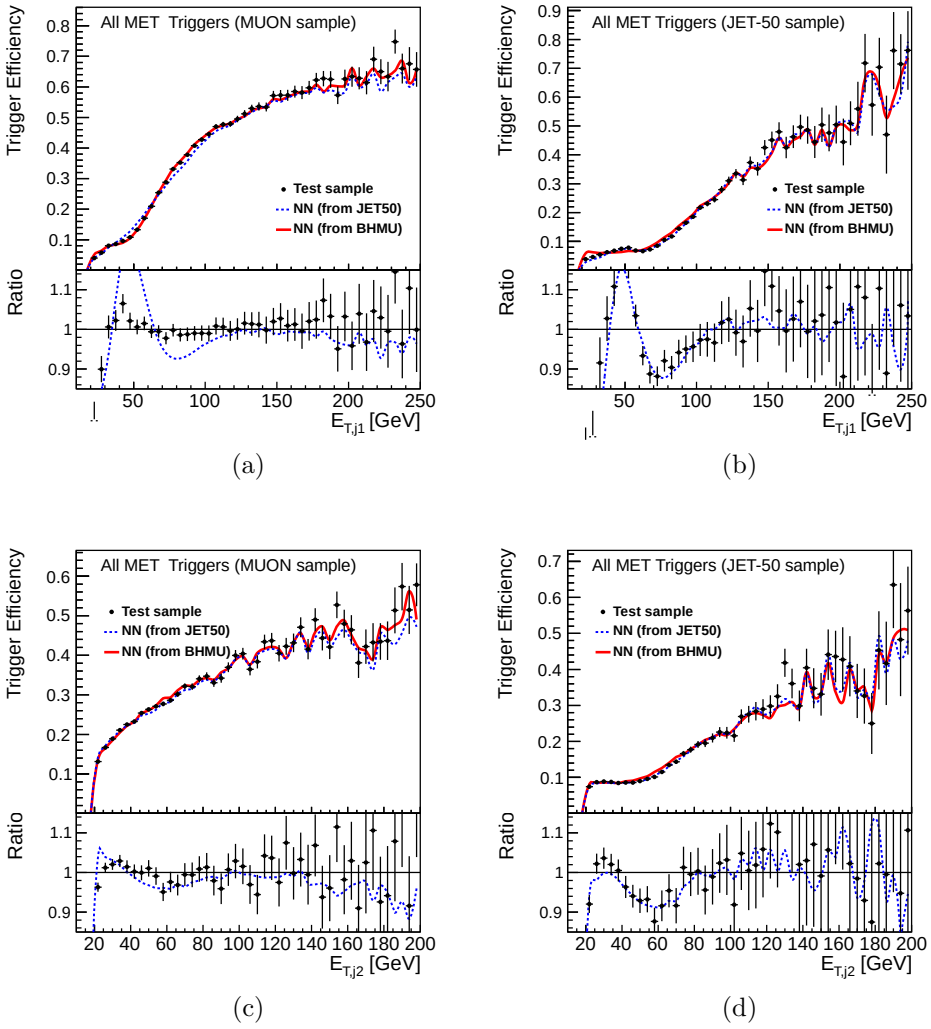


Figure 10.23.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) E_T^{j1}, and (c,d) E_T^{j2}.

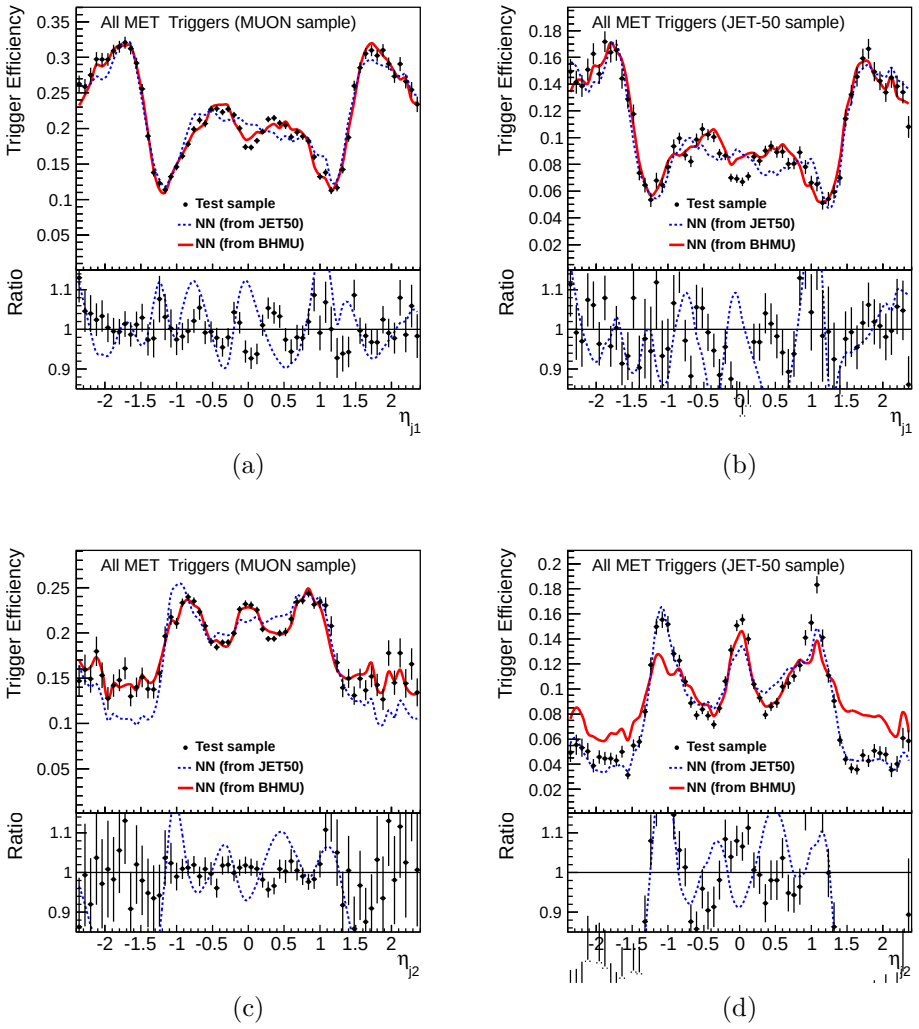
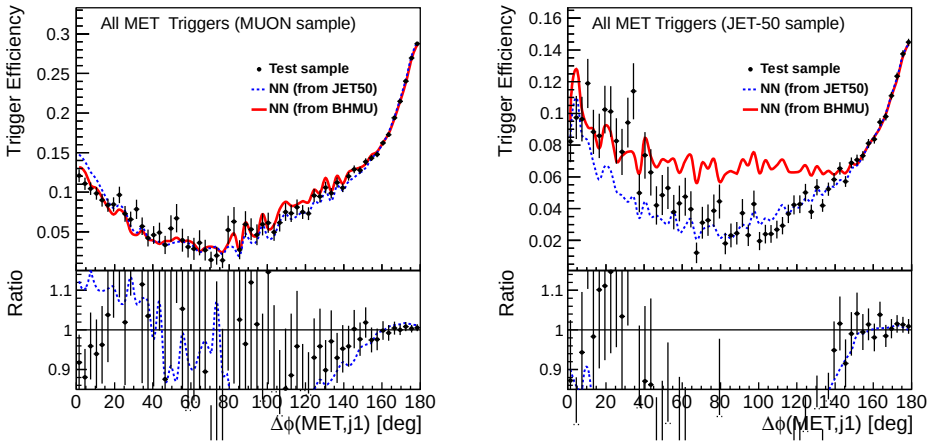
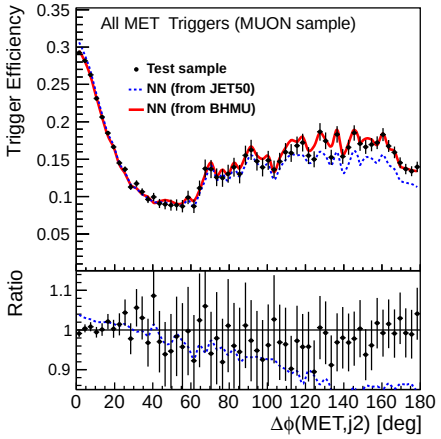


Figure 10.24.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) η^{j_1}, and (c,d) η^{j_2}.

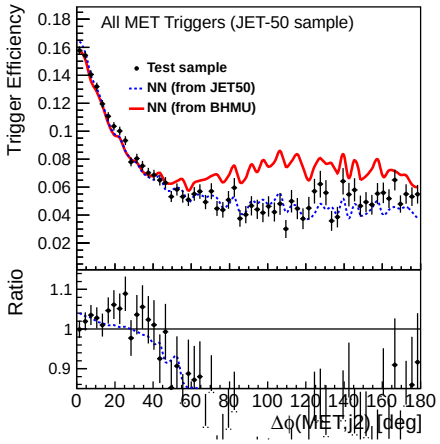


(a)

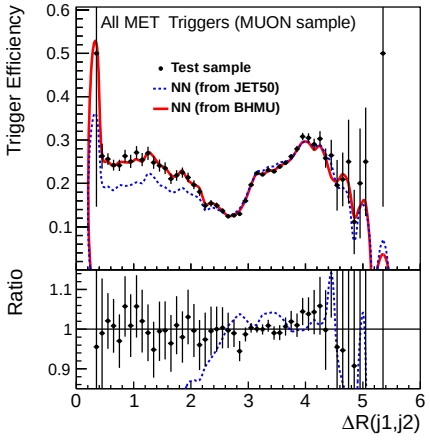
(b)



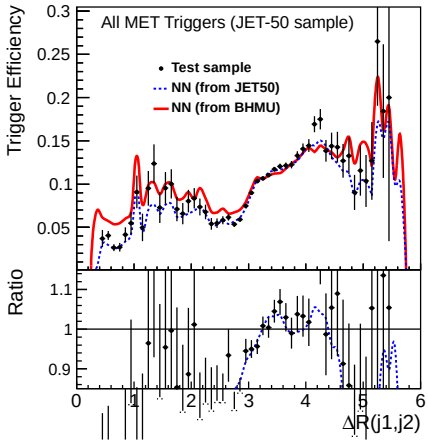
(c)



(d)



(e)



(f)

Figure 10.25.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) $\Delta\varphi(\not\!\!E_T, j_1)$; (c,d) $\Delta\varphi(\not\!\!E_T, j_2)$; and (e.f) $\Delta R(j_1, j_2)$.

We add some new features to improve the parameterization in the regions of interest: on top of slight changes to the structure of the network (additional nodes in the hidden layers), we train a network specifically for high-$\not\!\!E_T$ events in order to improve the precision. Additionally, we introduce two 3-jet parameterizations (of which one is dedicated to high-$\not\!\!E_T$ events), extending those for 2-jet events. For this purpose, we add five variables to the network: (a) the E_T of the third jet; (b) the pseudo-rapidity of the third jet; (c) the angular separation between the $\not\!\!E_T$ and the third jet in the transverse plane; (d) the separation in $\eta - \phi$ between the leading jet and the third jet $\Delta R(j_1, j_3)$, and (e) $\Delta R(j_2, j_3)$. We train a 14-input network with the following configuration $14 : 20 : 19 : 1$. These additions are presented in appendix **??**.

We now are in possession of a powerful, yet easy to implement, technique to parameterize the trigger efficiency. Moreover, this technique can be extended to the combination of any number of trigger paths. We have achieved our goal of reducing the complexity of the determination of the trigger efficiency. This technique is very generic and can be used in any experimental setup.

10.8 Systematic Uncertainties Associated with the Parameterizations

In this section we discuss several issues related to the uncertainty associated with the parameterization of the trigger efficiencies. Although not discussed here, there are effects introducing changes in the efficiency that must be taken into account even if these are not part of the parameterization; the jet energy scale is such an example.

We focus instead on the uncertainties related to the efficiency itself, which is the uncertainty associated to an event with fixed kinematic properties. The alternate parameterization (using the JET-50 sample) should be considered as plausible, since $\not\!\!E_T$ is a complicated quantity that could be due to a real undetected particle or to a mis-measurement. Additionally, this takes into account possible biases in the training samples, because the central and alternate samples can affect the estimation of the trigger efficiency differently.

Because the training process affects the parameterization, it is worth checking whether different input samples provide different effects. For instance, we checked that the 2- and 3-jet parameterizations are compatible with their high-$\not\!\!E_T$ counterpart, albeit for the issue related to the high-$\not\!\!E_T$ inefficiencies.

The structure of the network also plays a role in the parameterization. It is recommended to check alternative network structures and check how the parameterization is affected. Of course, computing time is a limiting factor here, for otherwise we would train networks with a higher number of nodes in the hidden layers.

In practice, we consider the outcome of the full analysis using each parameterization (central and alternate) and use the obtained distributions as the central and systematically varied shapes. This is easier to implement than an event-per-event uncertainty, which is hard to determine since the training samples have different event topologies.

10.9 Summary

In this chapter, we have described the $\not\!\!E_T$ triggers used to collect the events analyzed in this dissertation. We presented the process of event reconstruction at the various calorimeter levels, and proceeded with the derivation of a parameterization of the trigger efficiency.

Starting from a simple case, we moved towards more complex ones, ending with a full neural network-based parameterization of the efficiency. We showed evidence that the parameterization models the efficiency with very good accuracy, and provides clear benefits with respect to past methods. Additionally the new parameterization has the benefit of being much less labor-intensive, and of allowing the inclusion of additional triggers with a limited effort.

We do not blindly add multiple trigger paths. Reasonable selection cuts must also be put in place in order to derive (and apply) meaningful parameterizations. We recommend this technique for any analysis using a combination of triggers.

11. EVENT SELECTION

11.1 From Collisions to Events

The data processing model for the CDF experiment organizes the data collected into logical sets, each set being part of a *run*. Each run corresponds to a time period during which the detector and beam conditions are stable [193] and to which we assign an identifier denoted as *run number*. Every collision recorded by CDF is identified uniquely by its run and event numbers. The condition of the detector is continuously monitored during each data-taking period and quality bits are assigned. Whenever the detector was functional during an entire run or a section of it, the *good* bit is set, which means that the data can be used for physics analyses. Lists of good runs are periodically compiled by the Data Quality Group at CDF. These lists are organized by the various sub-detectors. For the analysis presented here, we require runs with good calorimeters and tracking system. These are needed for the proper identification of jets and leptons, as well as for b-tagging.[1] Figure 4.4 shows the amount of good quality data acquired as a function of time. The amount of analyzed data is computed from the integrated luminosities in each of the good runs used. For the analyses presented in this dissertation, we use $\mathcal{L}_{\text{int}} = 7.8\ \text{fb}^{-1}$, corresponding roughly to 80% of the total CDF Run II dataset marked good.

11.2 The Event Pre-selection

Events passing the $\not\!\!E_T$ triggers (cf. chapter 10) are required to be in the CDF list of good runs with good electromagnetic calorimeter and silicon, which ensures that all the necessary sub-detectors were operational during data taking. We exclude

[1] b-tagging is the process by which one attempts to identify jets originating from a b-quark.

runs where the detector was not fully operational.[2] Runs with missing beam lines are also removed. We now proceed with the presentation of the various selection cuts developed for the event pre-selection, which serves as basis to the analyses presented in this dissertation.

11.2.1 Trigger-driven Cuts

The systematic uncertainties originating from the choice of the trigger samples used in the efficiency calculations are large at small $\not\!\!E_T$, thus we require every event to have $\not\!\!E_T > 35$ GeV. This cut is significantly relaxed with respect to the 50 GeV cut in ref. [114, 161, 190]. This improvement is made possible by a multivariate parameterization of the trigger efficiency (cf. chapter 10) which allows for a better model of the turn-on region, and hence to relax the kinematic cuts.

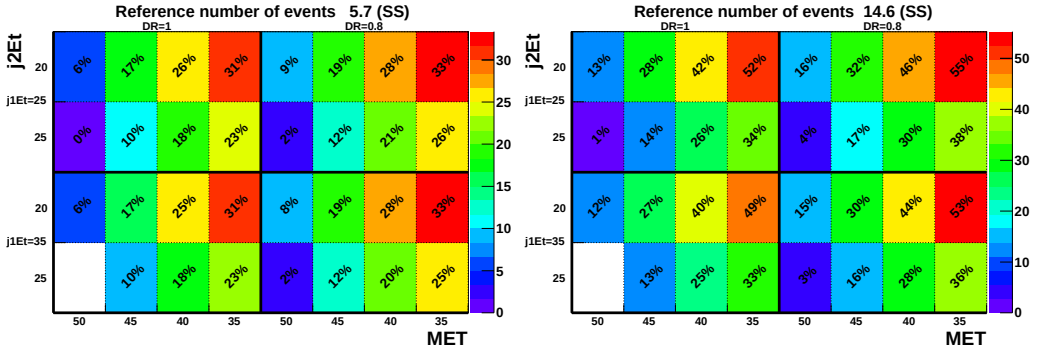
We require that the jets with the highest and second highest transverse energy satisfy the conditions $E_T^{j_1} > 25$ GeV and $E_T^{j_2} > 20$ GeV. Again this is a requirement that was loosened (from 35 GeV and 25 GeV cuts respectively) thanks to the improved trigger parameterization . We require the two leading jets to be in the central region ($|\eta^{j_i}| < 2$, with one of the two having $|\eta^{j_i}| < 0.9$) since we need tracking coverage to identify jets originating from b-quarks (cf. section 6).

To avoid calorimeter cluster merging at the L2 trigger (cf. section 10), we require that the separation between the two leading jets in the $\eta - \phi$ plane be larger than $\Delta R = 0.8$; this requirement is also loosened with respect to the previous iterations of the $\not\!\!E_T$+b-jets CDF analyses, which required $\Delta R > 1.0$.

The main motivation to relax the kinematic requirements is provided in figure 11.1 that shows an increase in the acceptance to rare standard model processes of $33 - 50\%$ (Higgs) and $55 - 70\%$ (diboson) in the most sensitive analysis category[3].The

[2] For instance, two run ranges are excluded due to the presence of hot towers biasing these periods of data-taking.

[3]Higher acceptance increases are obtain in less categories less pure than the one in figure 11.1.



(a) Higgs (b) Diboson

Figure 11.1.: Increase in event yields in the SS category in 5.7fb^{-1}: (a) Higgs, and (b) diboson. Each axis has two levels: $\Delta R(j_1, j_2)$ and ${E_T}^{j_1}$ define the quadrants, while $\not\!\!E_T$ and ${E_T}^{j_1}$ define the boxes within each quadrant, emphasizing their effect on the increase in acceptance.

acceptance increase for the electroweak production of single top is more moderate, at about $15 - 25\%$.

11.2.2 Lepton Identification Cuts

We do not need a high purity lepton sample, but the opposite, so we use a set of loose identification cuts to reject events with isolated leptons. We constructed these cuts to be loose enough to prevent double-counting signal events already considered in the channel with identified leptons. These requirements are described in section 5.5.

11.2.3 The Number of Jets

We accept events with 2 or 3 tight[4] jets. We expect mostly 2-jet events from the decay of the Higgs boson into $b\bar{b}$ pairs. However, we can increase our acceptance by roughly 50% by allowing events to have a third jet, which comes from additional radiation (initial- or final-state). Additionally, the inclusion of the third jet is useful in the context of a single top cross-section measurement, where t-channel production

[4]We denote as *tight* jets with $|\eta| < 2.4$ and uncorrected transverse energy $E_T > 15$ GeV/c^2.

yields three jets, while the s-channel yields mostly two. Also, in some cases the W decays to $e\nu$ and the electron fails the CDF electron identification algorithm, but is reconstructed as a jet by JetClu. This also occurs when the W decays to $\tau\nu$ and $\tau \rightarrow$ hadrons. Table 11.1 shows the contributions in signal region from WH processes in 2 and 3 jet events.

Process	All events		e, τ matched jet	
	2 jet	3 jet	2 jet	3 jet
$W \rightarrow \tau\nu$	44%	61%	2.8%	33%
$W \rightarrow e\nu$	38%	25%	0.6%	4%
$W \rightarrow \mu\nu$	18%	14%	–	–

Table 11.1: Contributions to 2/3jet events from different decay modes of the W boson in WH events.

The major drawback of accepting three-jet events lies in the increase of QCD multijet production and pair produced top background; the latter background is addressed at later stages in the analysis.

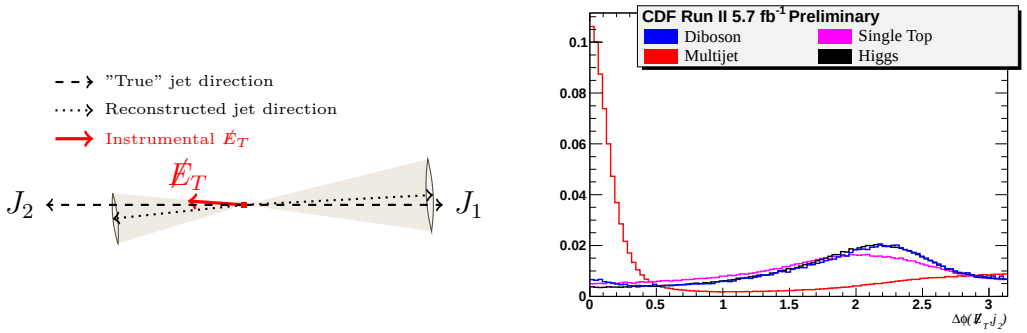
11.2.4 Rejecting Instrumental $\not\!\!E_T$

Most of background processes considered in this dissertation do produce real high $\not\!\!E_T$, e.g., W/Z decays to neutrinos or muons, which escape detection in the calorimeter. Additionally b-quarks produced in an event can decay semi-leptonically, which also produce real $\not\!\!E_T$. Mis-measurements in the calorimeter, on the other hand, can cause a QCD multijet event which has no real $\not\!\!E_T$ to appear with energy imbalance in the transverse plane. Since QCD multijet production has a very large cross-section, these events constitute a big fraction of the $\not\!\!E_T$+jets data sample. With the relatively low $\not\!\!E_T$ cut and no leptons, most of the data at this analysis stage is composed of QCD production of two or three jets, where one of the jets is poorly measured, resulting in a large transverse energy imbalance.

Angular Separation Between $\not\!\!E_T$ and the Jets

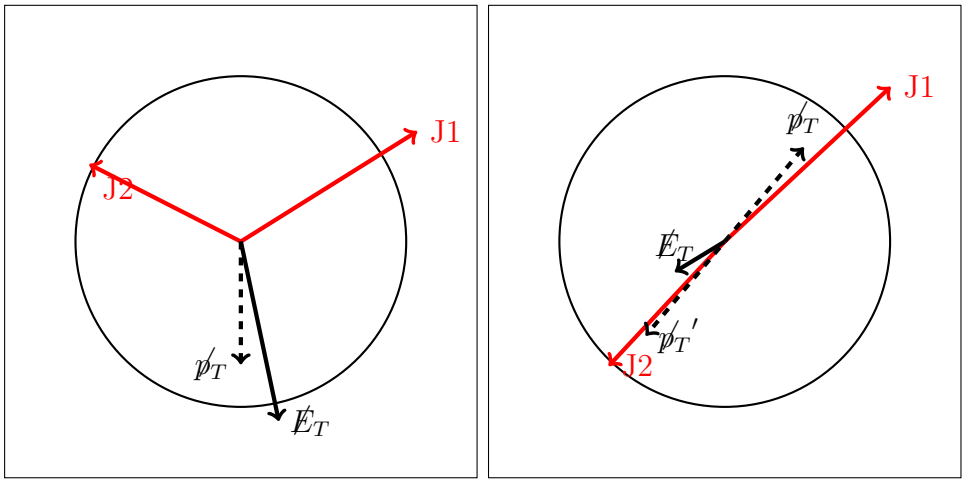
Due to this mis-measurement, most of the time the poorly measured jet will be the second highest E_T jet; therefore, the $\not\!\!E_T$ will be aligned to it in the transverse plane, as shown in figure 11.2. We label events having $\Delta\phi(\not\!\!E_T, j_2) < 0.4$ as belonging to the QCD control region; these events are used to derive a data-driven description of the multi-jet QCD background (cf. section 8.7.1) and to check the same technique on a very pure QCD region. For the pre-selection, where we want to isolate a signal, the above cut is reversed.

After that cut, QCD events where the $\not\!\!E_T$ is purely instrumental, and it is aligned with the jets, remain the dominant background. We cut on $\Delta\phi(\not\!\!E_T, j_1) > 1.5$ and $\Delta\phi(\not\!\!E_T, j_3) > 0.4$ to further suppress the QCD background. Those criteria remove a large amount of the QCD multijet background while having a small effect on the yields of signal events.



(a) (b)

Figure 11.2.: Identification of mis-measured jets. Schematic view of (a) a mis-measured QCD di-jet event, and (b) distribution of $\Delta\varphi(\not\!\!E_T, j_2)$ for several SM processes.



(a) (b)

Figure 11.3.: Transverse momentum flow imbalance in the tracker. Two processes: (a) $ZZ \to \nu\nu b\bar{b}$ (intrinsic $\not\!\!E_T$) and (b) $p\bar{p} \to b\bar{b}$ ($\not\!\!E_T$ from mis-measurement).

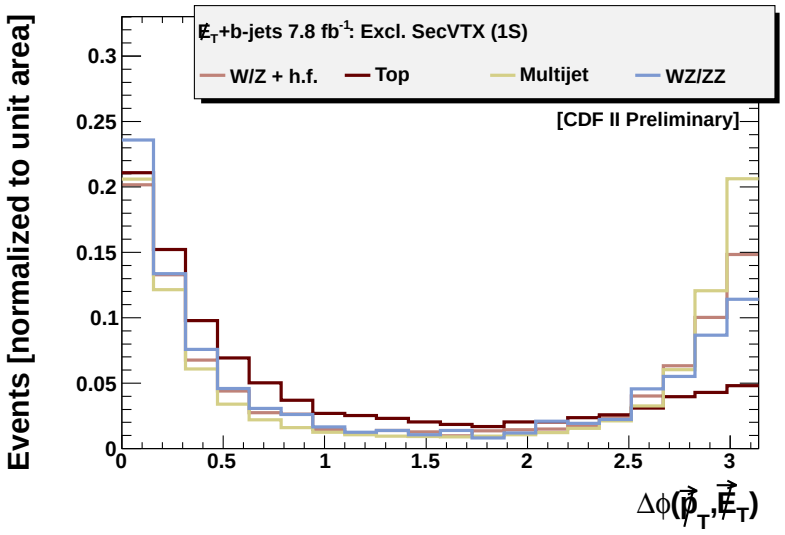


Figure 11.4.: $\Delta\phi(\not\!\!p_T, \not\!\!E_T)$ for events with no identified leptons, $35 < \not\!\!E_T < 70$ GeV, and with exactly one SecVtx-tagged jet the region where $\Delta\phi(\not\!\!E_T, j_2) < 0.4$ In this region, QCD multi-jet events are mis-measured, and therefore $\not\!\!p_T$ is either aligned or anti-aligned to $\not\!\!E_T$, while the other processes, which involve neutrinos, tend to have $\not\!\!p_T$ aligned to the $\not\!\!E_T$.

Tracker transverse momentum flow imbalance

We calculate the transverse momentum imbalance in the tracker ($\not\!\!p_T$), which is defined as negative vector sum of track p_T's:

$$\not\!\!p_T = \left| - \sum_{tracks} \vec{p}_T^{\,i} \right| \tag{11.1}$$

can be used to support the $\not\!\!E_T$ information and to discriminate between real and fake $\not\!\!E_T$. As shown in figure 11.4, $\not\!\!p_T$ is highly correlated with calorimeter $\not\!\!E_T$ for true $\not\!\!E_T$ events, while it is not for QCD events with mis-measured jets; this is shown schematically in figure 11.3. Thus, $\not\!\!p_T$ provides an additional handle to separate mis-measured events from real $\not\!\!E_T$ events.

The direction of $\not\!\!p_T$ is calculated similar to $\not\!\!E_T$

$$\varphi_{\not\!\!p_T} = \arctan\left(\frac{\not\!\!P_y^{tr}}{\not\!\!P_x^{tr}}\right) \tag{11.2}$$

where $\not\!\!P_x^{tr}$ and $\not\!\!P_y^{tr}$ are sums of track momenta in x and y directions respectively.

We found that the most discriminating variables from tracking are, ranked in order of separation power, the $\Delta\phi(\not\!\!p_T, \not\!\!E_T)$, the absolute amount of $\not\!\!p_T$, the $\Delta\phi(\not\!\!p_T, j_i)$, and use them all in the following.

11.2.5 Flavor categories

After the selection cuts described in the previous sections, we remain with a signal-to-background ratio of $S/B \sim 1/4000$ in the Higgs analysis. In order to improve this ratio further, we identify jets originating from a b-quark. We do so by employing both the SecVtx [102] (cf. section 6.1) and JetProb [119] (cf. section 6.2) b-tagging algorithms. The first identifies the displaced (secondary) vertex where the b quark decayed, while the second estimates the probability for a jet to not originate from the primary vertex. When using the SecVtx (JetProb) algorithm, we consider the

Table 11.2: Effect on the event selection on the yields for rare SM processes determined using 7.8 fb^{-1} of CDF data.

	Data	[%]	Higgs	[%]	Single Top	[%]	$WZ + ZZ$	[%]
Good Runs	14258901	100.0	156.4	100.0	2011.2	100.0	3362.9	100.0
Trigger driven selections	14111237	99.0	151.1	96.6	1953.0	97.1	3239.5	96.3
Lepton veto	13911664	97.6	105.1	67.2	1269.3	63.1	2544.2	75.7
$\not\!\!E_T$ not collinear to any jet	3709806	26.0	92.7	59.3	1111.4	55.3	2252.9	67.0
SecVtx + SecVtx	4800	0.0	12.1	7.7	68.5	3.4	34.0	1.0
SecVtx + JetProb	9609	0.1	11.8	7.6	74.3	3.7	37.7	1.1
Exclusive SecVtx	149675	1.0	28.7	18.3	386.0	19.2	150.7	4.5

"tight" operating point (a probability of less than 5% for the jet to originate from the primary vertex). We subdivide the sample into three orthogonal flavor categories: SS, where both jets are tagged by SecVtx; SJ, where one jet is tagged by SecVtx and the other by JetProb; and 1S, where one jet – and only one – is tagged by SecVtx. The double-tagged samples provide the most sensitivity in this analysis, while the single-tagged sample adds $\sim 10\%$ to the overall sensitivity. The effect on the signal-to-background ratio can be obtained form table 11.2.

11.2.6 Neural network to remove QCD

The main background in this search is the QCD production of two or three jets. We investigated the dynamics of the events in the sample, using a QCD heavy flavor Monte-Carlo simulation. Looking at a large set of variables, we keep here only the ones for which QCD has a very different behavior with respect to the signal and the remaining backgrounds; the idea is that we will *remove* events that are not signal-like with an artificial neural network (NN), and then use a second NN to *discriminate* signal from the surviving, more signal-like, backgrounds. The NN presented here is an improved version with that from the previous iterations, rejecting more backgrounds and keeping high signal acceptance. This approach to remove QCD backgrounds has been successfully used in the search for other signals, such as single top [161].

We train a mixture of 50% WH events and 50% ZH events with $m_H = 115$ GeV/c^2 against pre-tagged data weighted by our *Tag-Rate-Function*, which is the first step

of our data driven technique to estimate the QCD multi-jet background (cf. section 8.7.1). The variables used in the training are shown in table 11.3. We use the Multi Layer Perceptron (MLP) which is a simple feed-forward network, as implemented inside the TMVA [184, 187] package (v4.1.0). We will refer to the output of this network as NN_{QCD}.

Table 11.3: Input variables to the neural network to reject QCD NN_{QCD}.

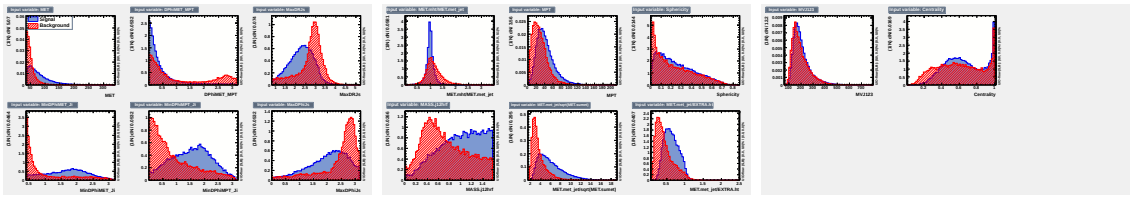
The energy imbalance in the calorimeter	$\not\!\!E_T$
The momentum flow imbalance in the tracker	$\not\!\!p_T$
The ratio between the vector sum of the jets and the $\not\!\!E_T$	$\not\!\!H_T/\not\!\!E_T$
The significance of the $\not\!\!E_T$	$\not\!\!E_T/\sum E_T$
The ration between $\not\!\!E_T$ and the scalar sum of the jets	$\not\!\!E_T/H_T$
The invariant mass of the $\not\!\!E_T$ and all jets in the event	$m(\not\!\!E_T, \text{jets})$
The sphericity of the event [1]	Sphericity
The centrality of the event $\sqrt{\sum(p_x^2+p_y^2)/\sum(p_x^2+p_y^2+p_z^2)}$	Centrality
The separation in ϕ between the direction of the two leading jets in their rest frame and the direction of the boost	ϕ^*
The angular separation between $\not\!\!E_T$ and $\not\!\!p_T$	$\Delta\phi(\not\!\!E_T, \not\!\!p_T)$
The maximum separation in R $(\eta-\phi)$ between the jets	$\text{Max}(\Delta R(j_i, j_k))$
The maximum separation in ϕ between the jets	$\text{Max}(\Delta\phi(j_i, j_k))$
The minimum separation in ϕ between the $\not\!\!E_T$ and any jet	$\text{Min}(\Delta\phi(\not\!\!E_T, j_i))$
The minimum separation in ϕ between the $\not\!\!p_T$ and any jet	$\text{Min}(\Delta\phi(\not\!\!p_T, j_i))$

[1] The momentum tensor is defined as $M_{lm} = \sum_i j_l^i j_m^i / \sum_i |\vec{j^i}|$, where $\vec{j^i}$ is the momentum of a reconstructed jet, and l and m are Cartesian coordinates. The index i runs over the number of jets in the event. The sphericity in an event is defined as $S = \frac{3}{2}(\lambda_2 + \lambda_3)$, where λ_2 and λ_3 are the smallest two eigenvalues of the normalized momentum tensor.

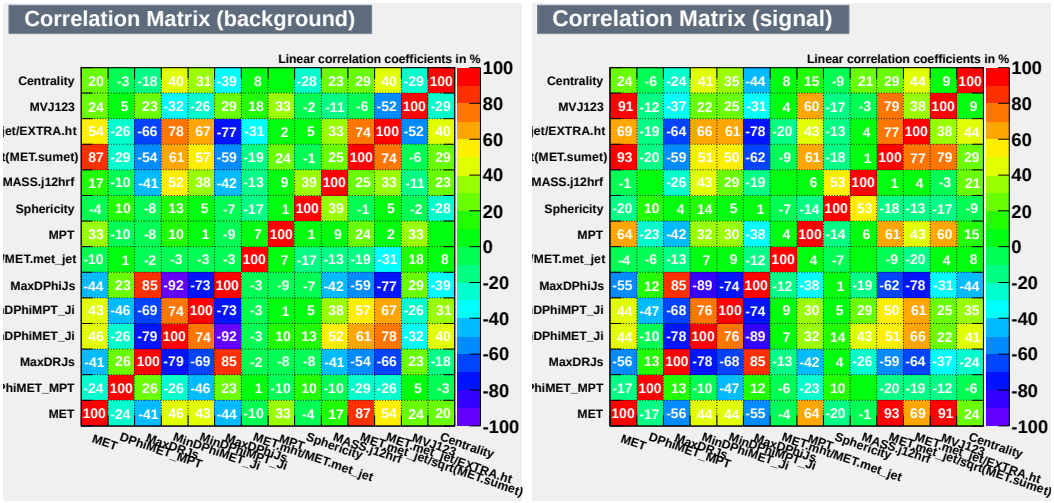
We use the output of this NN to define the final signal region ($NN_{QCD} > 0.45$. This cut reduces QCD by about one order of magnitude while keeping the signal efficiency between 90 and 95% (table 11.4).

Cat.	Sig. Acc.	Bkg. Rej.	QCD Rej.
1S	90.0%	72.6%	89.1%
SS	94.9%	48.2%	87.0%
SJ	93.6%	64.8%	87.4%

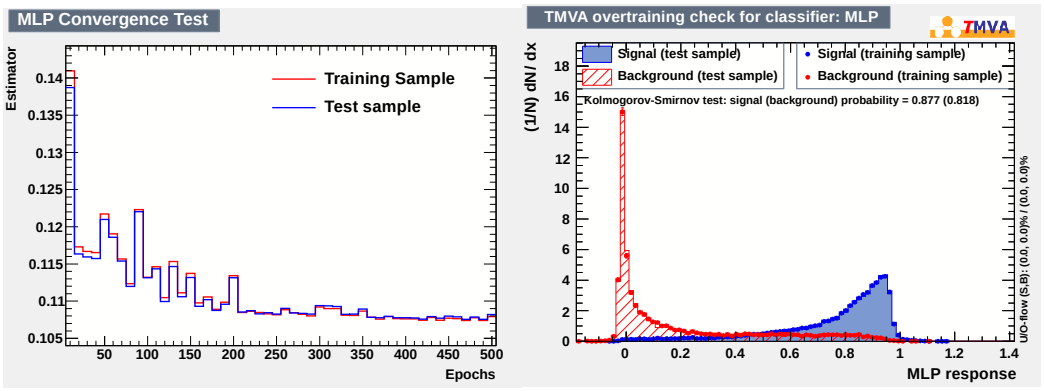
Table 11.4: Performance of the NN_{QCD} when discarding events with $NN_{QCD} \leq 0.45$.



(a) Input variables

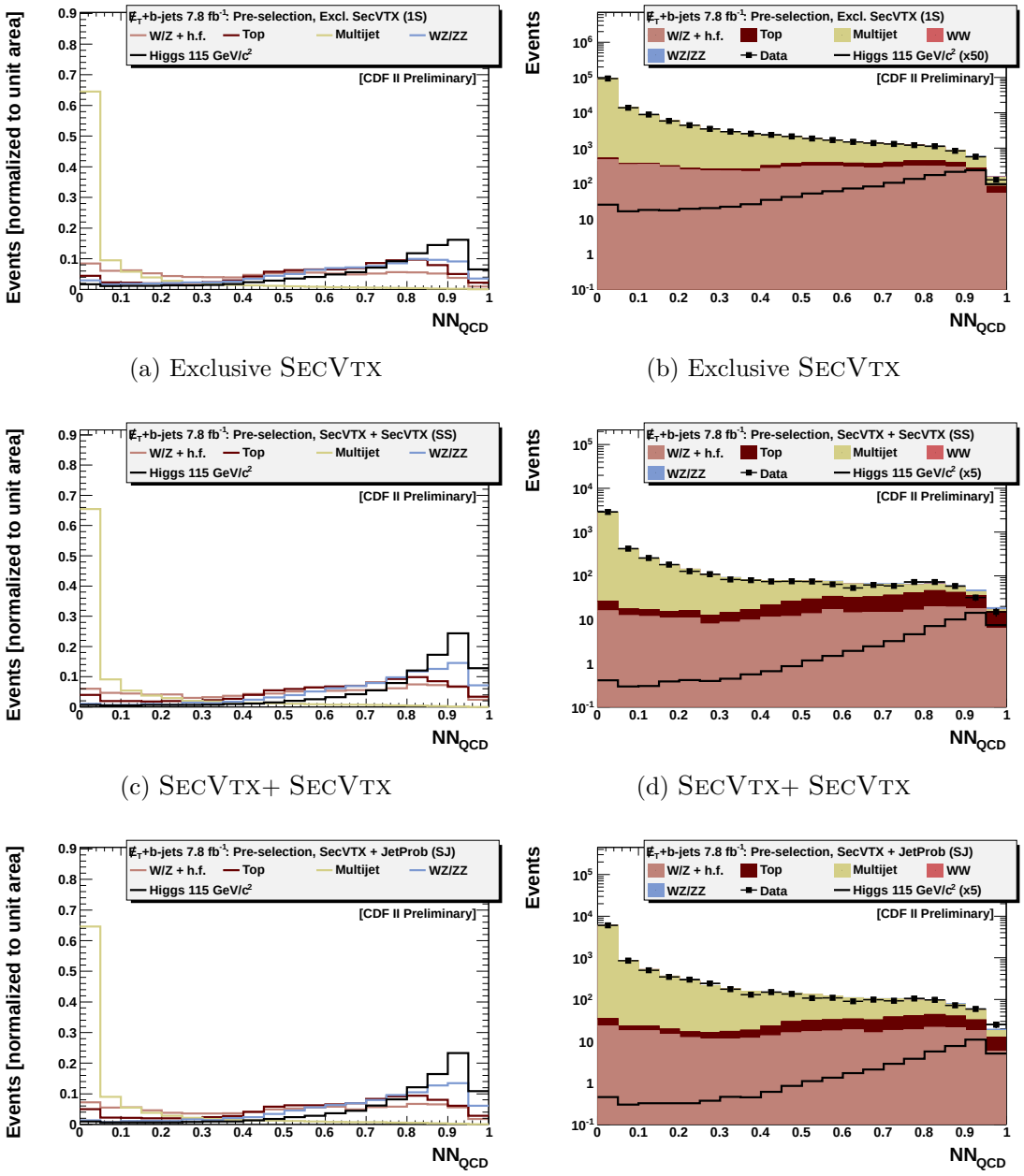


(b) Correlation matrices for the background (left) and the signal (right) samples



(c) NN convergence (d) NN ovetraining check

Figure 11.5.: Details on the training of the NN_{QCD} using events with $m_H = 115$ GeV/c^2: (a) inputs, (b) correlations, (c) convergence, and (d) output.



(a) Exclusive SecVtx (b) Exclusive SecVtx

(c) SecVtx+ SecVtx (d) SecVtx+ SecVtx

(e) SecVtx+ JetProb (f) SecVtx+ JetProb

Figure 11.6.: QCD Rejection Neural Network output

11.3 Control regions

As mentioned in section 8.7.1 the *Tag-Rate-Matrix* (TRM) used to model the multijet background has been derived from events with:

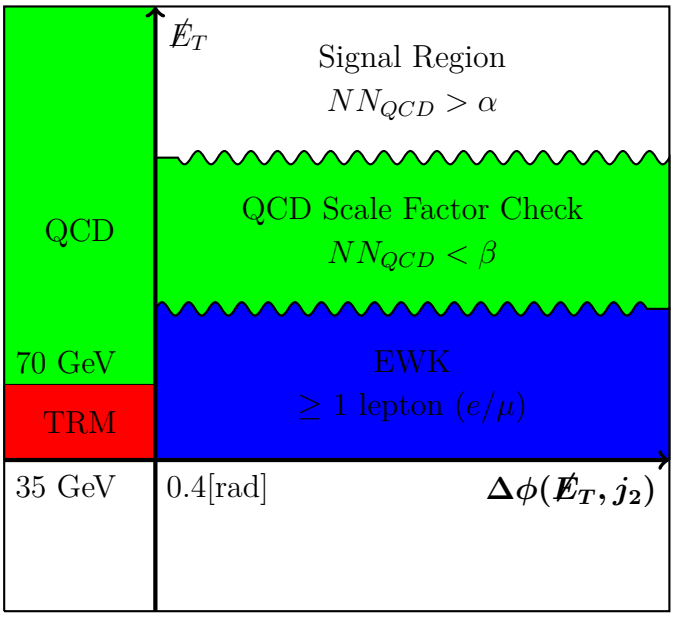


Figure 11.7.: The kinematic regions defined in the $\not\!\!E_T$+b-jets signature. The value of α is determined depending on the analysis goals, and is usually taken to reject at most 10% of the signal; it is equal to 0.45 for the Higgs analysis. We veto identified leptons in all regions except EWK. Additional selection criteria are listed in table 11.5.

- 2 or 3 jets, each with $|\eta| < 2.0$ and at least one being central ($|\eta| < 0.9$);
- $35 < \not\!\!E_T < 70$GeV, $E_T^{j_1} > 25$GeV, and $E_T^{j_2} > 20$GeV;
- $\Delta R(\vec{j_1}, \vec{j_2}) > 0.8$, and $\Delta\phi(\not\!\!E_T, j_2) < 0.4$;

This region is dominated by QCD, with minimal ($< 1\%$) contamination by other processes. The main characteristic of the selected events is that they are mis-measured.

We check our ability to predict the multijet backgrounds in two *control regions* (CR). The first, QCD CR1, is a high statistics region where we check the data-based model and evaluate the systematic uncertainties on the shapes of the various kinematic variables. In order to test our data-driven estimation in a more signal-like region, we define QCD CR2. This region is defined by reversing the cut on NN_{QCD} to remain blind to the signal region. One part of this region, $NN_{QCD} < 0.1$, is used to extract the multijet normalization as described in section 8.7.5. The region

QCD CR	EWK CR	QCD SF CR	Signal region
No lepton	At least one lepton	No lepton	No lepton
$\not\!\!E_T > 70$ GeV/c^2	$\not\!\!E_T > 35$ GeV/c^2	$\not\!\!E_T > 35$ GeV/c^2	$\not\!\!E_T > 35$ GeV/c^2
$\Delta\varphi(j_2, \not\!\!E_T) \leq 0.4$	$\Delta\varphi(j_2, \not\!\!E_T) > 0.4$	$\Delta\varphi(j_2, \not\!\!E_T) > 0.4$	$\Delta\varphi(j_2, \not\!\!E_T) > 0.4$
	$\Delta\varphi(j_3, \not\!\!E_T) > 0.4$	$\Delta\varphi(j_3, \not\!\!E_T) > 0.4$	$\Delta\varphi(j_3, \not\!\!E_T) > 0.4$
	$\Delta\varphi(j_1, \not\!\!E_T) > 1.5$	$\Delta\varphi(j_1, \not\!\!E_T) > 1.5$	$\Delta\varphi(j_1, \not\!\!E_T) > 1.5$
		$NN_{QCD} < 0.1$	$NN_{QCD} > 0.45$

Table 11.5: Main kinematic selection requirements for each of the control regions and the signal region.

with $0.1 < NN_{QCD} < 0.45$ serves as a medium statistics cross check of the multi0jet normalization with non-negligible electroweak contribution.

Since in the signal region we expect backgrounds originating from events with real high $\not\!\!E_T$, such as W/Z+jets, $t\bar{t}$, single top production and diboson production, we test our ability to predict these in an other control region, EWK CR. To remain unbiased to our final region, we test electroweak/top backgrounds in the kinematic region similar to signal region, with the exception of requiring at least one lepton in the event (all events with leptons are rejected from the signal region).

Table 11.5 summarizes the information regarding each control region and the final signal region. Tables 11.6 to 11.9 show the expected[5] and observed event yields in the control regions, while table 11.10 shows the *prior* expected and observed event yields in the signal region. The posterior yields are very close to the prior expected yields (cf. next chapter). This is an indication of the quality of our modeling and its ability to adequately represent the data.

The distributions for several kinematic variables used as inputs to NN_{QCD} and NN_{SIG} are shown in figures 11.8 to 11.11. Additional plots are provided in appendix C[6]. The simulation is in good agreement with the data.

[5]By construction, the multi-jet background is normalized to the difference between the data and the MC simulation in the control regions, while it is normalized according to a factor derived from the region $NN_{QCD} < 0.1$ in the signal region.

[6]We also validate the inputs to the NN_{QCD} in the signal region (cf. figure 14.2)

$\not\!\!E_T$+b-jets 7.8 fb^{-1}: TRM Region [CDF II Preliminary]

	Exclusive SecVtx	SecVtx + SecVtx	SecVtx + JetProb
Diboson	27.9±5.0	1.9±0.5	2.6±0.6
Single Top	21.2±4.0	4.3±1.0	3.7±0.8
Top Pair	51.3±6.7	10.4±1.8	10.5±1.8
Zbb	915.6±301.9	123.7±45.0	145.1±51.5
Z+HF	27.2±9.1	3.8±1.5	4.2±1.6
Z+LF	40.8±14.0	0.4±0.3	1.6±0.9
W+HF	112.1±38.3	8.4±3.5	12.1±5.0
W+LF	118.3±44.5	0.3±0.3	1.9±1.8
Multijet	654408.4±1210.4	38777.9±273.1	67212.2±328.0
Mistags	160.2±13.5	1.2±0.1	3.8±0.5
Exp. B.	655883.1±1215.0	38932.3±273.1	67397.9±328.1
Data	655858	38933	67397
Higgs	1.5±0.2	0.7±0.1	0.6±0.1
Multijet S.F.	0.973±0.002	1.164±0.008	1.23±0.006
S/B	1/427668	1/60345	1/119431
$S/\sqrt{B}$	0.002	0.003	0.002
Significance	0.005	0.015	0.008

Table 11.6: Event yields in the TRM region.

$\not\!\!E_T$+b-jets 7.8 fb^{-1}: QCD Control Region [CDF II Preliminary]

	Exclusive SecVtx	SecVtx + SecVtx	SecVtx + JetProb
Diboson	6.0±1.5	0.2±0.1	0.5±0.2
Single Top	6.2±1.3	1.4±0.4	1.5±0.4
Top Pair	32.7±4.6	6.1±1.2	7.5±1.4
Zbb	68.5±27.5	7.3±4.4	10.5±6.6
Z+HF	11.6±4.2	1.6±0.7	1.8±0.8
Z+LF	9.7±3.9	0.0±0.0	0.3±0.2
W+HF	30.8±11.2	2.6±1.3	4.2±2.3
W+LF	55.3±24.0	0.4±0.4	1.8±1.7
Multijet	24765.6±240.2	1550.5±54.4	3511.7±83.2
Mistags	83.4±7.8	0.8±0.1	2.3±0.3
Exp. B.	25069.8±244.2	1571.0±54.5	3542.0±83.4
Data	25070	1571	3542
Higgs	0.6±0.1	0.1±0.0	0.2±0.0
Multijet S.F.	0.938±0.009	1.177±0.041	1.422±0.034
S/B	1/43348	1/10153	1/20214
$S/\sqrt{B}$	0.004	0.004	0.003
Significance	0.013	0.027	0.017

Table 11.7: Event yields in QCD control region.

$\not\!\!E_T$+b-jets 7.8 fb^{-1}: EWK Region [CDF II Preliminary]

	Exclusive SecVtx	SecVtx + SecVtx	SecVtx + JetProb
Diboson	145.2±24.2	12.7±2.6	11.8±2.4
Single Top	235.2±42.4	45.9±9.2	37.1±7.5
Top Pair	906.9±111.9	255.2±37.2	215.9±32.1
Zbb	5.0±3.4	0.7±0.7	0.3±0.3
Z+HF	84.0±27.4	12.1±4.3	12.4±4.4
Z+LF	55.8±18.5	0.1±0.1	0.6±0.4
W+HF	866.0±288.9	83.7±29.4	83.6±29.4
W+LF	249.0±91.5	1.3±1.2	1.3±1.1
Multijet	1699.3±463.1	64.4±62.8	184.7±59.9
Mistags	1167.6±94.0	7.9±0.9	26.2±3.0
Exp. B.	5414.0±648.5	484.0±85.6	574.0±80.1
Data	5414	484	574
Higgs	13.1±1.1	5.7±0.7	4.5±0.5
Multijet S.F.	0.704±0.192	0.817±0.797	0.968±0.314
S/B	1/414	1/85	1/129
$S/\sqrt{B}$	0.178	0.258	0.186
Significance	0.266	0.412	0.318

Table 11.8: Event yields in EWK control region.

$\not\!\!E_T$+b-jets 7.8 fb^{-1}: Pre-selection Region [CDF II Preliminary]

	Exclusive SecVtx	SecVtx + SecVtx	SecVtx + JetProb
Diboson	336.0±54.5	33.9±6.3	32.7±6.2
Single Top	428.0±76.8	77.2±15.2	65.2±13.0
Top Pair	940.3±116.0	197.1±28.9	180.0±26.9
Zbb	398.4±136.4	24.4±13.4	22.7±11.1
Z+HF	631.6±204.6	85.0±28.8	87.0±29.5
Z+LF	358.9±117.1	1.5±1.1	2.7±1.4
W+HF	1819.7±606.5	143.5±50.4	185.6±65.1
W+LF	2310.4±777.0	7.4±5.3	20.4±11.6
Multijet	141069.8±1828.1	4259.8±128.9	9026.6±162.9
Mistags	2884.3±229.2	19.2±2.1	64.0±7.3
Exp. B.	151177.4±2507.9	4849.1±161.0	9686.9±199.3
Data	151180	4849	9687
Higgs	29.3±2.4	11.7±1.3	9.3±1.1
Multijet S.F.	1.003±0.013	0.952±0.029	1.014±0.018
S/B	1/5160	1/416	1/1046
$S/\sqrt{B}$	0.075	0.167	0.094
Significance	0.343	0.723	0.477

Table 11.9: Event yields in pre-selection.

$\not\!\!E_T$+b-jets 7.8 fb^{-1}: Signal Region [CDF II Preliminary]

	Exclusive SecVtx	SecVtx + SecVtx	SecVtx + JetProb
Diboson	268.2±43.7	30.1±5.7	28.0±5.3
Single Top	314.1±56.5	60.2±11.9	49.1±9.8
Top Pair	727.2±89.9	153.9±22.7	137.4±20.6
Zbb	74.9±31.6	6.2±5.0	5.2±4.2
Z+HF	394.5±128.0	56.5±19.3	54.6±18.7
Z+LF	209.2±68.7	0.9±0.7	1.6±1.0
W+HF	1064.8±355.4	94.5±33.7	111.5±39.7
W+LF	1173.8±399.3	5.3±4.2	13.3±8.4
Multijet	7566.3±36.5	275.7±5.7	623.5±8.5
Mistags	1955.6±155.9	14.2±1.6	46.0±5.2
Exp. B.	13748.6±964.3	697.4±68.0	1070.0±75.0
Data	13791	636	994
Higgs	25.4±2.1	10.9±1.2	8.6±1.0
Multijet S.F.	0.994±0.005	0.969±0.02	1.017±0.014
S/B	1/542	1/64	1/125
$S/\sqrt{B}$	0.216	0.413	0.261
Significance	0.349	0.725	0.479

Table 11.10: Prior event yields in the signal region.

11.4 Summary

In this chapter, we have presented the main event selection for the analysis. We use kinematic selection criteria and b-jet identification to remove the bulk of the backgrounds and to isolate the Higgs signal. These reduce the total background by more than two orders of magnitude. We train a neural network to separate the remaining multijet background from the signal, and retain only the events having $NN_{QCD} > 0.45$, reducing the background by yet another order of magnitude. We carefully validate the variables used in this analysis in several control regions, and observe excellent agreement between our simulation and the data.

The search for the SM Higgs boson is described in detail in chapter 14.

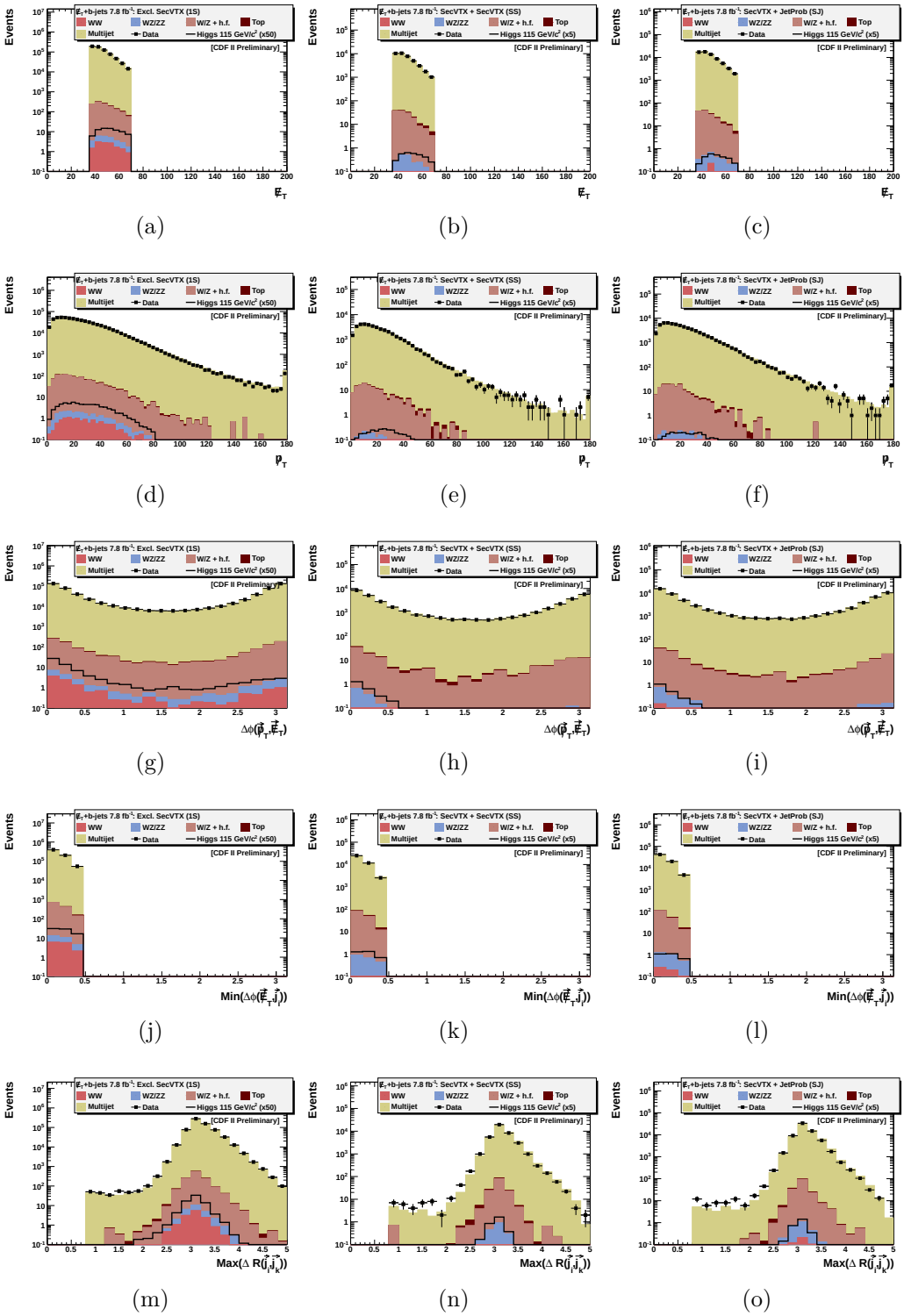


Figure 11.8.: Distribution of several input variables to NN_{QCD} in the TRM region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data (see text).

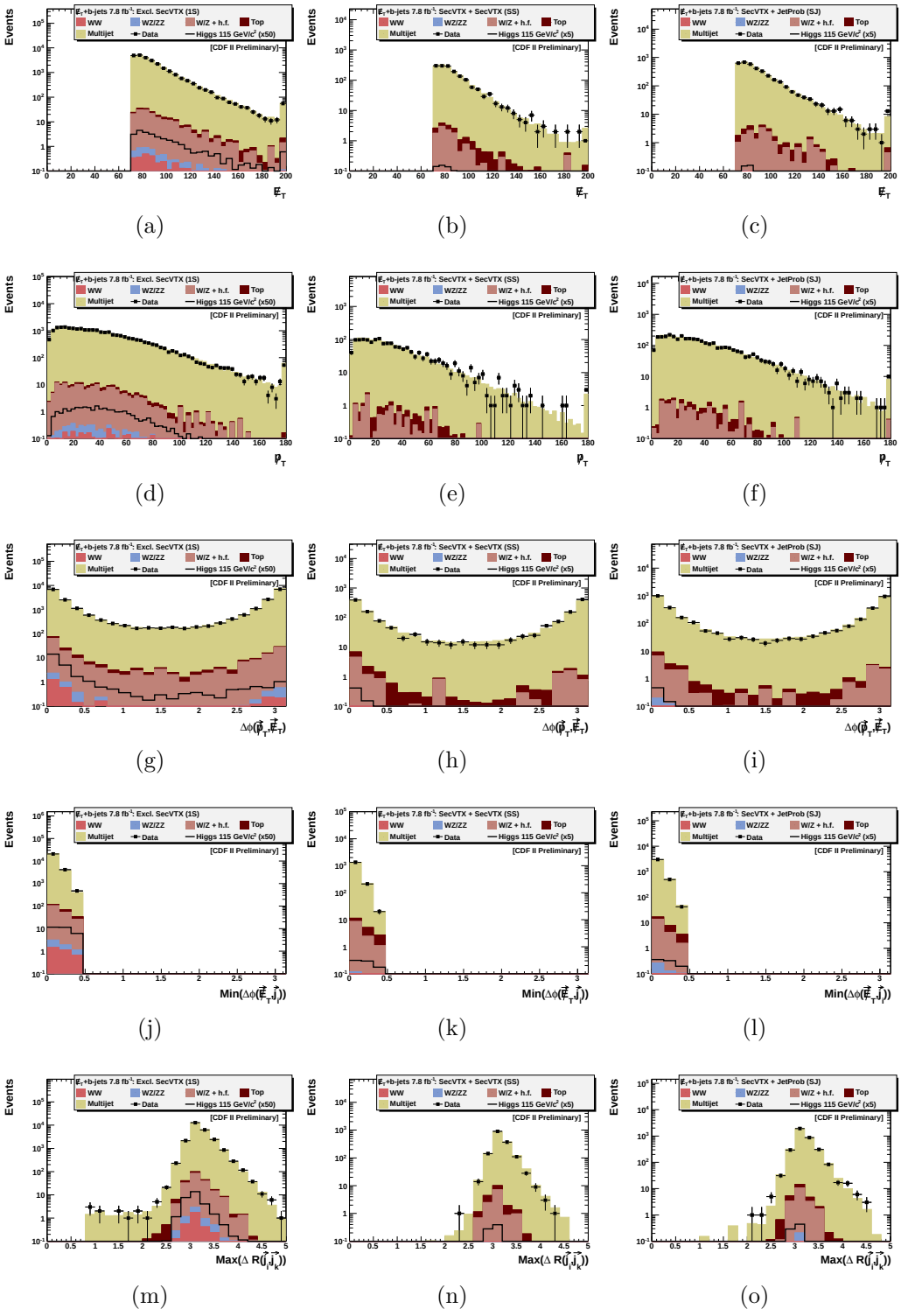


Figure 11.9.: Distribution of several input variables to NN_{QCD} in the QCD CR1 region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data (see text).

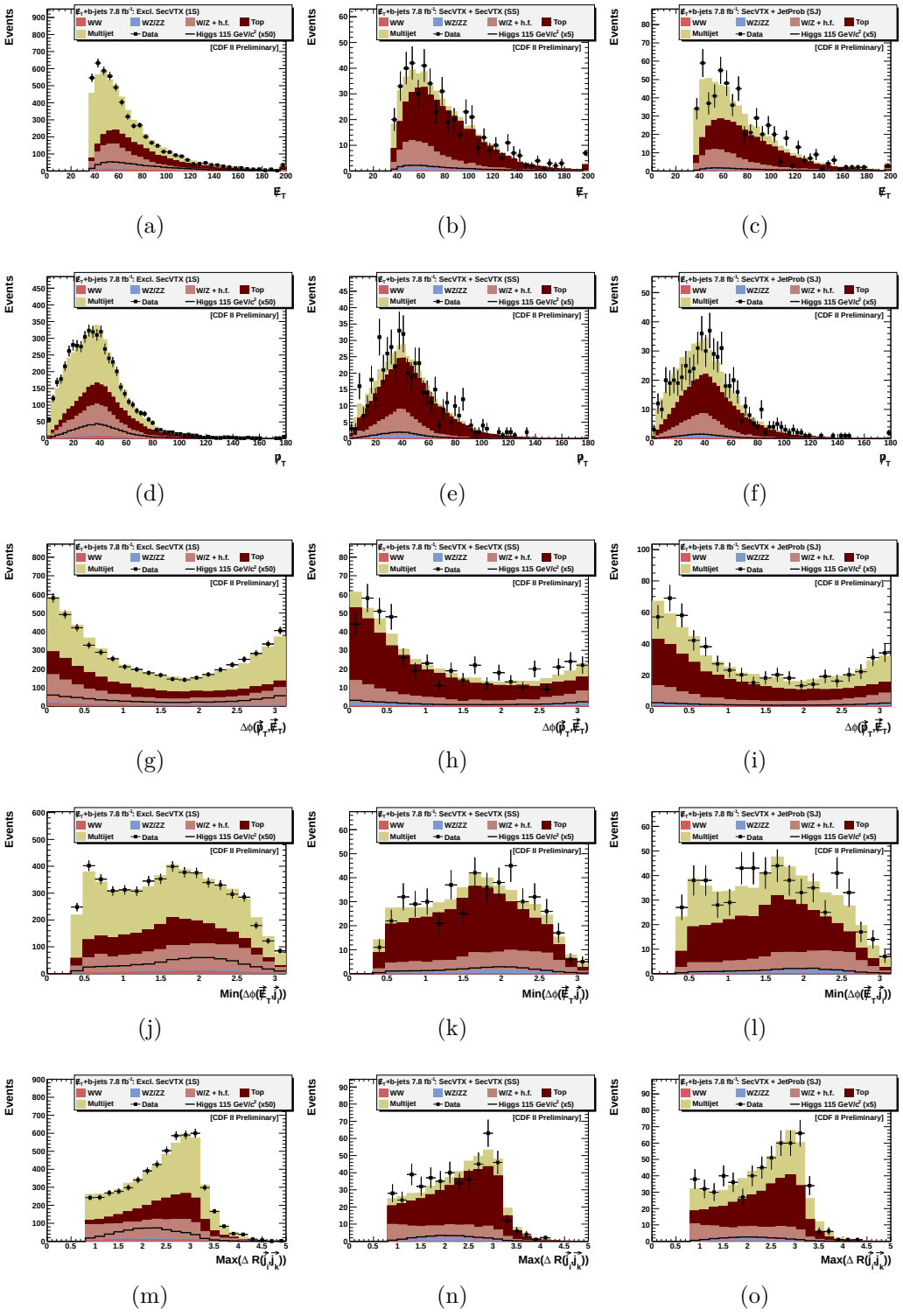


Figure 11.10.: Distribution of several input variables to NN_{QCD} in the EWK region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data (see text).

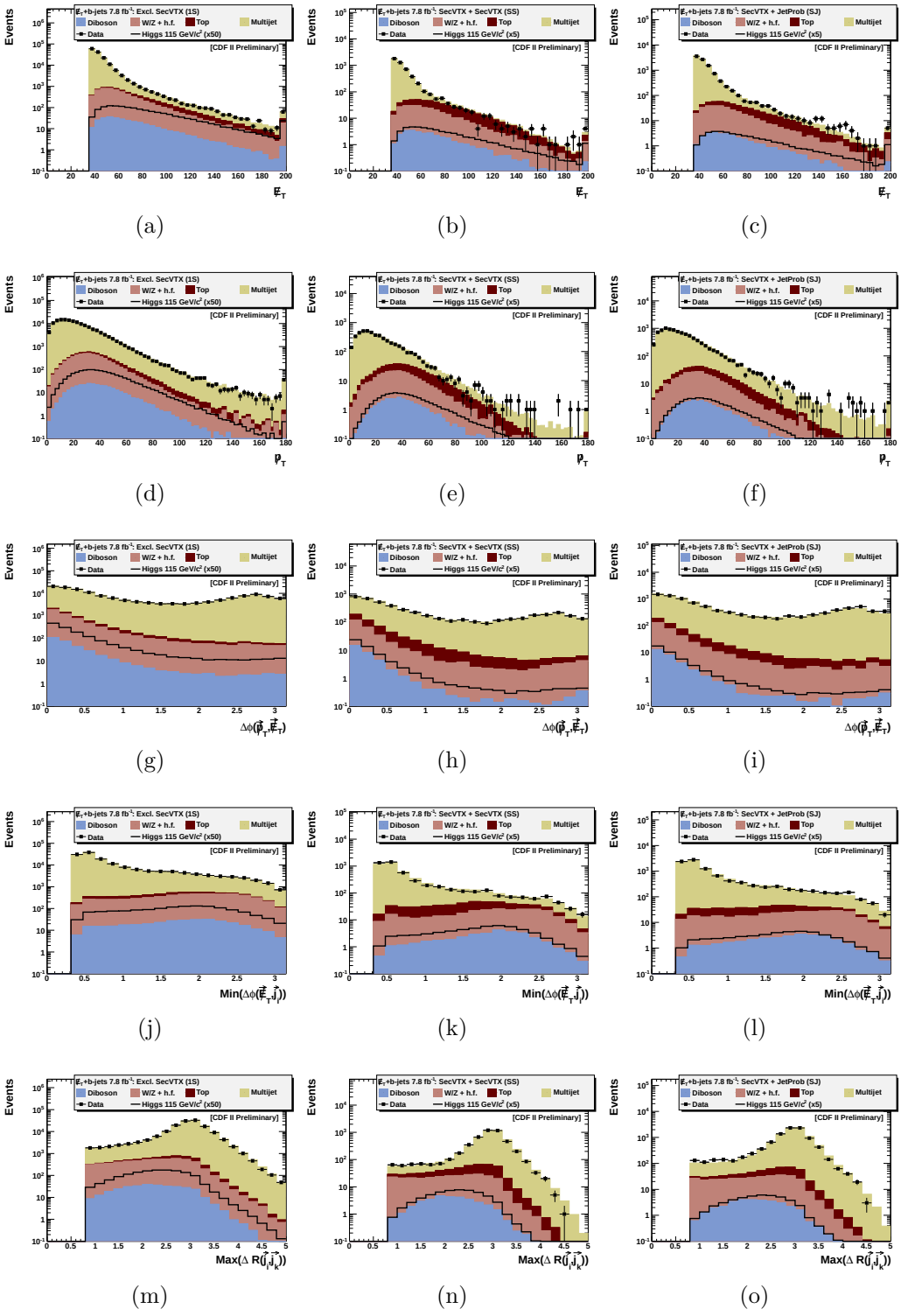


Figure 11.11.: Distribution of several input variables to NN_{QCD} in the pre-selection region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data (see text).

12. STATISTICAL INTERPRETATION

The experimental method is based on the comparison of experimental data to the predictions made by the theories being tested. In our case, this takes the form of a bin-by-bin comparison of predicted distributions of a given quantity (using simulated events, cf. chapter 7) to the same distribution in the recorded data. This chapter describes how this comparison is done. We present the key concepts behind *statistical interpretation* and present the techniques used to estimate the compatibility of the SM predictions with experimental data.

In this dissertation, we perform two kind of statistical analyses. When probing the SM Higgs production (cf. chapter 14), in the absence of an observation, we set an upper limit on the production cross-section times branching fraction at 95% confidence level (C.L.). When validating the analysis techniques by probing known standard model processes (cf. chapter 15), we directly measure the production cross-section, estimate its posterior probability density, and estimate the significance of our result. All these tasks share a common statistical framework: that of *Bayesian inference*. We do not use frequentist approaches unless explicitly mentioned. An thorough overview of confidence limits is presented in ref. [194].

Achieving good separation of signal events from background and reducing the systematic uncertainties on our predictions are key ingredients to improving measurements and significance evaluations. Due to the smallness of the signal we seek to measure, the simultaneous use of the contents of all bins to compare the observations to the predictions is necessary. The uncertainties on signal and background affect the extrapolation of the background fits to the signal regions. The contents of low signal to background bins serve as signal depleted regions to constrain the backgrounds predictions. They are an important aspect of the interpretation of thigh signal to background bins.

These aspects are addressed below. Systematic uncertainties are handled with Bayesian tools, i.e. assigning flat priors to the values of the uncertain nuisance parameters. The impact of the latter on the predictions are evaluated, and integrals computed over the values of the nuisance parameters.

12.1 Bayesian Inference

The agreement between the simulation and the data in a given bin[1] is quantified by the probability to expect μ events given we observe D events, $P(\mu|n)$. This quantity is not readily available, and difficult to estimate. Bayes' theorem [175] allows evaluating this probability assuming prior knowledge; it is expressed as

$$P(\mu|D) = \frac{L(D|\mu)\pi(\mu)}{P(n)} \tag{12.1}$$

where $P(\mu|D)$ is the posterior probability on μ, $L(D|\mu)$ is the probability to measure n events given the expectation μ events given by the Poisson distribution

$$L(D|\mu) = P(D|\mu) = \frac{\mu^D e^{-\mu}}{D!}, \tag{12.2}$$

$\pi(\mu)$ is the prior probability on μ (the effect of statistical and systematic uncertainties), and $P(D)$ is the probability to observe D events, and is given by the condition that the sum of all probabilities should sum to unity. Using equation 12.1,

$$\int P(\mu|D)d\mu = 1 = \int \frac{L(D|\mu)\pi(\mu)}{P(D)}d\mu, \tag{12.3}$$

which leads to

$$P(D) = \int L(D|\mu)\pi(\mu)d\mu. \tag{12.4}$$

[1]A bin is a specific region in variable space, for which we collect the number of expected (μ) and observed (D) events.

In the next section, we specify μ in terms of signal and background contributions, and extract a probability for the number of signal events. From this, we estimate an upper limit on the production cross-section times branching fraction.

12.2 Upper Limit on the Production Cross-section

The number of expected events in a given bin μ consists in the sum of the expected number of signal (s) and background (b) events. Given these, we have

$$P(s+b|D) = \kappa(n,\mu)\frac{(s+b)^D \cdot e^{-(s+b)}}{n!} \tag{12.5}$$

where $\kappa(D,\mu) = \pi(\mu)/P(D)$.

The predictions of the SM have all been confirmed experimentally, except for the SM Higgs, which we search for. Thus, we want to express the probability for the number of signal events s given b and D, i.e. $P(s|D,b)$. Additionally, we want to know how this number relates to the SM prediction s_{SM} and introduce the ratio $\beta = s/s_{\mathrm{SM}}$. Our goal is to determine a 95% C.L. upper limit ℓ on β so that

$$\int_0^{\ell} P(\beta|n, \sum_k b_k, s_{\mathrm{SM}})d\beta = 0.95 \tag{12.6}$$

where we have expanded b in terms of the expected event yields for each background process b_k. Equation (12.6) can be fully expanded into

$$\frac{\int_0^{\ell} \left(\sum_k b_k + \beta \cdot s_{\mathrm{SM}}\right)^D \cdot \exp\left(\sum_k b_k + \beta \cdot s_{\mathrm{SM}}\right)d\beta}{\int_0^{\infty} \left(\sum_k b_k + \beta \cdot s_{\mathrm{SM}}\right)^n \cdot \exp\left(\sum_k b_k + \beta \cdot s_{\mathrm{SM}}\right)d\beta} = 0.95. \tag{12.7}$$

This equation cannot be solved analytically, so we resort to numerical integration. We consider increasing values of β, compute the integral, and stop when the value 0.95 is reached.

In this section, we have presented the basic concept behind the estimation of upper limits. The case of a single bin corresponds to a counting experiment. Additional

sensitivity is provided by splitting a kinematic distribution in many bins, and compute the likelihood L for each bin, assuming that each bin is independent from the others. The same approach is used to include additional search regions (the flavor categories in our case).

The derivation of this section does not take any systematic uncertainties into account. These express our confidence in the simulations, and must be included in the calculations. We include them in the next section.

12.2.1 Systematic Uncertainties

Systematic uncertainties affect the predicted event yields (s and b_k) in two ways: *rate* uncertainties apply to all the bins of a distribution in a correlated way while *shape* uncertainties apply separately to each bin, and thus alter the shape of a distribution. In this section, we describe how these uncertainties affect the calculations of the previous section. We do not discuss the origin of these uncertainties; this is the topic of chapter 13. Shape uncertainties are not discussed here, since for they are a simple rate uncertainty in the case of a single bin; they will be added in the next section.

To each independent source of systematic uncertainty m corresponds a *nuisance parameter* θ_m, which is a coefficient to the number of events for a particular process. The number of expected events can now be expressed as

$$\mu = \sum_k \left(\left(\prod_{m \in \mathcal{S}_k} \theta_m \right) \cdot b_k \right) + \beta \cdot \left(\prod_{m \in \mathcal{S}} \theta_m \right) \cdot s_{\mathrm{SM}} \tag{12.8}$$

where $\mathcal{S}_k$ ($\mathcal{S}$) is the set of systematic uncertainties that affect the kth background process (the signal). Each nuisance parameter is modeled using a truncated Gaussian distribution

$$G(\theta_m | \tilde{\mu}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(\theta_m - \tilde{\mu})^2}{2\sigma^2}. \tag{12.9}$$

The truncation ensures that no unphysical predictions are made (i.e. no negative predictions). A process with a 3% systematic uncertainty will have $\tilde{\mu} = 1.0$ and $\sigma = 0.03$.

The systematic uncertainties are taken into account by convoluting each nuisance parameter in the likelihood function and integrating over the nuisance parameters (cf. statistical section of the Review of Particle Physics [7]). The likelihood now becomes

$$L(D|\beta, \sum_k b_k, s_{\mathrm{SM}}, \sum_m \theta_m) = \int \dots \int \frac{1}{D!}\mu^D \cdot e^{-\mu} \times \prod_m G(\theta_m|1.0, \sigma_m) \cdot \prod_m d\theta_m \quad (12.10)$$

where $\mu = \mu(\beta, b_k, s_{\mathrm{SM}}, \theta_m)$. When the integration over each nuisance parameters is performed, the likelihood L does not depend any more on the nuisance parameters.

The integrals in (12.10) take a lot of computing resources, especially because points with small integrants contribute very little to the total likelihood. Conventional sampling methods for Monte Carlo integration do not sufficiently probe the peak of the nuisance parameters prior distributions $G(\theta_m|1.0, \sigma_m)$. Ideally, one would like to sample the peaks of G. In this dissertation, we resort to Markov Chain Monte Carlo (MCMC) [7] for this purpose (cf. section 12.4).

12.2.2 Multiple Bins and Channels

This section expand upon the previous to include all the bins of the distribution used to compare the simulation to the data. This significantly increases the sensitivity of the analysis, benefitting from different s/b ratios in different bins, i.e. high (low) s/b for high (low) values of the discriminant function.

When multiple bins are considered, shape uncertainties have an important role, because they can alter the predictions across many bins. Shape uncertainties are conveniently viewed as a bin-specific rate uncertainty. The correlation between the values of the uncertainty for different bins of the same distribution in the same region

are implied, but are irrelevant for the determination of the likelihood functions. To each bin i we associate a nuisance parameter η_{ik} for each process k, convolute the parameter into the likelihood function, and integrate over the nuisance parameter. For each bin, the range for the integration over each parameter is determined from distributions expressing the upper and lower fluctuations due to the systematic uncertainty being considered; for each bin, these distributions are compared to the central template (without any systematic uncertainty included).

The total likelihood is obtained assuming that all bins are independent, and is given by

$$L = \prod_i L_i \tag{12.11}$$

where L_i is the likelihood for the ith bin. The case of multiple channels is straightforward: one simply considers the additional bins as if they were from the same distribution. The only difference is that the values for the nuisance parameters may be different for different regions. Usually one performs the estimation of the upper limit in each channel independently, and then performs the combination to extract a more precise upper limit.

12.3 The complete picture

In this section, we describe the complete recipe used to extract the signal cross-section given the data and the backgrounds (with the associated uncertainties) [167].

We use a likelihood function for the extraction of the signal cross-section as well as for the determination of the significance. This function is the product of the Poisson probabilities for each bin of each histogram in every channel. These Poisson probabilities are function of d_i, the number of observed events in each bin, and μ_i, the predictions in each bin; i ranges from 1 to n_{bins}. The likelihood function is given by

$$L = \prod_{i=1}^{n_{\text{bins}}} \frac{\mu_i^{d_i} e^{-\mu_i}}{d_i!}. \tag{12.12}$$

The prediction in each bin is a sum over signal and background contributions:

$$\mu_i = \sum_{k=1}^{n_{\rm bkg}} b_{ik} + \sum_{k=1}^{n_{\rm sig}} s_{ik} \tag{12.13}$$

where b_{ik} (s_{ik}) is the background (signal) prediction in bin i for process k.

Uncertain nuisance parameters θ_m (one for each nuisance parameter) affect the signal and background predictions. The nuisance parameters are applied starting from shape uncertainties, followed by bin-by-bin uncertainties, and overall rate uncertainties. Shape and rate effects due to a given nuisance parameter are treated as 100% correlated. The overall rate uncertainties do not include the statistical uncertainties, which are taken into account by bin-by-bin uncertainties, which are independent from each other and from any other source of systematic uncertainty.

The likelihood is then a function of the observed data $\mathbf{D} = \{d_i\}$, the signal scale factors $\boldsymbol{\beta}$ (one for each signal process), the nuisance parameters $\boldsymbol{\theta} = \{\theta_m\}$ and $\boldsymbol{\eta} = \{\eta_{s,ik}, \eta_{b,ik}\}$, as well as the central values of the signal $\boldsymbol{s} = \{s^0_{ik}\}$ and background $\boldsymbol{b} = \{b^0_{ik}\}$ predictions, and the rate $\boldsymbol{\rho} = \{\rho^{m\pm}_{b,ik}, \rho^{m\pm}_{s,ik}\}$, shape $\boldsymbol{\kappa} = \{\kappa^{m\pm}_{b,ik}, \kappa^{m\pm}_{s,ik}\}$, and bin-by-bin $\boldsymbol{\delta} = \{\delta^0_{b,ik}, \delta^{m\pm}_{b,ik}, \delta^0_{s,ik}, \delta^{m\pm}_{s,ik}\}$ uncertainties:

$$L = L(\mathbf{D}|\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{s}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\kappa}, \boldsymbol{\delta}). \tag{12.14}$$

We use the McLimit software package [195,196] to numerically integrate (12.14) over the nuisance parameters

$$L(\mathbf{D}|\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{s}, \boldsymbol{b}) = \int \ldots \int L(\mathbf{D}|\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{s}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\kappa}, \boldsymbol{\delta}) d\boldsymbol{\rho} d\boldsymbol{\kappa} d\boldsymbol{\delta} \tag{12.15}$$

and to determine the 95% C.L. upper limit ℓ satisfying

$$\frac{\int_0^\ell L(\mathbf{D}|\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{s}, \boldsymbol{b}) d\boldsymbol{\beta}}{\int_0^\infty L(\mathbf{D}|\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{s}, \boldsymbol{b}) d\boldsymbol{\beta}} = 0.95. \tag{12.16}$$

The contributions to a bin's prediction from a source of background are modified by shape uncertainty parameters θ_m by linearly interpolating and extrapolating the difference between the central prediction b^0_{ik} and the prediction in a histogram corresponding to a $+1\sigma$ variation $\kappa^{m+}_{b,ik}$ if $\theta_m > 0$, and performing a similar operation using a -1σ varied histogram if $\theta_m < 0$.

$$b^1_{ik} = b^0_{ik} + \sum_{m=1}^{n_n} \begin{cases} (\kappa^{m+}_{b,ik} - b^0_{ik})\theta_m & : \quad \theta_m \geq 0 \\ (b^0_{ik} - \kappa^{m-}_{b,ik})\theta_m & : \quad \theta_m < 0 \end{cases} . \tag{12.17}$$

An identical prescription is applied to the standard model signal predictions s^0_{ik}, summed over the same nuisance parameters, using alternate shape histograms $\kappa^{m+}_{s,ik}$ and $\kappa^{m-}_{s,ik}$. Thus, a nuisance parameter may affect both the signal and background predictions in a correlated way. The application of shape uncertainties is not allowed to produce a negative prediction in any bin for any source of background or signal:

$$b^2_{ik} = \max(0, b^1_{ik}), \tag{12.18}$$

and similarly for the signal predictions. Each template histogram, the central prediction and the shape variations, has a bin-by-bin uncertainty in each bin. These bin-by-bin uncertainties are linearly interpolated in each bin in exactly the same way as the predicted values. This procedure works well when the shape-variation templates share all or most of the same events, but overestimates the bin-by-bin uncertainties when the alternate shape templates are filled with independent Monte Carlo samples. If the bin-by-bin uncertainty on b^0_{ik} is $\delta^0_{b,ik}$, and the bin-by-bin uncertainty on $b^{m\pm}_{ik}$ is $\delta^{m\pm}_{b,ik}$, then

$$\delta^1_{b,ik} = \delta^0_{b,ik} + \sum_{m=1}^{n_n} \begin{cases} (\delta^{m+}_{b,ik} - \delta^0_{b,ik})\theta_m & : \quad \theta_m \geq 0 \\ (\delta^0_{b,ik} - \delta^{m-}_{b,ik})\theta_m & : \quad \theta_m < 0 \end{cases} . \tag{12.19}$$

A similar operation is applied to interpolate and extrapolate the bin-by-bin uncertainties on the signal $\delta^1_{s,ik}$. Bin-by-bin uncertainties are also not allowed to be negative:

$$\delta^2_{b,ik} = \max(0, \delta^1_{b,ik}), \tag{12.20}$$

and similarly for the signal. Each bin of each background has a nuisance parameter $\eta_{b,ik}$ associated with it.

$$b^3_{ik} = b^2_{ik} + \delta^2_{b,ik}\eta_{b,ik}, \tag{12.21}$$

where $\eta_{b,ik}$ is drawn from a Gaussian centered on zero with unit width when integrating over it. If $b^3_{ik} < 0$, then $\eta_{b,ik}$ is re-drawn from that Gaussian. Similarly for the signal,

$$s^3_{ik} = s^2_{ik} + \delta^2_{s,ik}\eta_{s,ik}. \tag{12.22}$$

Finally, rate uncertainties are applied multiplicatively. If the fractional uncertainty on b^0_{ik} due to nuisance parameter m is $\rho^{m+}_{b,ik}$ for a $+1\sigma$ variation and it is $\rho^{m-}_{b,ik}$ for a negative variation, then a quadratic function is determined to make a smooth application of the nuisance parameter to the predicted value:

$$b_{ik} = b^3_{ik} \prod_{m=1}^{n_\mathrm{n}} \left(1 + \frac{\rho^{m+}_{b,ik} + \rho^{m-}_{b,ik}}{2}\theta^2_m + \frac{\rho^{m+}_{b,ik} + \rho^{m-}_{b,ik}}{2}\theta_m \right), \tag{12.23}$$

and similarly for the signal, where the product is over the same nuisance parameters as the background.

$$s^4_{ik} = s^3_{ik} \prod_{m=1}^{n_\mathrm{n}} \left(1 + \frac{\rho^{m+}_{s,ik} + \rho^{m-}_{s,ik}}{2}\theta^2_m + \frac{\rho^{m+}_{s,ik} + \rho^{m-}_{s,ik}}{2}\theta_m \right). \tag{12.24}$$

The rate uncertainties are applied multiplicatively because most of them affect the rates by scale factors, such as luminosity uncertainty, or acceptance uncertainties, which also multiply the Monte Carlo statistical uncertainties and hence are applied at the end. Multiple shape uncertainties are treated additively because most of them correspond to events migrating from one bin to another.

The signal predictions $s_{ik}^4 \beta_k$ are based on their standard model rates. These are scaled to test other values of the single top production cross-sections:

$$s_{ik} = s_{ik}^4 \beta_k \tag{12.25}$$

where β_k scales the kth signal.

12.4 Markov Chain Integration

Integrating (12.10) is computationally very expensive because it requires sampling a nuisance parameter space of large dimensions. Conventional MC integration methods uniformly sample this space and evaluate the integrand at each point. The problem with such techniques is that most integration points contribute very little to the likelihood.

Markov Chain Monte Carlo (MCMC) methods sample the nuisance parameter space following a *Markov chain*, which spends more time around points which contribute to the integral, but nonetheless samples enough of the parameter space. Following the Metropolis-Hastings algorithm [197, 198], we start with a random point and select the next point using a Gaussian proposal function. We then compare the probability densities for both points: when $p(\text{new})/p(\text{old}) > 1$, we move to the new point; if not, we move to the new point with the probability $\Pi = p(\text{new})/p(\text{old})$.

The length of the chain is an important parameter of the technique. Long chains are needed to reach the stationary distribution. In practice, we compute the likelihood using different seeds for the random number generator and increasing the chain length. We stop when the output does not change when using different seeds. We typically need chain lengths of 500,000 to 1,000,000 steps.

12.5 Expected and Observed Limits

Before using the data to set an upper limit to the signal production cross-section, we evaluate the sensitivity of our analysis to the signal. The latter is a blind analysis because it does not carry any information from the data in the signal region[2].

We simulate a random *pseudo-experiment* by drawing the *pseudo-data* for each bin from a Poisson probability density function with parameter $\hat{\mu}$, which is determined by allowing fluctuations of the background yields within one standard deviation (we assume there is no signal at all). The value of D is different for each pseudo-experiment. We then obtain an upper limit as previously described.

The median expected limit is the median of the distribution of upper limits from pseudo-experiments; it characterizes the sensitivity of an analysis. The uncertainty on the limit corresponds to boundary of a 68% interval of the distribution of upper limits. We draw sufficient pseudo-experiments to ensure enough statistics, typically about five to ten thousand. Additionally, we compute the upper limit for the *Asimov dataset* [199] which corresponds to the median background expectation and is exactly the outcome of the simulation. This limit should be very close to the median expected limit. The observed limit, i.e. the result quoted for an analysis, is obtained from the data (one experiment).

Figure 12.1 shows the distribution of the upper 95% C.L. limit for 5,000 pseudo-experiments for a Higgs mass of 115 GeV/c^2 (more details are given in chapter 14). The black curve is a fit to the density function $d(\rho) = p_1(\rho - p_2)^{p_3} \exp^{-p_4\rho}$. We also show the median expected, Asimov and observed upper limits. The bands around the median represent the smallest 68% and 95% intervals for the distribution of pseudo-experiments.

[2]Data-driven models contain carry information obtained from the data, but these are from control regions that are orthogonal to the signal region, where the computations are performed.

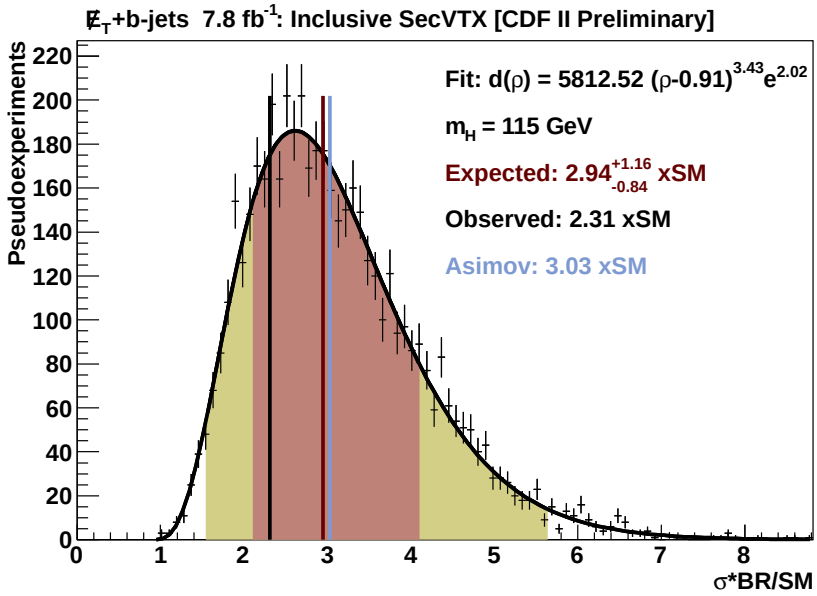


Figure 12.1.: Distribution of the upper 95% C.L. limit for 5,000 pseudo-experiments for $m_H = 115$ GeV/c^2. The black curve is a fit to the density function $d(\rho)$. We also show the median expected (dark red), Asimov (blue) and observed (black) upper limits. The bands around the median represent the smallest 68% (red) and 95% (gold) intervals for the distribution of pseudo-experiments.

12.6 Cross-section Measurement

The measurement of the cross-section uses the same likelihood function as limit setting, but a different quantity is extracted.

Starting from

$$L'(\beta) = \int L(\mathbf{D}|\beta, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{s}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\kappa}, \boldsymbol{\delta})\pi(\boldsymbol{\theta})\pi(\boldsymbol{\eta})d\boldsymbol{\theta}d\boldsymbol{\eta}, \tag{12.26}$$

where the π functions are the Bayesian priors assigned to each nuisance parameter. The L' notation is condensed and not all parameters are present. The measured cross-section corresponds to the maximum of L', which occurs at $\beta^{\max}$:

$$\sigma_{\text{meas}} = \beta^{\max} \cdot \sigma_{\text{SM}}. \tag{12.27}$$

The asymmetric uncertainties quoted correspond to the shortest interval containing 68% of the integral of the posterior, assuming a uniform prior in β, $\pi(\beta) = 1$:

$$0.68 = \frac{\int_{\beta_{\text{low}}}^{\beta_{\text{high}}} L'(\beta)\pi(\beta)d\beta}{\int_0^{\infty} L'(\beta)\pi(\beta)d\beta}.$$

12.7 Significance Calculation

In addition to measuring cross-sections or setting limits, it sometimes important to establish evidence of or observation of a given process. To do so, we use the p-value, which is the probability of observing an outcome of our experiment at least as signal-like as the one observed, or more, assuming that a signal is absent. By convention, an observed p-value of less than 1.35×10^{-3} constitutes evidence for a signal, and an observed p-value less than 2.87×10^{-7} constitutes a discovery. These are the one-sided integrals of the tails of a unit Gaussian distribution beyond $+3\sigma$ and $+5\sigma$, respectively.

The experimental outcomes are ranked on a one-dimentional scale using the likelihood ratio

$$-2\ln Q = -2\ln\frac{L(\mathbf{D}|\beta,\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\eta}},\boldsymbol{s}=\mathbf{s}^{\mathrm{SM}},\boldsymbol{b},\boldsymbol{\rho},\boldsymbol{\kappa},\boldsymbol{\delta})}{L(\mathbf{D}|\beta,\hat{\hat{\boldsymbol{\theta}}},\hat{\hat{\boldsymbol{\eta}}},\boldsymbol{s}=0,\boldsymbol{b},\boldsymbol{\rho},\boldsymbol{\kappa},\boldsymbol{\delta})} \tag{12.28}$$

where $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$ are the best-fit values of the nuisance parameters which maximize L given the data $\mathbf{D}$, assuming presence of single top signal with a standard model rate, and $\hat{\hat{\boldsymbol{\theta}}}$ and $\hat{\hat{\boldsymbol{\eta}}}$ are the best-fit values of the nuisance parameters which maximize L assuming that no single top signal. However, fits are computationally cumbersome and are not needed for a correct incorporation of systematic uncertainties but optimize the sensitivity. Fitting to nuisance parameters for which the discriminant distributions provide little or no constraint do not appreciably improve the sensitivity of the search, and are not performed. Only the most important nuisance parameters, W/Z + heavy flavor jets and multi-jet, are fit for.

The desired p-value p is then

$$p = p(-2\ln Q \leq -2\ln Q_{\mathrm{obs}}|\mathbf{s}=0). \tag{12.29}$$

Systematic uncertainties are included in the expected distributions of $-2\ln Q$ assuming $\mathbf{s}=0$ or $\mathbf{s}=\mathbf{s}_{\mathrm{SM}}$ in a Bayesian way, averaging the distributions of $-2\ln Q$ over variations of the nuisance parameters, weighted by their priors.

To do this in practice, we use pseudoexperiments, which consist of a simulation of the signal and background processes, which are then mixed together and are randomly fluctuated according to the distribution of the uncertainties. In practice, each source of uncertainty is assigned a value, whose effect is then incorporated in the predictions. The latter are combined, providing a distribution close to the real data. Many pseudoexperiments are require in order to investigate subtle effects in the measurement. Each pseudoexperiment passes through the same machinery as the real data would have been through, yielding one outcome.

Thus, in order to include systematic uncertainties, histograms of $-2\ln Q$ are filled with the results of simulated pseudoexperiments. Each of these pseudoexperiments

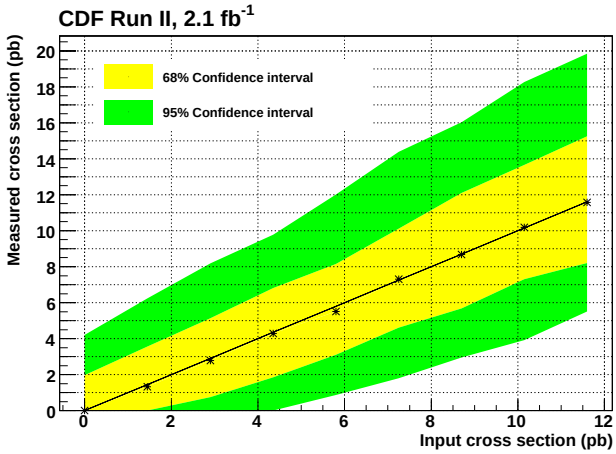


Figure 12.2.: Distributions of 68% and 95% of extracted cross-sections centered on the medians are shown as a function of the input single top cross-section. The fit technique used does not introduce bias.

is drawn from predicted distributions varying the nuisance parameters drawn from prior distributions. Fitting to the main nuisance parameters prevents fluctuations of $-2 \ln Q$ and optimizes the sensitivity in the presence of uncertainty.

In the computation of the p-value, we include all sources of systematic uncertainties, including the theoretical uncertainty in the signal cross-section.

12.8 Sensitivity and Linearity

The observed cross-section and p-value depend on the true cross-sections but also the random outcome of the data. The performance of our techniques is evaluated with the expected distribution of outcomes, assuming a signal is present (sensitivity). For the cross-section measurement, the sensitivity is given by the median expected total uncertainty on the cross-section. The sensitivity of the significance determination is given by the median expected significance. These sensitivities are computed from distributions of Monte Carlo pseudoexperiments with all the nuisance parameters fluctuated according to their priors. The choice and optimization of the analysis method was done without using the observed data, but working with expected sensitivities.

We check our cross-section fit method using Monte Carlo pseudoexperiments generated assuming several signal scalings and varied systematic uncertainties. These are fit for the cross-section.

The procedure used cannot produce a negative cross-section measurement, since the priors are zero for negative values. For an input cross-section of zero, half of the measured cross-sections then are exactly zero, and the other half form a distribution of positive fit cross-sections. The median fit cross-section is used to avoid average effects on our linearity check. Distributions of 68% and 95% of extracted cross-sections centered on the medians are then shown as a function of the input cross-section in figure 12.2. The fit technique used here does not introduce bias.

12.9 Summary

In this chapter, we have presented the statistical tools used to interpret the data, and to measure the signal production cross-section (or an upper limit thereof). We use these techniques in the subsequent chapters to set an upper limit on the Higgs production cross-section, and to validate the analysis technique by measuring the production cross-section of rare ($\sim$pb) SM processes.

13. SYSTEMATIC UNCERTAINTIES

To address the fact that our predictions for various observables differ from their "true" values, we assign systematic uncertainties and propagate their effects on the measured cross-sections and on the significance of the signal.

13.1 Types of systematic uncertainties

Systematic uncertainties are grouped by their source. A given source of uncertainty may affect several templates (background and/or signal). The various systematic uncertainties related to a same source are considered 100% correlated. There are three categories of systematic uncertainties: (a) rate uncertainties, which are related to the predicted production rates of the various signal and background processes; (b) shape uncertainties, which express differences in the distributions due to a given systematic source, and (c) bin-by-bin uncertainties, which are statistical fluctuations

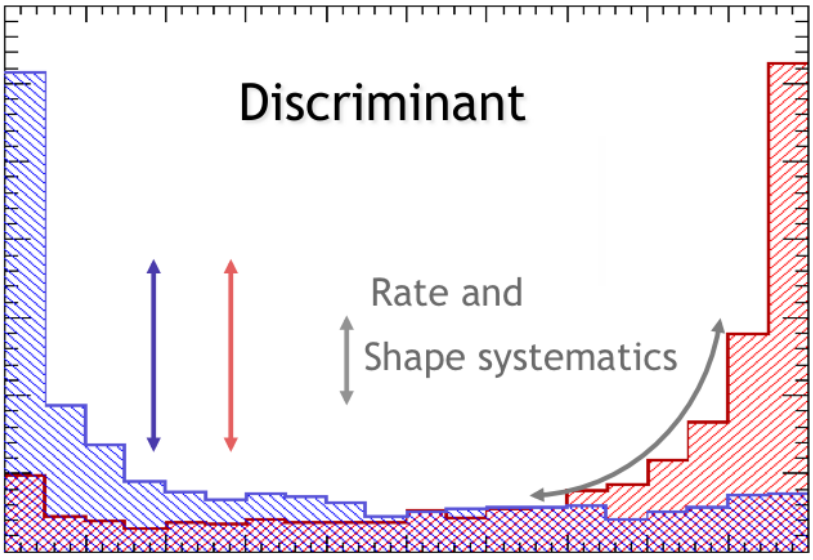


Figure 13.1.: Schematic view of shape and rate systematic uncertainties.

arising from the limited size of the Monte Carlo or data samples (where applicable). Figure 13.1 shows the effect of rate and shape systematics on the distribution of a discriminant variable.

13.2 Sources of systematic uncertainty

Some sources of systematic uncertainty affect both rates and shapes. All rate uncertainties are assigned a truncated Gaussian prior, preventing negative predictions (cf. chapter 12. For shape uncertainties, if the prediction for a given bin is negative, it is set to zero. To reduce the impact of limited MC statistics, the shape uncertainties are "median smoothed", using a window of five bins.

We now describe the individual sources of uncertainty:

- **Theoretical cross-sections** The Monte Carlo (MC) predictions for the signal and background processes are made using various event generators for evaluating the tree-level diagrams and PYTHIA for the parton shower. These simulations predict the kinematic shapes well (cf. section 11.3). The total cross-sections at LO or NLO are scaled to NLO or better models, and we propagate the associated uncertainties. We use a $\pm 10\%$ uncertainty for top pair production [200], a $\pm$ 30% uncertainty for the W and Z background processes [201], and $\pm$ 6% for the diboson prediction [202].

- **Integrated Luminosity** An uncertainty of $\pm 6\%$ is applied to all the predictions based on Monte Carlo. This uncertainty accounts for the uncertainty in the $p\bar{p}$ inelastic cross-section and for the uncertainty in the acceptance of the luminosity monitor of CDF [89].

- **Trigger Efficiency** Since we are using data below the full efficient region of the trigger, we apply a parameterization of the trigger efficiency to the MC simulated samples. We assign a systematic uncertainty by considering an alternate shape (JET-50 parameterization) to vary the central one (MUON parameteriza-

tion); both are described in chapter 10. The overall uncertainty associated with the parameterization is of the order of 2-4%, depending on the process.

- **b-tagging efficiency** The b-tagging efficiency affects the predicted rates of MC simulated processes. Known differences between the data and the simulation are corrected for by scaling the simulation, and uncertainties on these scale factors are collected together in one source of uncertainty (they affect the predictions in the same way). We assign an uncertainty between 5.2% and 10.4%, depending on the heavy flavor category.

- **Lepton Veto** The uncertainty in the efficiency of the cuts used to veto leptons was determined to be less than 2%.

- **Initial and Final State Radiation (ISR/FSR)** To evaluate the ISR/FSR, PYTHIA uses a model of "backwards evolution" [111,203,204]. New Monte Carlo samples are generated for single top signals and $t\bar{t}$ with Λ_{QCD} doubled (more ISR) or divided in half (less ISR) and with the initial transverse momentum scale and the hard scattering of the shower both multiplied (more ISR) or divided (less ISR) by four. The model used by PYTHIA for gluon radiation from partons emitted from the hard-scattering interaction has been tuned with LEP data [111]. The parameters for the final-state showering are also adjusted in PYTHIA, except for the hard-scattering scale. The uncertainties are then computed comparing the ISR/FSR templates to the central ones. The effects of variations in ISR and FSR are only applied to the signal, and are treated as 100% correlated with each other. Their effect is of the order of 2-3%.

- **Jet Energy Scale (JES)** Each step in the calibration of the calorimeter response involves an uncertainty (cf. section 5.6.2), which is propagated to the final JES. The raw jet energies are corrected for test-beam scales, non-uniformity of the detector, multiple interactions and energy outside of the jet cone which is not assigned to the jet [108]. Since the influence of JES variation is different

for each background component, the uncertainties on JES are incorporated by varying the jet energy scale in all Monte Carlo samples twice: one upwards, one downwards. JES uncertainties can also modify the shapes of the distributions, which we take into account.

- **Parton Distribution Functions (PDF)** Lack of precise knowledge of the PDFs is a source of theoretical uncertainty for the amount of signal produced. The uncertainty is estimated using different sets of PDF eigenvectors. The central PDF's used in this analysis are the CTEQ5L set [41]. The uncertainty is determined comparing:
  - two PDF parameterization sets, CTEQ5L (NLO) and MRST72 [43] (LO);
  - MRST72 and MRST75 with different Λ_{QCD};
  - each of the 20 signed eigenvectors of CTEQ6M with the central PDF.

  The total PDF uncertainty is obtained adding the larger of the 20 eigenvectors' uncertainty or the MRST72 and CTEQ5L PDF sets uncertainty in quadrature with the Λ_{QCD} uncertainty. A 2% uncertainty was found to be sufficient for all the backgrounds. The signal's uncertainties were ranging from 1% to 2%. Shape variations of PDF were not considered for the Higgs signal.

- **QCD multijet model** The data-driven model for multijet production predicts the shapes of the distributions. The rates are obtained from a control region (the low end of the NN_{QCD} distribution). In this region, we scale the multijet production to the difference between the rates predicted by data and the MC (cf. section 8.7.5). We then obtain the uncertainty on this scaling using propagation of errors. The variations in the tag rate probability parameterization used to estimate the multijet background also modify the shapes of the distributions. The shape uncertainty is obtained by varying the tag rate probability in each bin of the matrix.

We also take into account the normalization uncertainty on the processes which are part of the background in the region from which we get the QCD normalization. Those uncertainties are anti-correlated with respect to the normalization on these processes, and weighted appropriately. In the $NN_{QCD} < 0.1$ control sample used to derive the normalization, the biggest contamination sources come from W + heavy flavor jets (5% of the 1S and $< 1.5\%$ of the SS and SJ samples), $t\bar{t}$ (2.7% of the SS sample and $< 1.5\%$ of the two other samples), and Z + heavy flavor jets ($< 1.8\%$ of the three tagging samples). Diboson contamination is negligible ($< 0.5\%$).

To avoid double counting non QCD multijet events in our estimation of the QCD multijet background, we need to apply the tag rate parameterization to the MC predictions and subtract the output from the data. When measuring known SM processes to validate the analysis technique (cf. chapter 15), this subtraction also includes removing the signal MC. When the signal contribution to this subtraction isn't negligible, we associate a shape systematic uncertainty to this removal by varying the amount of signal we subtract by about three times the theoretical uncertainty on its production cross-section. This is not done for the Higgs.

13.3 Summary

In this chapter, we discussed the sources of systematic uncertainty that we include in the analysis. We have described how we estimate them, and shown how they affect our predictions. We are now ready to use the techniques presented earlier to search for the SM Higgs boson (next chapter) and other SM processes (cf. chapter 15).

Table 13.1: Systematic uncertainties

	$\mathcal{L}$ (CDF)	$\mathcal{L}$ (Tevatron)	$\sigma(W/Z+h.f.)$	$\sigma(WH/ZH)$	$\sigma(WW/WZ/ZZ)$	$\sigma(t)$	B-Tagging	ISR/FSR	Jet Energy Scale	Lepton ID	Lepton Veto	Mistags	PDF	Tag Rate	Trigger Efficiency
Excl. SecVtx															
$ZH \to \nu\nu b\bar{b}$	4.4/-4.4	3.8/-3.8	-	5/-5	-	-	5.2/-5.2	-2/2	**2.6/-2.6**	-	2/-2	-	3/-3	-	**-2.3/0**
$WH \to \ell\nu b\bar{b}$	4.4/-4.4	3.8/-3.8	-	5/-5	-	-	5.2/-5.2	-2/2	**3.3/-3.1**	2/-2	-	-	3/-3	-	**-3.4/0**
$ZH \to \ell\ell b\bar{b}$	4.4/-4.4	3.8/-3.8	-	5/-5	-	-	5.2/-5.2	-2/2	**3.7/-3.8**	-	-	-	3/-3	-	**-3.8/0**
Single top (s)	4.4/-4.4	3.8/-3.8	-	-	-	10/-10	5.2/-5.2	-	**2.7/-2.8**	2/-2	-	-	3/-3	-	**-3.5/0**
Single top (t)	4.4/-4.4	3.8/-3.8	-	-	-	10/-10	5.2/-5.2	-	**2.8/-2.8**	2/-2	-	-	3/-3	-	**-3.7/0**
Top pair	4.4/-4.4	3.8/-3.8	-	-	-	10/-10	5.2/-5.2	-	**-0.8/0.6**	2/-2	-	-	3/-3	-	**-1.9/0**
WW	4.4/-4.4	3.8/-3.8	-	-	6/-6	-	5.2/-5.2	-	**4.9/-5.4**	2/-2	-	-	3/-3	-	**-4.1/0**
WZ	4.4/-4.4	3.8/-3.8	-	-	6/-6	-	5.2/-5.2	-	**5.3/-5.2**	2/-2	-	-	3/-3	-	**-4/0**
ZZ	4.4/-4.4	3.8/-3.8	-	-	6/-6	-	5.2/-5.2	-	**5/-4.8**	2/-2	-	-	3/-3	-	**-3.2/0**
$W + h.f.$	4.4/-4.4	3.8/-3.8	30/-30	-	-	-	5.2/-5.2	-	**8.2/-6.8**	2/-2	-	-	3/-3	-	**-5.2/0**
$Z + h.f.$	4.4/-4.4	3.8/-3.8	30/-30	-	-	-	5.2/-5.2	-	**10.8/-3.4**	2/-2	-	-	3/-3	-	**-3.7/0**
Multijet	-	-	-	-	-	-	-	-	-	-	-	-	-	0.7/-0.7	**3.2/0**
Mistags	-	-	-	-	-	-	-	-	-	-	-	**17.9/-17.4**	-	-	-
SecVtx + JetProb															
$ZH \to \nu\nu b\bar{b}$	4.4/-4.4	3.8/-3.8	-	5/-5	-	-	8.3/-8.3	-2.4/2.4	**1.9/-1.9**	-	2/-2	-	3/-3	-	**-1.7/0**
$WH \to \ell\nu b\bar{b}$	4.4/-4.4	3.8/-3.8	-	5/-5	-	-	8.3/-8.3	-2.4/2.4	**2.4/-2.4**	2/-2	-	-	3/-3	-	**-2.6/0**
$ZH \to \ell\ell b\bar{b}$	4.4/-4.4	3.8/-3.8	-	5/-5	-	-	8.3/-8.3	-2.4/2.4	**2.9/-3.4**	-	-	-	3/-3	-	**-3.1/0**
Single top (s)	4.4/-4.4	3.8/-3.8	-	-	-	10/-10	8.3/-8.3	-	**2.2/-2.3**	2/-2	-	-	3/-3	-	**-2.9/0**
Single top (t)	4.4/-4.4	3.8/-3.8	-	-	-	10/-10	8.3/-8.3	-	**3.9/-3.4**	2/-2	-	-	3/-3	-	**-3.9/0**
Top pair	4.4/-4.4	3.8/-3.8	-	-	-	10/-10	8.3/-8.3	-	**-0.6/0.2**	2/-2	-	-	3/-3	-	**-1.8/0**
WW	4.4/-4.4	3.8/-3.8	-	-	6/-6	-	8.3/-8.3	-	**0.05/-3.9**	2/-2	-	-	3/-3	-	**-5.4/0**
WZ	4.4/-4.4	3.8/-3.8	-	-	6/-6	-	8.3/-8.3	-	**4.7/-4.9**	2/-2	-	-	3/-3	-	**-3.4/0**
ZZ	4.4/-4.4	3.8/-3.8	-	-	6/-6	-	8.3/-8.3	-	**3.8/-3.9**	2/-2	-	-	3/-3	-	**-2.2/0**
$W + h.f.$	4.4/-4.4	3.8/-3.8	30/-30	-	-	-	8.3/-8.3	-	**6.8/-5.9**	2/-2	-	-	3/-3	-	**-4.4/0**
$Z + h.f.$	4.4/-4.4	3.8/-3.8	30/-30	-	-	-	8.3/-8.3	-	**8.3/-3.1**	2/-2	-	-	3/-3	-	**-2.8/0**
Multijet	-	-	-	-	-	-	-	-	-	-	-	-	-	1.6/-1.6	**1.6/0**
Mistags	-	-	-	-	-	-	-	-	-	-	-	**65.2/-38.5**	-	-	-
SecVtx + SecVtx															
$ZH \to \nu\nu b\bar{b}$	4.4/-4.4	3.8/-3.8	-	5/-5	-	-	10.4/-10.4	-3/3	**1.7/-1.8**	-	2/-2	-	3/-3	-	**-1.8/0**
$WH \to \ell\nu b\bar{b}$	4.4/-4.4	3.8/-3.8	-	5/-5	-	-	10.4/-10.4	-3/3	**2.4/-2.3**	2/-2	-	-	3/-3	-	**-2.7/0**
$ZH \to \ell\ell b\bar{b}$	4.4/-4.4	3.8/-3.8	-	5/-5	-	-	10.4/-10.4	-3/3	**3.2/-3.1**	-	-	-	3/-3	-	**-3.3/0**
Single top (s)	4.4/-4.4	3.8/-3.8	-	-	-	10/-10	10.4/-10.4	-	**2/-2.2**	2/-2	-	-	3/-3	-	**-2.9/0**
Single top (t)	4.4/-4.4	3.8/-3.8	-	-	-	10/-10	10.4/-10.4	-	**3/-2.8**	2/-2	-	-	3/-3	-	**-3.6/0**
Top pair	4.4/-4.4	3.8/-3.8	-	-	-	10/-10	10.4/-10.4	-	**0/0.02**	2/-2	-	-	3/-3	-	**-1.8/0**
WW	4.4/-4.4	3.8/-3.8	-	-	6/-6	-	10.4/-10.4	-	**4.2/6.1**	2/-2	-	-	3/-3	-	**-5.9/0**
WZ	4.4/-4.4	3.8/-3.8	-	-	6/-6	-	10.4/-10.4	-	**3.7/-4.3**	2/-2	-	-	3/-3	-	**-3.2/0**
ZZ	4.4/-4.4	3.8/-3.8	-	-	6/-6	-	10.4/-10.4	-	**4/-4.1**	2/-2	-	-	3/-3	-	**-2.3/0**
$W + h.f.$	4.4/-4.4	3.8/-3.8	30/-30	-	-	-	10.4/-10.4	-	**4.3/-4.6**	2/-2	-	-	3/-3	-	**-3.6/0**
$Z + h.f.$	4.4/-4.4	3.8/-3.8	30/-30	-	-	-	10.4/-10.4	-	**8.8/-3.2**	2/-2	-	-	3/-3	-	**-2.3/0**
Multijet	-	-	-	-	-	-	-	-	-	-	-	-	-	2.5/-2.5	**1.9/0**
Mistags	-	-	-	-	-	-	-	-	-	-	-	**36.7/-30**	-	-	-

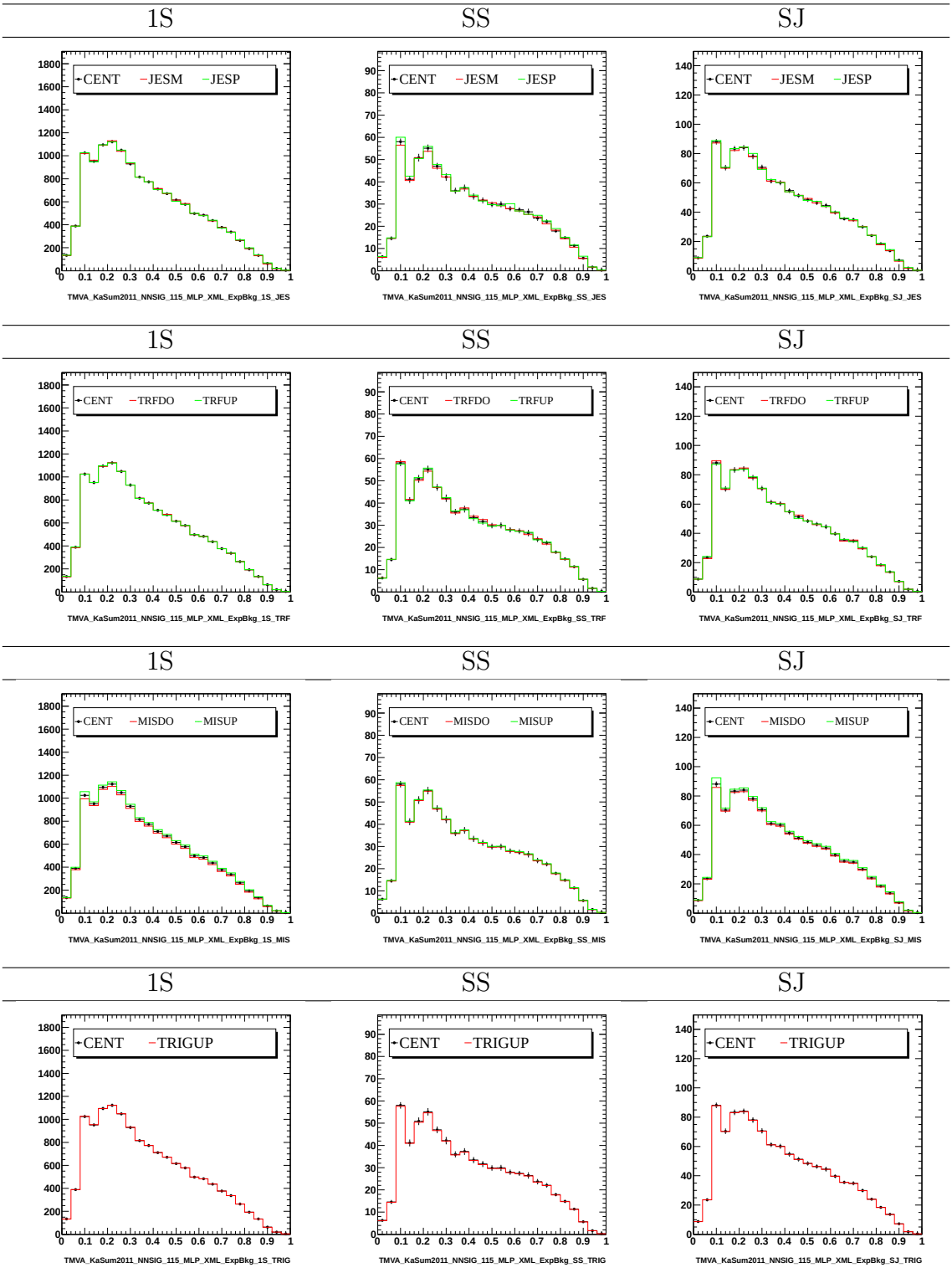


Figure 13.2.: Shape systematic uncertainties for the total background in each flavor category: JES (1st row), TRM (2nd row), mis-tag (3rd row) and trigger efficiency (4th row). The black, red and green lines represent the central, up- and down-shifted distributions respectively.

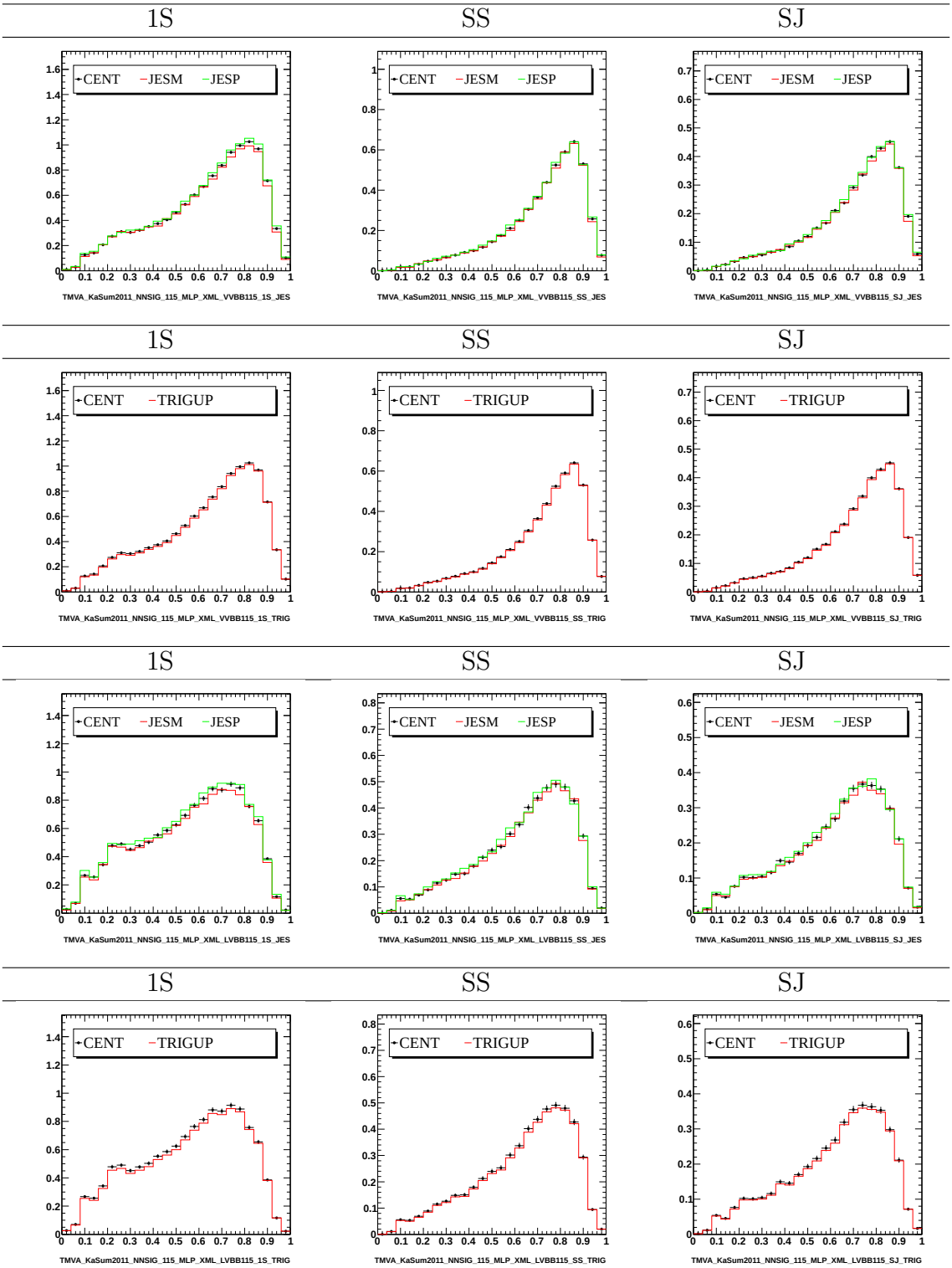


Figure 13.3.: Shape systematic uncertainties for the $ZH \to \nu\nu b\bar{b}$ (top two rows) and $WH \to \ell\nu b\bar{b}$ (bottom rows) in each flavor category: JES (rows 1 and 3) and trigger efficiency (rows 2 and 4). The black, red and green lines represent the central, up- and down-shifted distributions respectively.

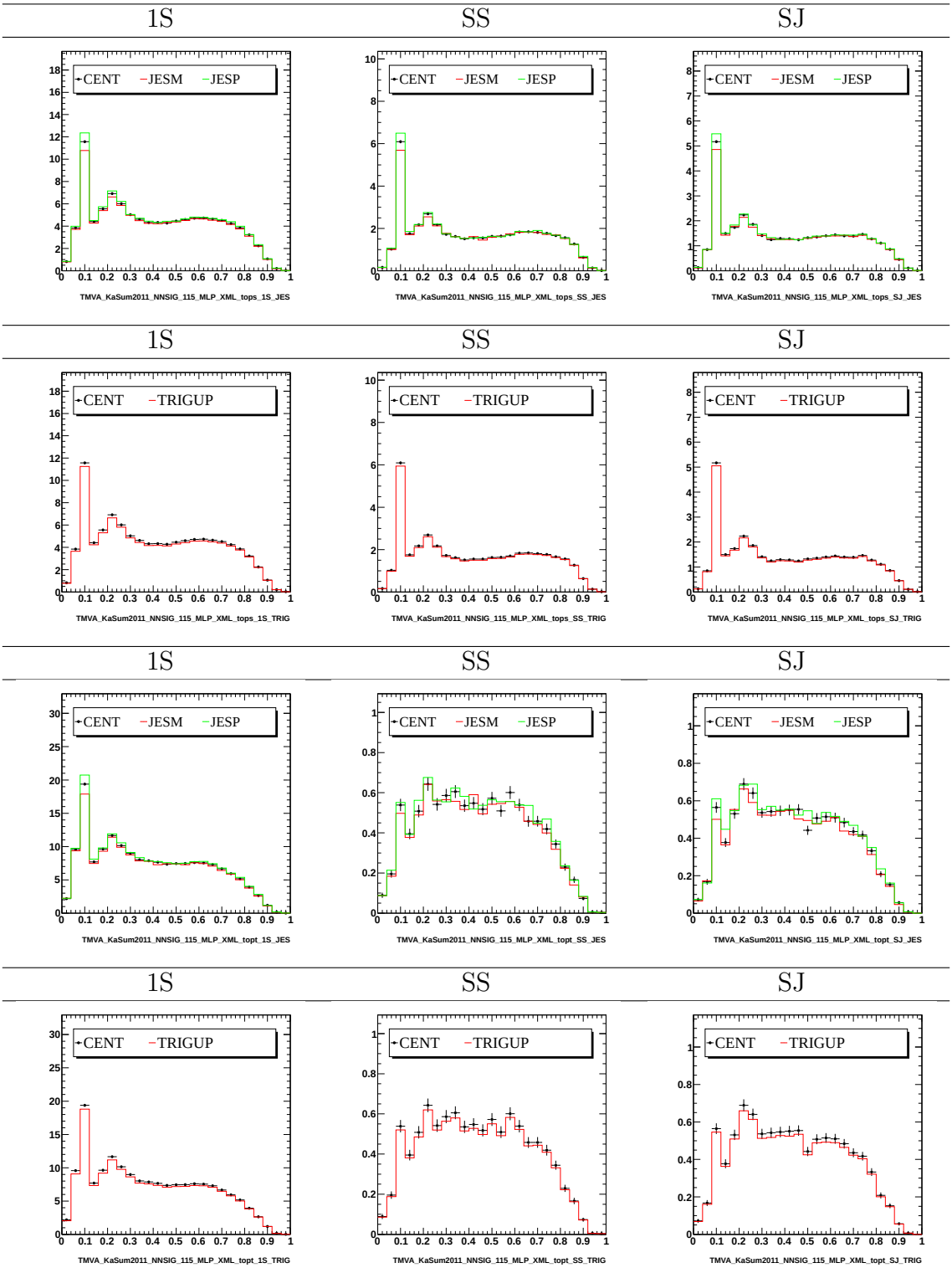


Figure 13.4.: Shape systematic uncertainties for the s- (top two rows) and t-channel (bottom rows) single top production in each flavor category: JES (rows 1 and 3) and trigger efficiency (rows 2 and 4). The black, red and green lines represent the central, up- and down-shifted distributions respectively.

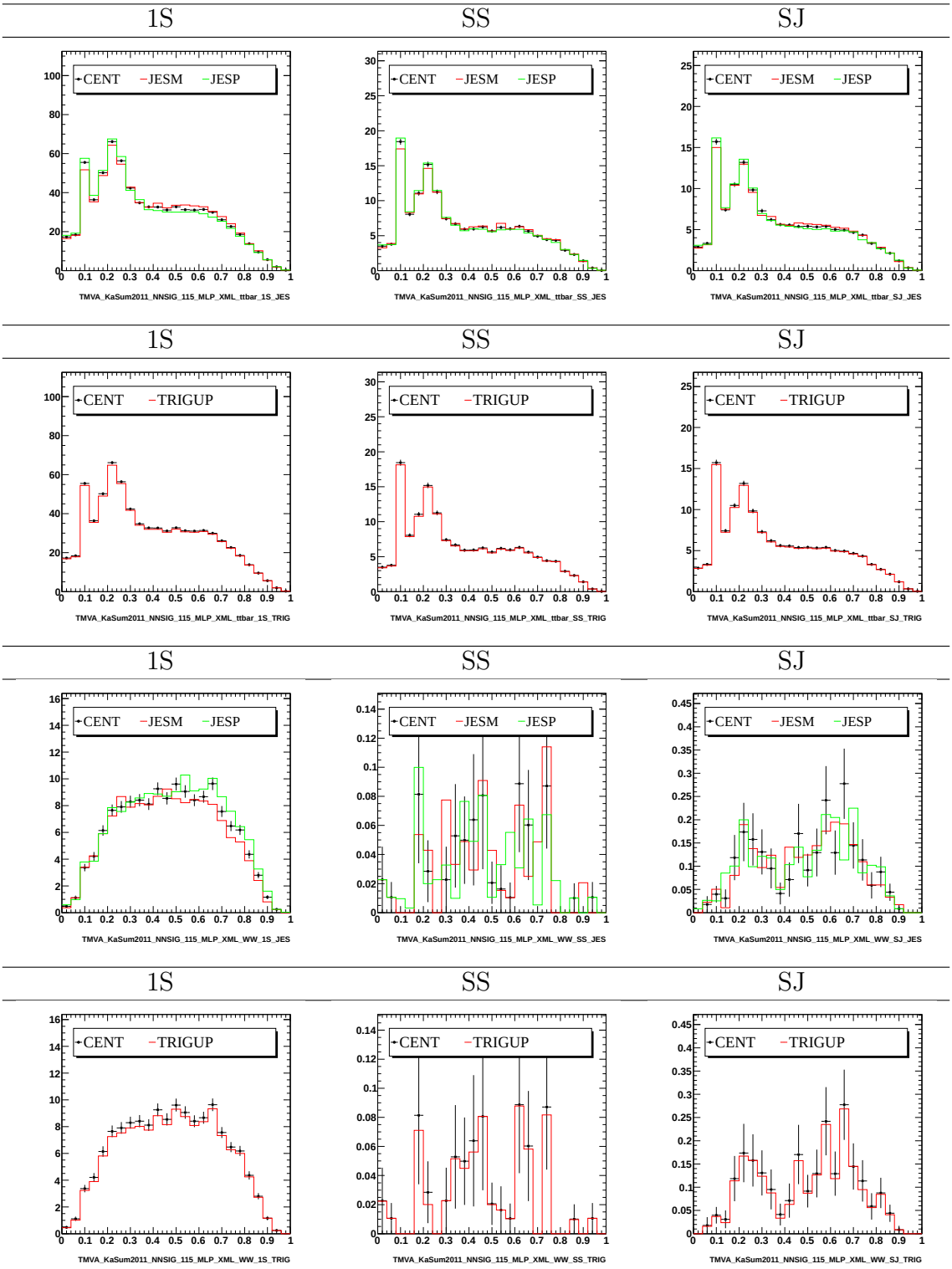


Figure 13.5.: Shape systematic uncertainties for the $t\bar{t}$ (top two rows) and WW (bottom rows) in each flavor category: JES (rows 1 and 3) and trigger efficiency (rows 2 and 4). The black, red and green lines represent the central, up- and down-shifted distributions respectively.

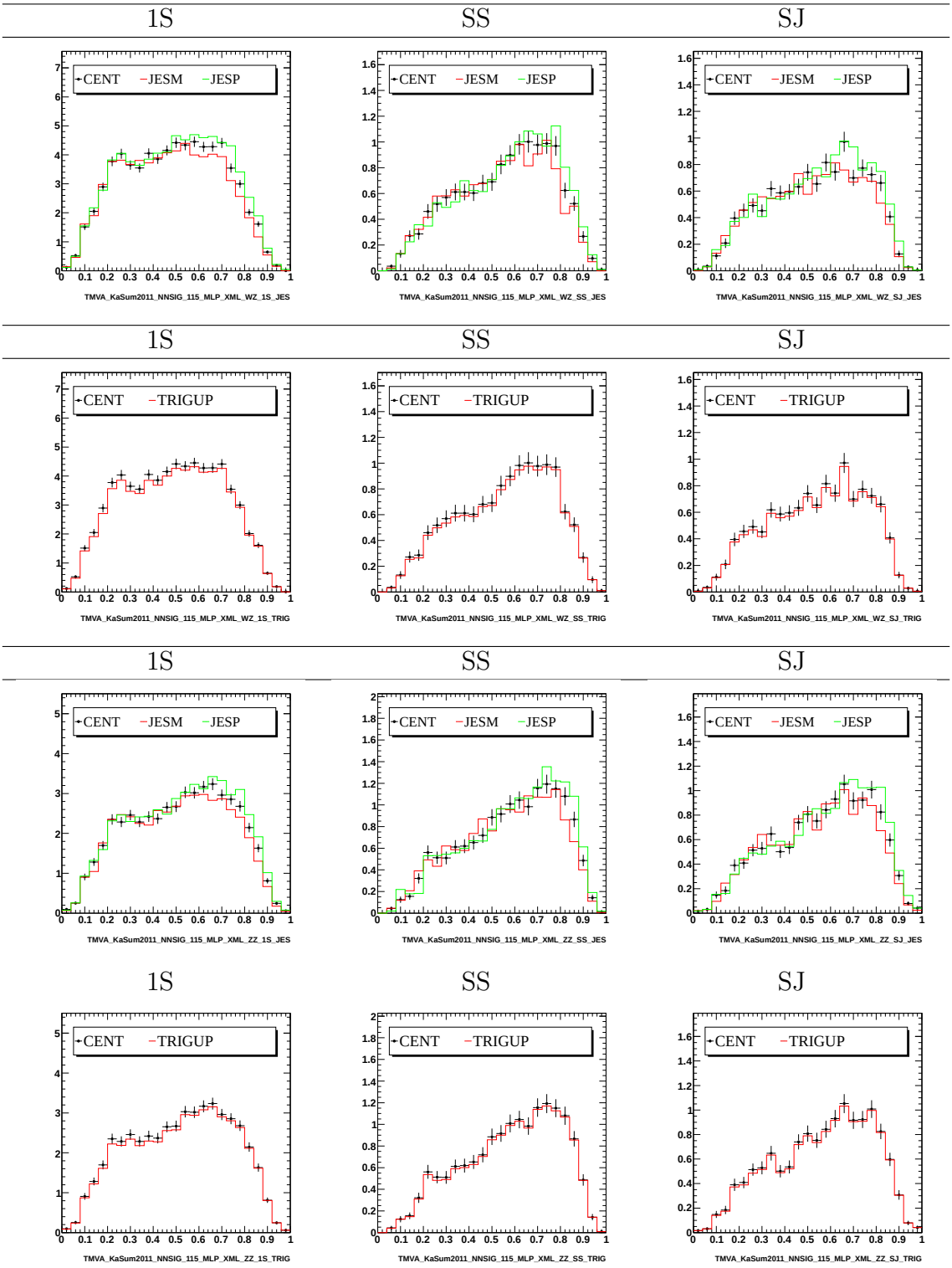


Figure 13.6.: Shape systematic uncertainties for the WZ (top two rows) and ZZ (bottom rows) in each flavor category: JES (rows 1 and 3) and trigger efficiency (rows 2 and 4). The black, red and green lines represent the central, up- and down-shifted distributions respectively.

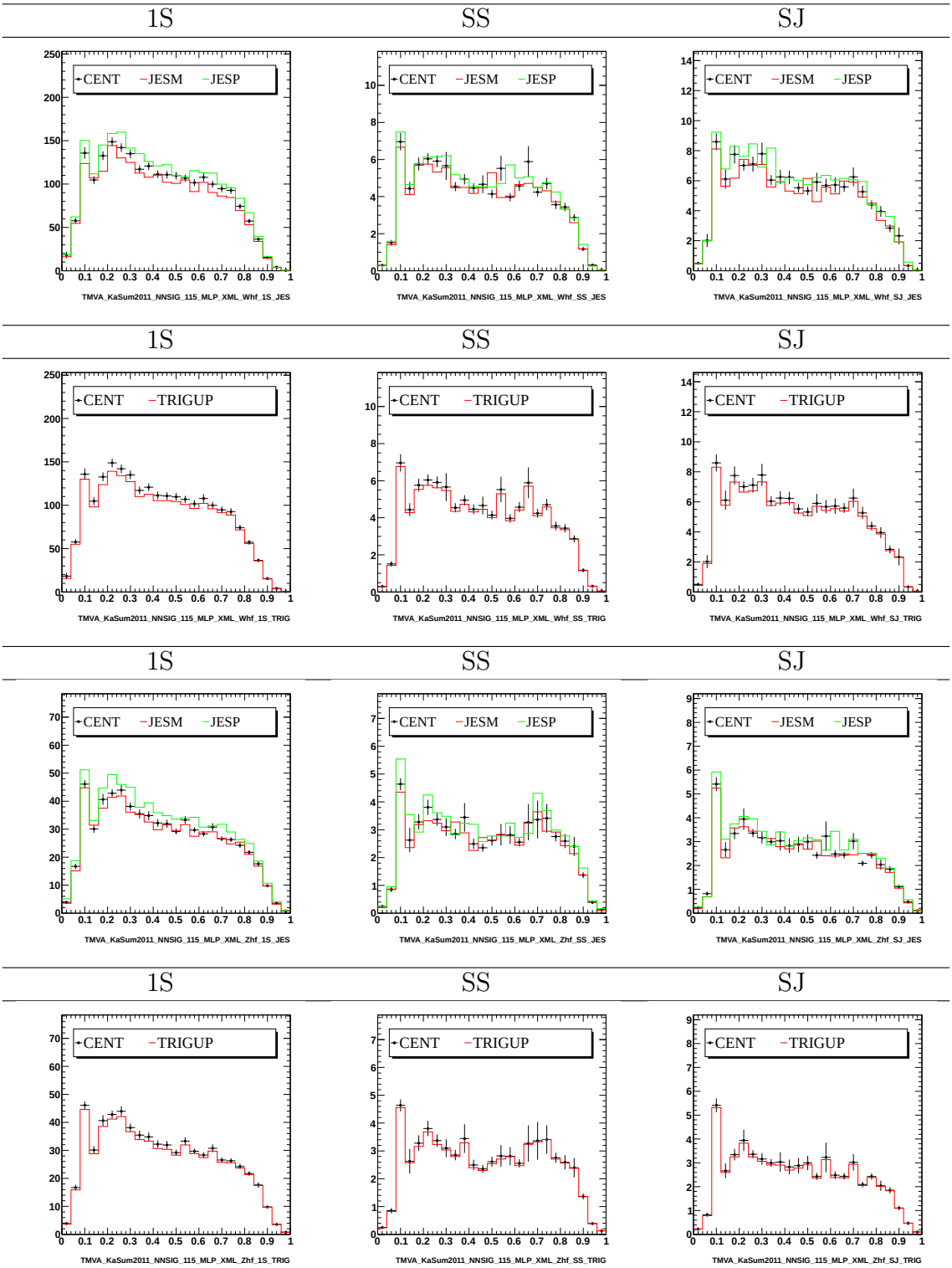


Figure 13.7.: Shape systematic uncertainties for the $W + h.f.$ (top two rows) and $Z + h.f.$ (bottom rows) in each flavor category: JES (rows 1 and 3) and trigger efficiency (rows 2 and 4). The black, red and green lines represent the central, up- and down-shifted distributions respectively.

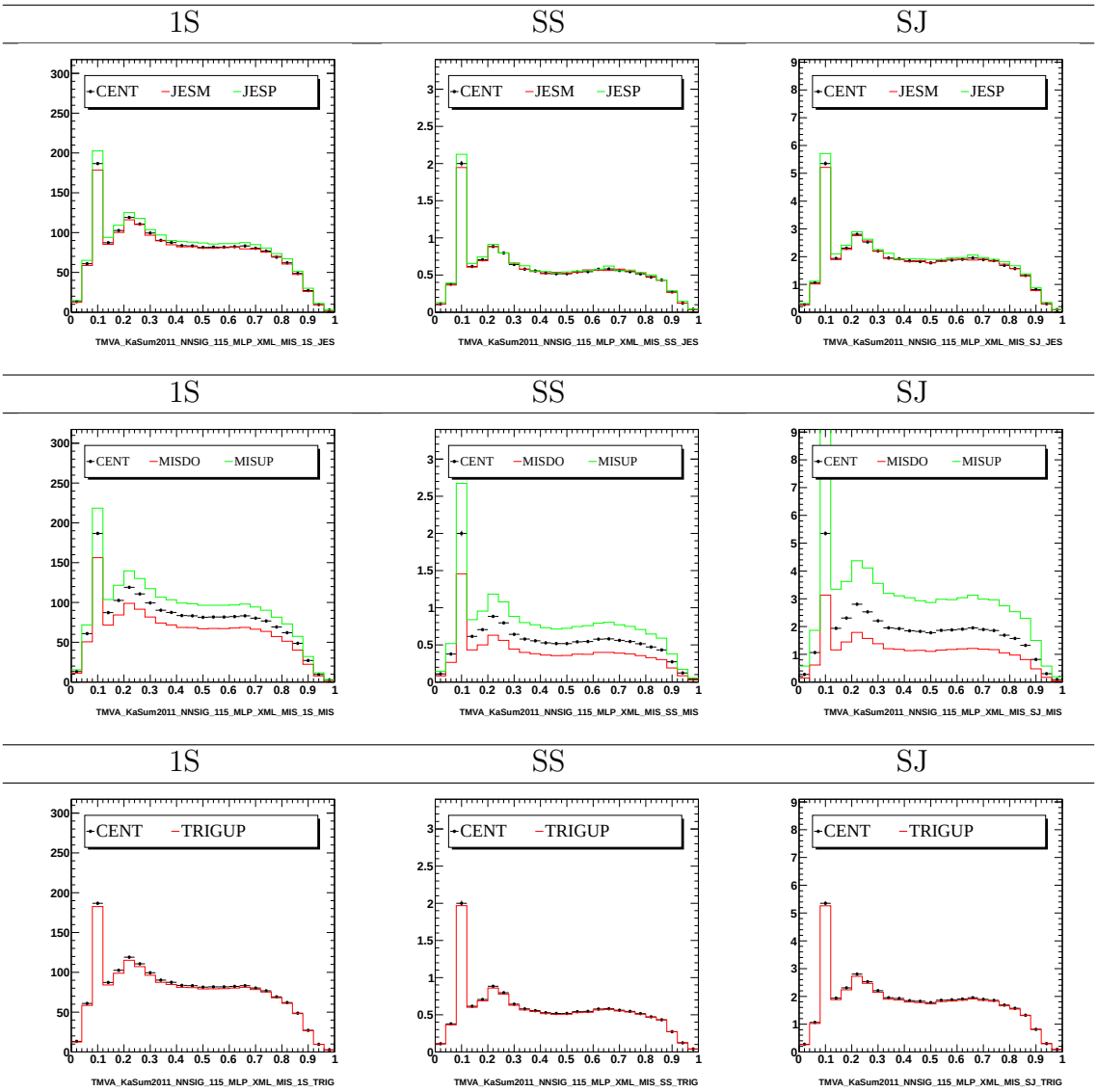


Figure 13.8.: Shape systematic uncertainties for the electroweak mis-tags background in each flavor category: JES (1st row), mis-tag (2nd row) and trigger efficiency (3rd row). The black, red and green lines represent the central, up- and down-shifted distributions respectively.

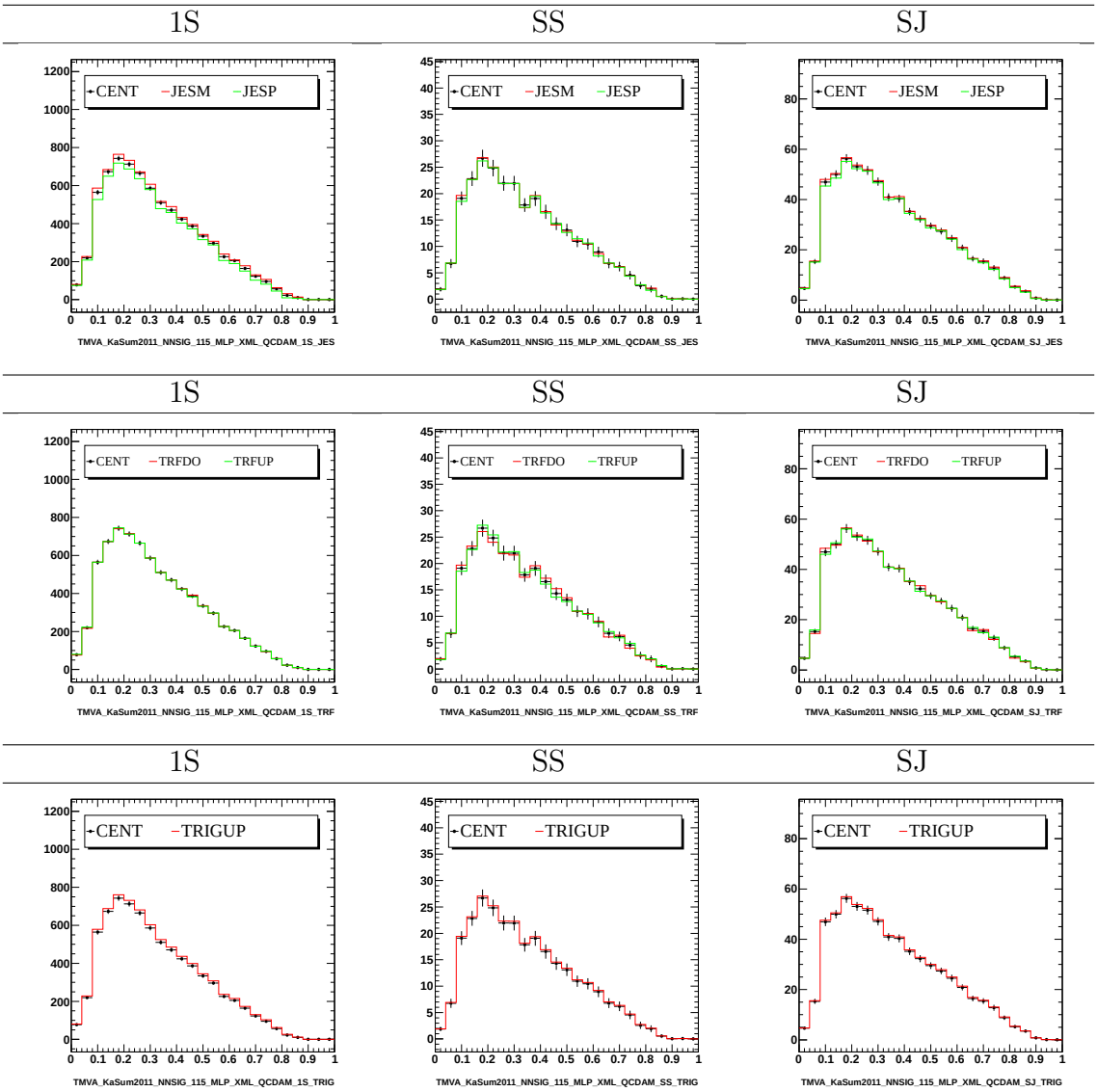


Figure 13.9.: Shape systematic uncertainties for the QCD multi-jet background in each flavor category: JES (1st row), TRM (2nd row) and trigger efficiency (3rd row). The black, red and green lines represent the central, up- and down-shifted distributions respectively.

/

14. SEARCH FOR THE SM HIGGS BOSON

The search for the Higgs boson (H) is currently the most active areas of research at the Tevatron. The electroweak fits to standard model (SM) parameters, performed including the latest Tevatron top mass averaged measurements [30], point to the value $m_H = 92^{+34}_{-26}$ GeV/c^2, or $m_H < 161$ GeV/c^2 [53]. In the mass region above 135 GeV/c^2 the searches focus on $gg \to H$ where $H \to WW$, because of the high cross section and the "low" backgrounds when the W's decay leptonically. At low mass the searches focus on the production of H associated with either a Z or a W boson. It has to be noted that while both CDF and DØ have excluded the presence of the Higgs boson in the mass region $156 < m_H < 177$ GeV/c^2 [55], the low mass searches are harder because of higher backgrounds and lower signal efficiency. In fact, the individual searches in the various low mass channels have yet to reach sensitivity to the SM

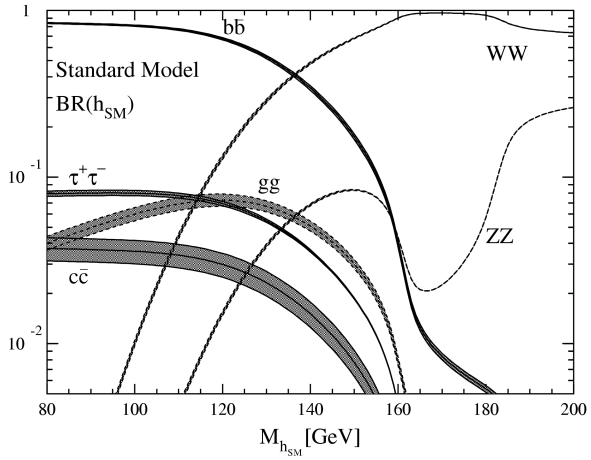


Figure 14.1.: Branching fractions for the decay of the SM Higgs boson. Source: [205]

Higgs cross section[1]. Nonetheless by combining these searches from CDF and DØ, the collaborations might have a chance to exclude or find a low mass Higgs boson.

This chapter describes a search for the SM Higgs boson production in association with a Z or W boson in $\bar{p}p$ collisions recorded by CDF. We consider a scenario where $Z \to \nu\nu$, or $W \to l\nu$ and the electron or muon escape detection; the Higgs boson decays into a $b\bar{b}$ pair. We significantly relax the kinematic requirements and accept 30-40% more acceptance to the WH/ZH signal with respect to the previous analysis [114, 161, 190]. Moreover, we now collect events from three different triggers and parametrize the complex turn-on using a dedicated artificial neural network. We split the data sample into various control regions (cf. section 11.3) and a signal region. The observed data in signal region is analyzed only after all background predictions and final event selection are determined.

The tools used in this analysis were used to measure the single top production cross-section for the first time in this channel [161]. This result was part of the recent observation of the single top quark by CDF [31]. We thus are at the stage where even the smallest backgrounds in this channel have been measured, apart from diboson production, which is as challenging as finding the Higgs boson, due to the lower invariant mass and the small branching ratio ($\mathcal{B}(Z \to b\bar{b}) = 0.2$ while $\mathcal{B}(H \to bb) = 0.75$). Moreover, these tools, and the QCD removing neural network, make this channel one of the most sensitive at low mass.

14.1 Review of the Event Selection

This analysis is based on an integrated luminosity of 7.8 fb^{-1}, collected with the CDF II detector between February 2002 and March 2011. The data are collected using a logical combination of the trigger paths presented in chapter 10, which require events to have large $\not\!\!E_T$ (MET45), or $\not\!\!E_T$ and two jets (MET35+JET, MET+DIJET). These new triggers help collect 10% more data. When using multiple trigger paths,

[1] As of the Summer of 2011, the best individual low mass Higgs analysis sets a 95% confidence level limit of 2.3 times the SM prediction.

it is crucial to properly model their combined effect on the data collection, and propagate it to our simulations. For this purpose, we have developed the neural network parameterization described in chapter 10.

From this inclusive dataset we select events offline with the following requirements: (a) a $\not\!\!E_T > 35$ GeV to avoid low trigger efficiencies and a too large increase in the backgrounds; (b) the two leading jets within $|\eta| < 2.0$, and of which at least one is central, i.e. having $|\eta| < 0.9$; (c) the leading jet with a $E_T > 25$ GeV, and the second jet with $E_T > 20$ GeV; (d) a separation in the $\eta - \varphi$ plane between the two leading jets of $\Delta R(j_1, j_2) > 0.8$, and (e) no events with 4 or more jets with $E_T > 15$ GeV/c^2 in the $|\eta| < 2.4$ region.

Additionally, we veto events with identified leptons and we require $\Delta\varphi(\not\!\!E_T, j_2) > 0.4$, $\Delta\varphi(\not\!\!E_T, j_3) > 0.4$ and $\Delta\varphi(\not\!\!E_T, j_1) < 1.5$ (to reject mis-measured QCD events). This constitutes the pre-selection (cf. chapter 11.2).

14.2 The search for the signal

The signal region is consists in pre-selection events with $NN_{QCD} > 0.45$. As mentioned above, we selected this criterion to maximize signal significance. The biggest background rejected is QCD events faking high $\not\!\!E_T$. The dominant backgrounds in the signal region are QCD, mis-tags, W/Z+jets and $t\bar{t}$ in similar proportions. We study the dynamic of those events to develop a NN with the goal of discriminating the surviving backgrounds from the interesting signal. Table 14.1 shows the event yields in the signal region. Figures 14.3 and 14.2 show the distribution of several kinematic variables. Additional plots are provided in appendix C.

$\not\!\!E_T$+b-jets 7.8 fb^{-1}: Signal Region [CDF II Preliminary]

	Exclusive SecVtx	SecVtx + SecVtx	SecVtx + JetProb
WW	147.6±20.9	0.7±0.6	2.4±1.6
WZ/ZZ	120.9±15.5	29.4±6.5	25.7±5.6
Single Top	277.3±50.3	51.5±10.6	43.3±9.0
Top Pair	728.2±93.1	154.0±25.0	137.6±22.6
$Z + h.f.$	678.5±249.3	63.5±28.2	61.3±26.6
$W + h.f.$	2234.7±793.1	99.6±41.8	124.5±56.5
Multijet	7567.6±62.8	279.8±8.6	639.0±11.9
Mistags	1910.4±157.5	14.0±1.7	44.9±5.3
Total	13665.2±1054.4	692.5±79.3	1078.8±89.8
Data	13791	636	994
m_H	Exclusive SecVtx	SecVtx + SecVtx	SecVtx + JetProb
100	37.6±3.5	15.8±2.1	12.4±1.7
105	33.7±3.1	14.2±1.8	11.2±1.5
110	29.8±2.7	12.7±1.6	9.9±1.3
115	25.4±2.3	10.9±1.4	8.6±1.1
120	21.1±1.8	9.0±1.1	7.1±0.9
125	17.1±1.6	7.3±0.9	5.8±0.8
130	13.2±1.2	5.8±0.7	4.5±0.6
135	9.8±0.9	4.2±0.5	3.3±0.4
140	6.8±0.6	3.0±0.4	2.3±0.3
145	4.5±0.4	2.0±0.2	1.6±0.2
150	2.8±0.2	1.2±0.1	0.9±0.1

Table 14.1: Number of expected and observed events in the signal region in all flavor categories.

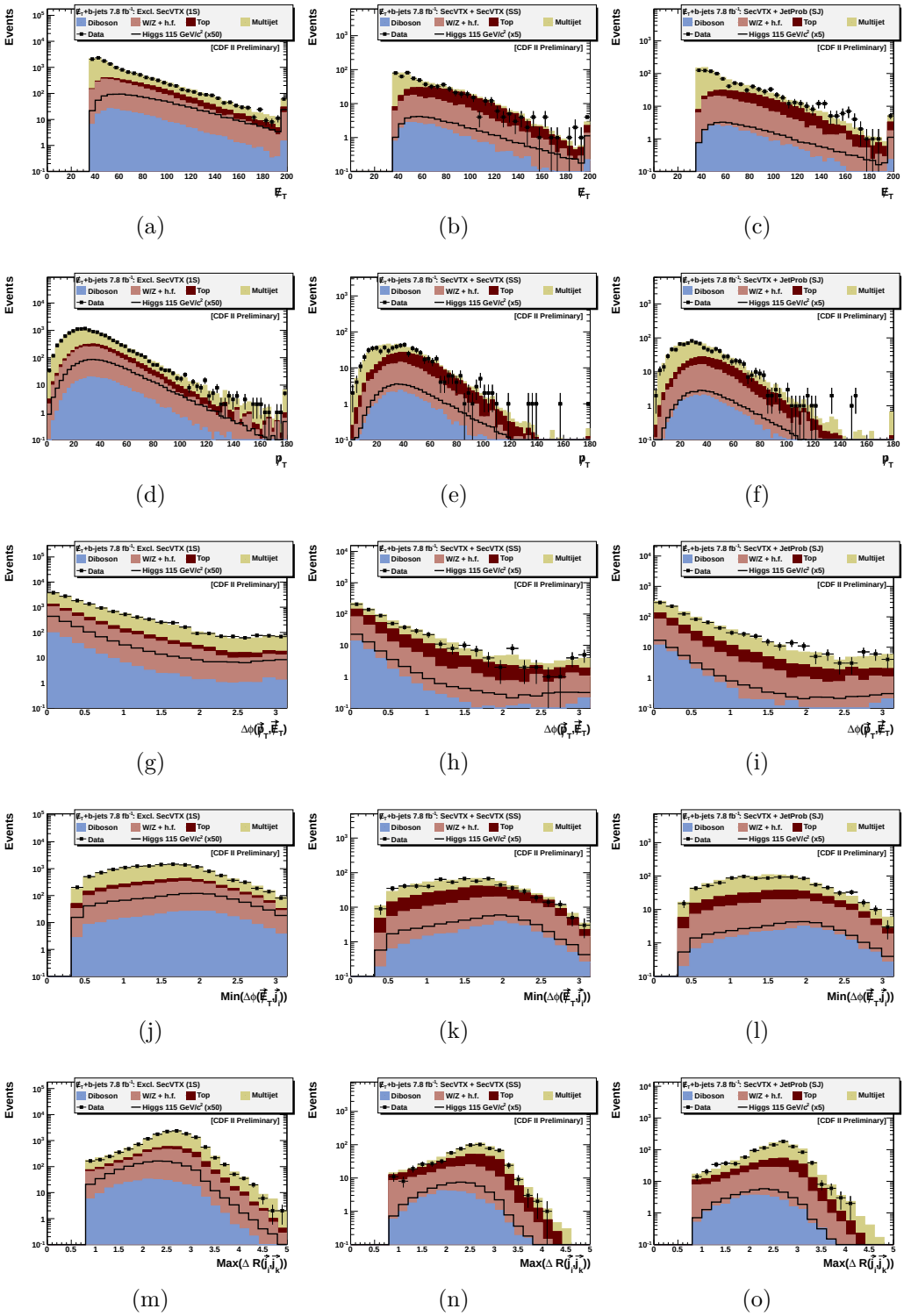


Figure 14.2.: Distribution of several input variables to NN_{QCD} in the signal region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The normalization of the multijet background is derived from a control region ($NN_{QCD} < 0.1$).

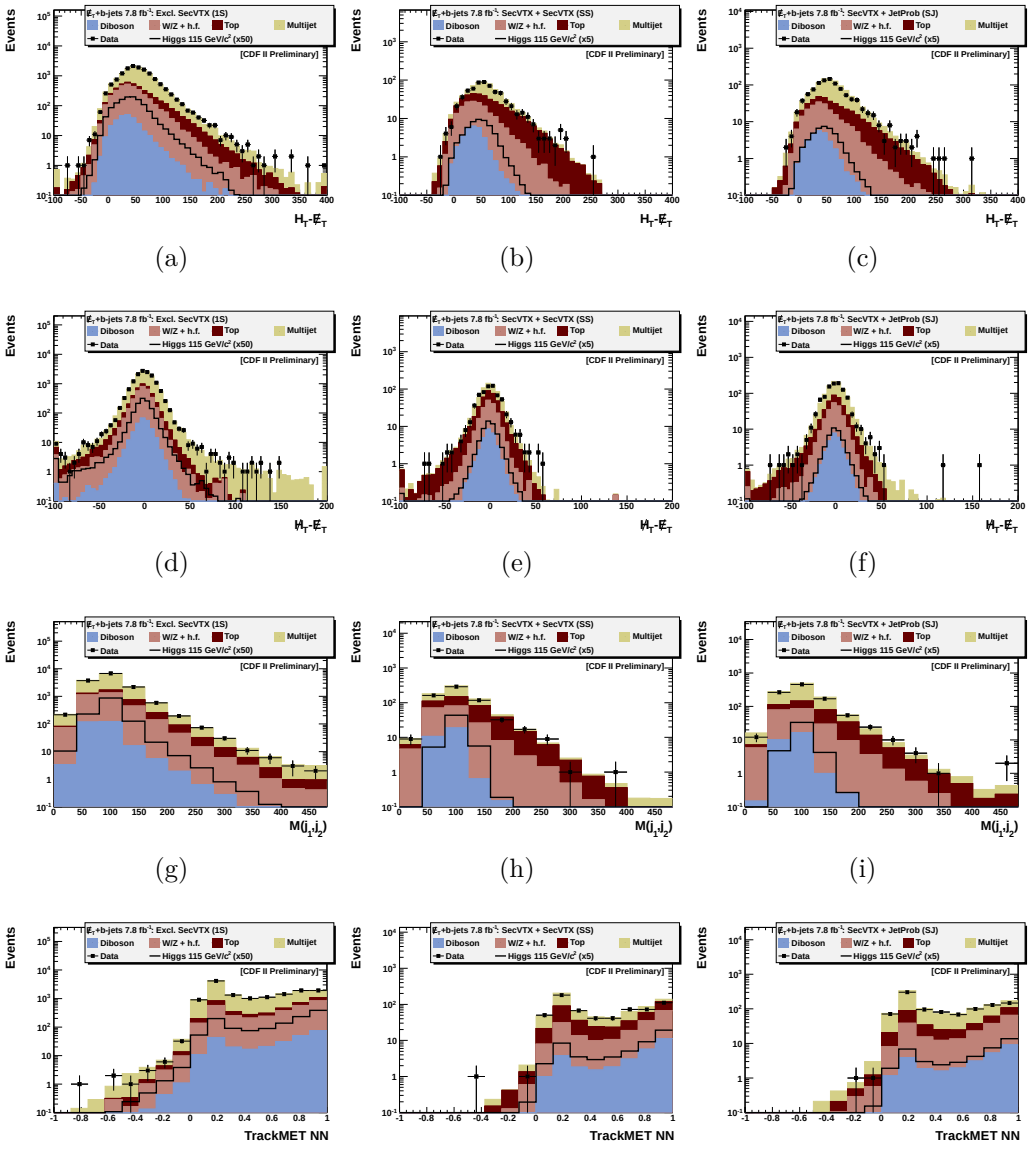
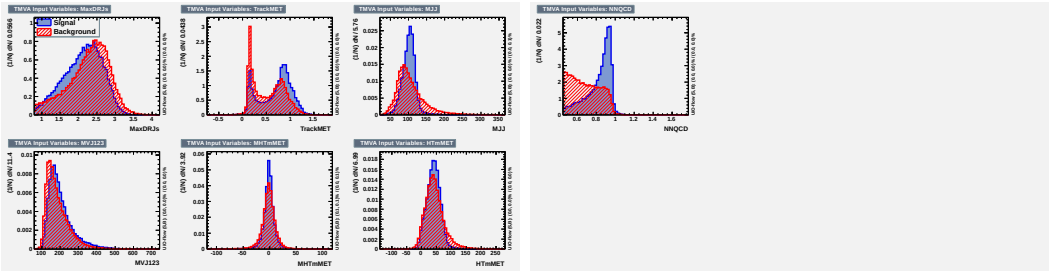
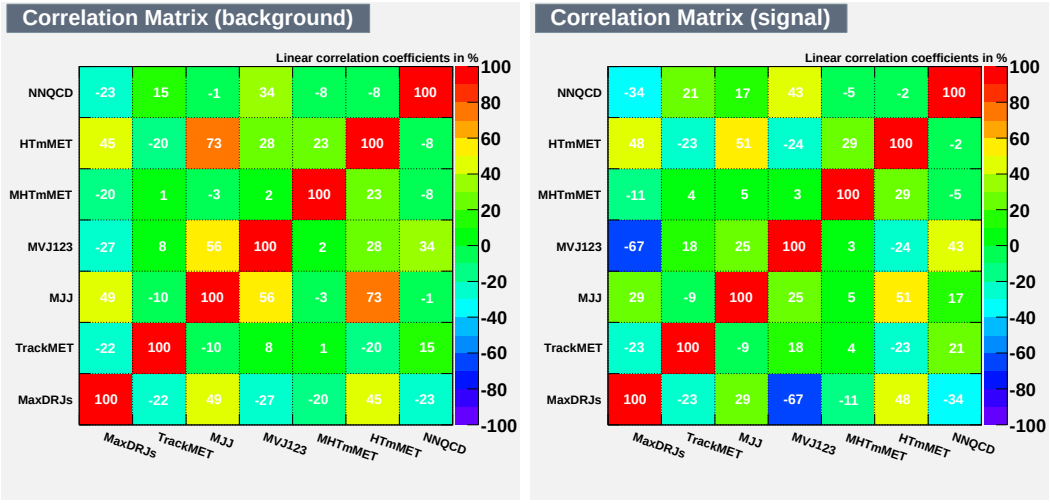


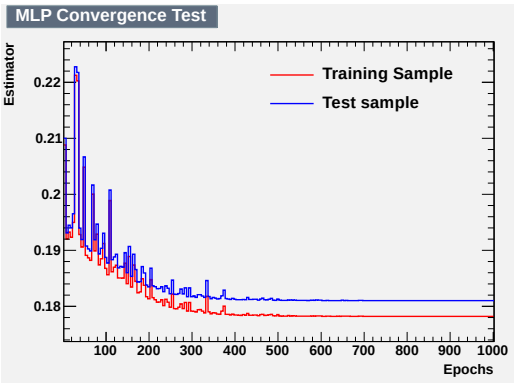
Figure 14.3.: Distribution of several input variables to NN_{SIG} in the signal region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The normalization of the multijet background is derived from a control region ($NN_{QCD} < 0.1$).



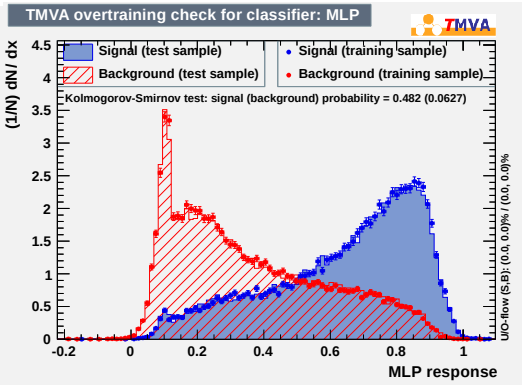
(a) Input variables



(b) Correlation matrices for the background (left) and the signal (right) samples

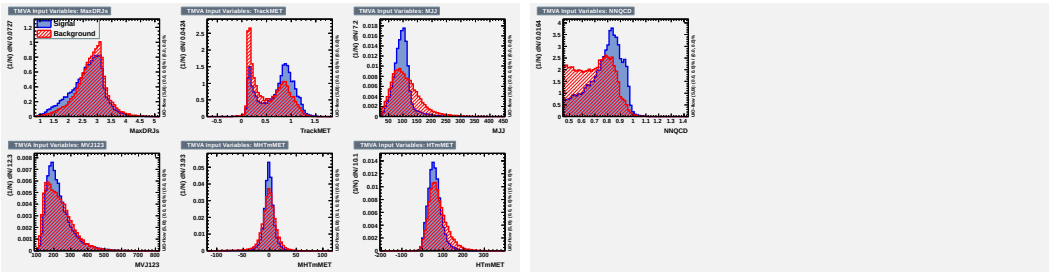


(c) NN convergence

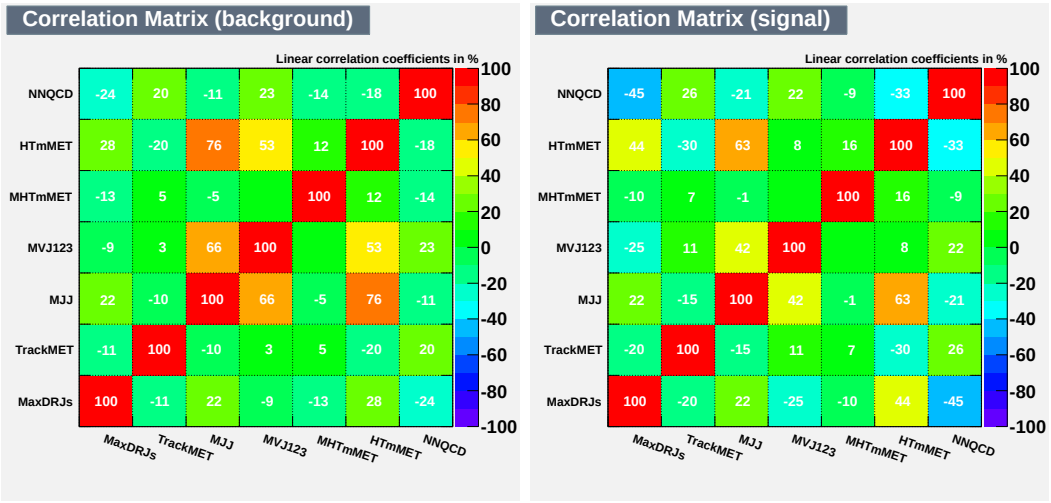


(d) NN ovetraining check

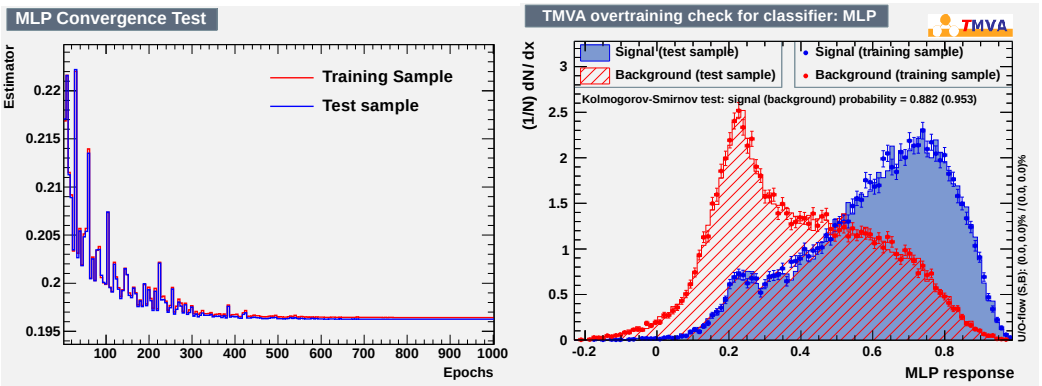
Figure 14.4.: Details on the training of the NN_{SIG} using two-jet events with $m_H = 115$ GeV/c^2: (a) inputs, (b) correlations, (c) convergence, and (d) output.



(a) Input variables



(b) Correlation matrices for the background (left) and the signal (right) samples



(c) NN convergence (d) NN ovetraining check

Figure 14.5.: Details on the training of the NN_{SIG} using three-jet events with $m_H = 115$ GeV/c^2: (a) inputs, (b) correlations, (c) convergence, and (d) output.

The invariant mass between the two leading jets	M_jj
The invariant mass of the $\not\!\!E_T$and all jets in the event	$M(\not\!\!E_T, \text{jets})$
The difference between $\not\!\!H_T$and $\not\!\!E_T$	$\not\!\!H_T - \not\!\!E_T$
The difference between the scalar sum of the jets in the event and $\not\!\!E_T$	$H_T - \not\!\!E_T$
A track-based NN sensitive to the Higgs	TrackNN
The maximum separation in $\Delta\phi$ between the jets	$\text{Max}(\Delta\phi(j_i, j_k))$
The QCD MJ rejection NN output	NN_{QCD}

Table 14.2: Input variables to the discriminant neural network NN_{SIG}.

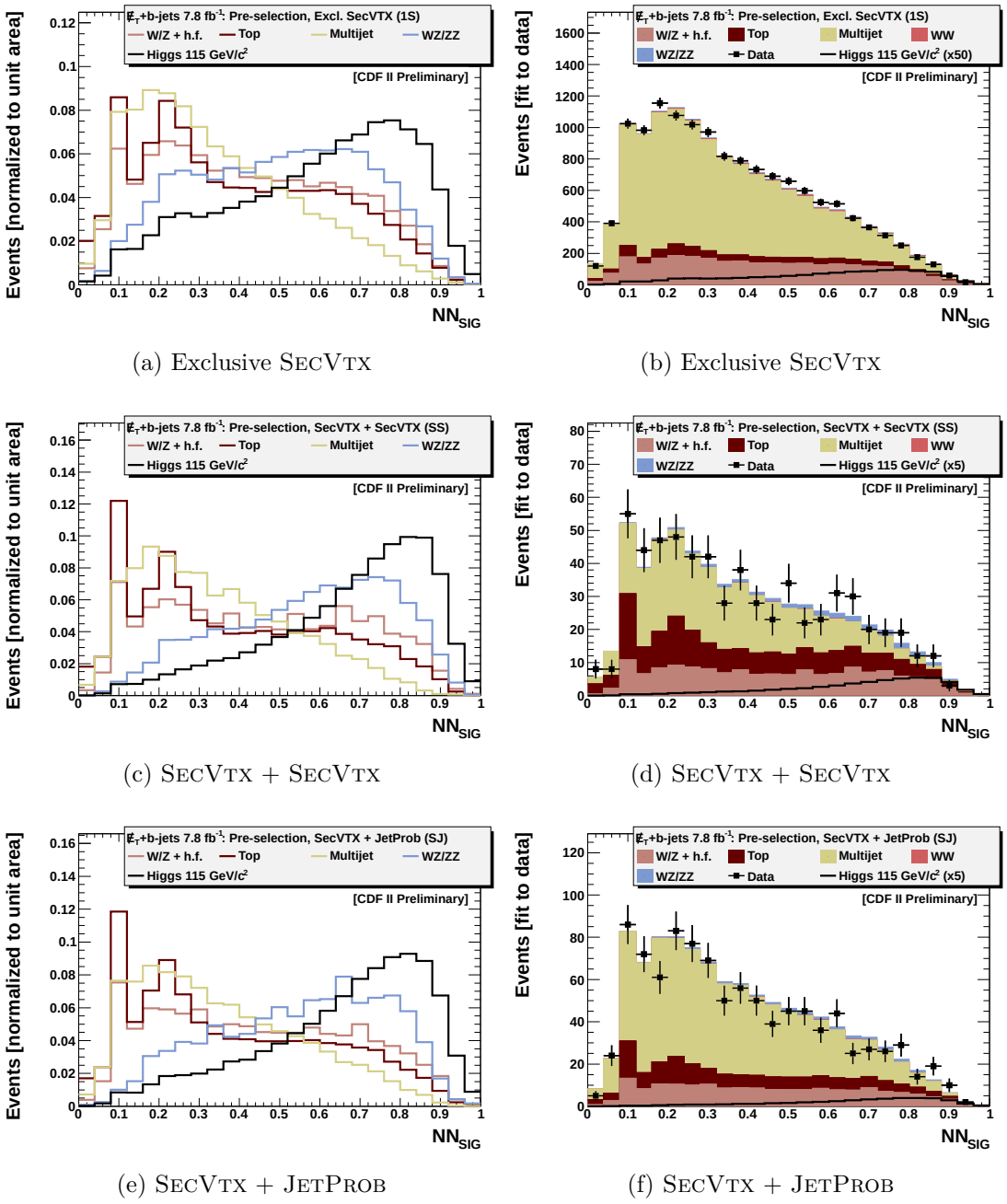
A second NN to discriminate the signal from the backgrounds

Since the background composition is different in events with 2 or 3 jets, we train separate networks in each category. The outputs of these networks are combined in when searching for the signal. For the NN training of 2-jet (3-jet) events we use a background sample made of 75% (50%) of $\not\!\!E_T$+jets untagged data (none of the jets in the event are b-tagged) and 25% (50%) of $t\bar{t}$ events. The Higgs signal used for the training is a mixture of 50% WH events and 50% ZH events. We train one network for the ten mass points we probe (from 100 to 150 GeV/c^2 in steps of 5 GeV/c^2). In order to increase the separating power of the NN, we implement a track-based discriminant, TrackMET, which was trained to optimize the separation of both ZH and WH events from QCD and $t\bar{t}$ backgrounds. A detailed description of the method can be found in ref. [206].

The neural network chosen here is a Multi Layer perceptron (MLP, cf. section 9.3). The 7 input variables are presented in Table 14.2. Figure 14.6 shows the NN output which we will use to scan for the presence of a signal.

14.3 Systematic Uncertainties

The systematic uncertainties are classified as correlated and uncorrelated errors according to the relations between the signal and the background processes. The uncorrelated systematic uncertainties are: QCD multijet normalization (0.7% in single tagged, 2.5% in SecVtx+SecVtx, 1.6% in SecVtx+JetProb samples), and the MC



(a) Exclusive SecVtx (b) Exclusive SecVtx

(c) SecVtx + SecVtx (d) SecVtx + SecVtx

(e) SecVtx + JetProb (f) SecVtx + JetProb

Figure 14.6.: Final discriminant NN_{SIG} output distribution in signal region with the binning used to perform the likelihood fit.

statistical fluctuations. Additionally, the statistical variations in TRM, which is used to estimate the multijet background, can also modify the distributions. It is taken

$\not\!\!E_T$+b-jets 7.8 fb^{-1} [CDF II Preliminary]		
m_H	Expected	Observed
100	$2.25^{+0.97}_{-0.67}$	1.79
105	$2.42^{+1.04}_{-0.69}$	1.80
110	$2.60^{+1.05}_{-0.75}$	2.24
115	$2.95^{+1.21}_{-0.84}$	2.31
120	$3.36^{+1.42}_{-0.96}$	3.28
125	$4.00^{+1.68}_{-1.12}$	5.39
130	$4.95^{+1.99}_{-1.40}$	4.99
135	$6.49^{+2.81}_{-1.86}$	8.00
140	$8.69^{+3.71}_{-2.46}$	11.56
145	$13.33^{+5.36}_{-3.90}$	16.70
150	$20.87^{+8.64}_{-5.88}$	30.44

Table 14.3: The predicted and observed cross-section limits of the ZH and WH processes combined when $H \to b\bar{b}$ divided by the SM cross-section.

into account by varying the TRM probability in each bin of the matrix by $\pm 1\sigma$, and the alternative shapes are used in the limit calculation. The correlated systematics are: luminosity (6.0%), b-tagging efficiency scale factor between data and Monte Carlo (5.2% for single and 10.4% for SecVtx+SecVtx, 8.3% for SecVtx+JetProb samples), trigger efficiency ($< 3\%$), lepton veto efficiency (2%), PDF uncertainty (3%) and jet energy scale. ISR/FSR systematic uncertainties (between 2% and 3%) are applied on the signal.

14.4 Results

Observing no significant excess in the data, we place 95% confidence level upper limits on the Higgs boson production cross section. Considering the systematic uncertainties listed above, we computed the expected limit for the Higgs cross-section when the Higgs is produced with a Z/W boson and decays to two b quarks where Z decays to neutrinos and W to leptons. We use a Bayesian method for deriving the limits [196] (cf. chapter 12). Table 14.3 shows the final results. In figure 14.7, we show these results graphically. The dotted line joins the median expected 95% C.L. upper

limits for each mass point. This represents the sensitivity of our analysis. The bands (which join the intervals at each mass point) indicate the 68% and 95% probability regions where the limits can fluctuate, in the absence of signal. The solid line joins the observed 95% C.L. upper limits obtained from the data. All the cross-sections times branching fraction are ratios with respect to the standard model cross-section.

Figure 14.8 shows the upper limits on SM Higgs production for each analysis channel at CDF together with the outcome of their combination. This analysis in the $\not\!\!E_T$+b-jets signature is one of the most sensitive. These results are further combined with those from the DØ collaboration (cf. figure 3.9b).

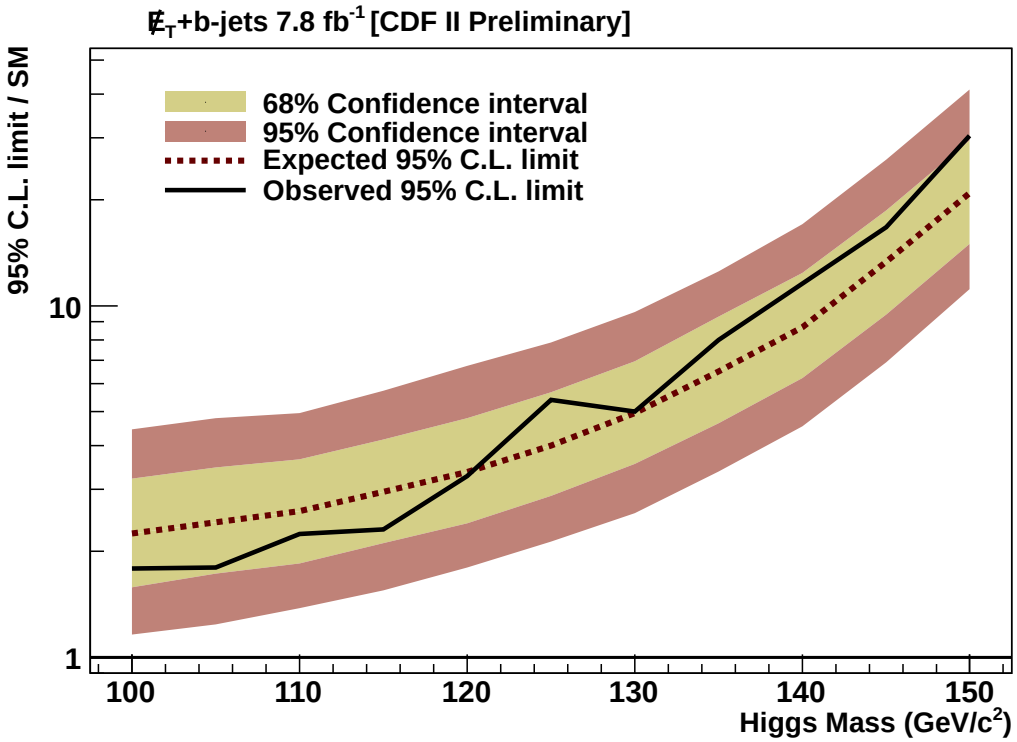


Figure 14.7.: Observed (solid line) and expected (dotted) median 95% C.L. upper limits (for the background-only hypothesis) on the VH associated production cross-section divided by the SM prediction in the $VH \to \not\!\!E_T b\bar{b}$ channel for $100 < m_H < 150$ GeV/c^2. The points are joined by straight lines for better readability. The bands indicate the 68% and 95% probability regions where the limits can fluctuate, in the absence of signal. The limits displayed in this figure are obtained with the Bayesian calculation described in chapter 12.

14.5 Summary

In this chapter, we presented an updated search of the Standard Model Higgs boson in events $VH \to \not\!\!E_T + b\bar{b}$ using 7.8 fb^{-1} of CDF data. We use an artificial neural network, NN_{QCD}, to suppress the dominant QCD background. Another network,

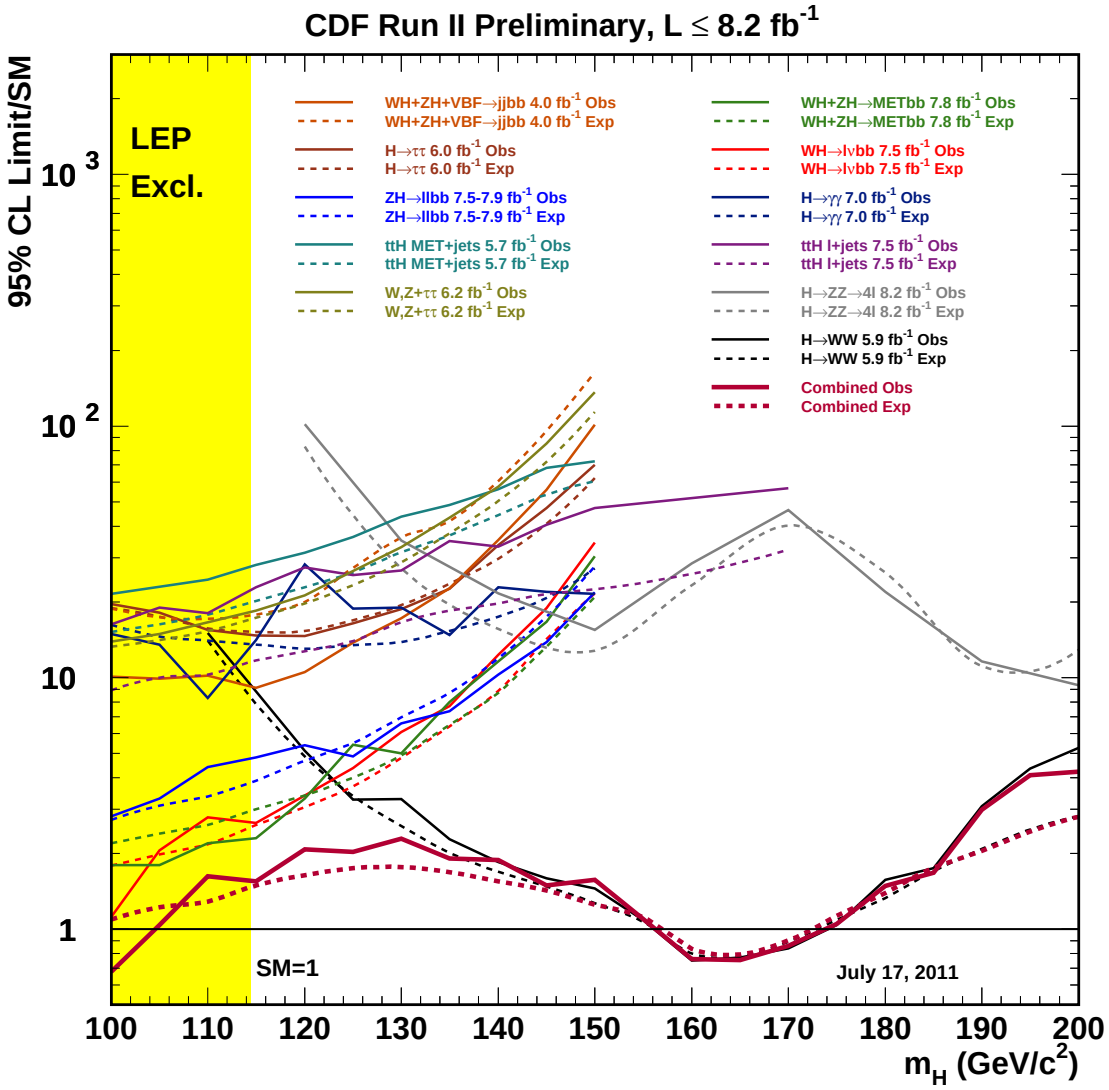


Figure 14.8.: The combined upper limit as a function of the Higgs boson mass between 100 and 200 GeV/c^2. Solid lines indicate the observed upper limit for each channel and combined result (dark red). Dashed lines indicate the median expected upper limit for each channel and the combined result (dark red). Credit: The CDF Collaboration

NN_{SIG}, is used to discriminate the signal from the surviving backgrounds. We have improved our background modeling and have analyzed additional data using a new trigger path. We expect to set a limit on the SM Higgs cross section times the branching ratio of 2.9 in the hypothesis of $m_H = 115$ GeV/c^2. In absence of a significant signal excess, we observe a limit of 2.3 times the SM prediction.

This channel suffers from larger systematic uncertainties than lepton+jets one but is nonetheless one of the most sensitive at the Tevatron.

15. VALIDATION OF THE ANALYSIS FRAMEWORK

Besides checking numerous validation plots to make sure the backgrounds are well understood, it is useful to use the analysis technique presented in this dissertation to measure known SM processes with a low production cross-section, but at least 10 times larger than that of VH associated production. In this chapter we present the outcome of cross-section measurements for three SM processes: (a) the electroweak production of single top; (b) the production of top pairs ($t\bar{t}$), and (c) the production of two vector bosons, i.e. WZ and ZZ[1].

15.1 The Single Top Cross-section and the V_{tb} Parameter

Measuring the electroweak single top production cross-section (s- and t-channel) was the first cross-check of our analysis technique performed in the $\not\!\!E_T$+b-jets signature[2]. While being significantly less sensitive than its lepton+jets counterpart [167], in which the W decay – and thus the top quark – can be reconstructed, this search is an interesting check of the methods outlined in the past chapters. In this section, we briefly describe the result obtained analyzing 2.1 fb^{-1} of CDF data [161].

Figure 15.1 shows the distribution of the discriminant neural network after the rejection of the QCD background using a dedicated neural network. A scan of this discriminant using the likelihood method described in chapter 12 yields an expected cross-section measurement of

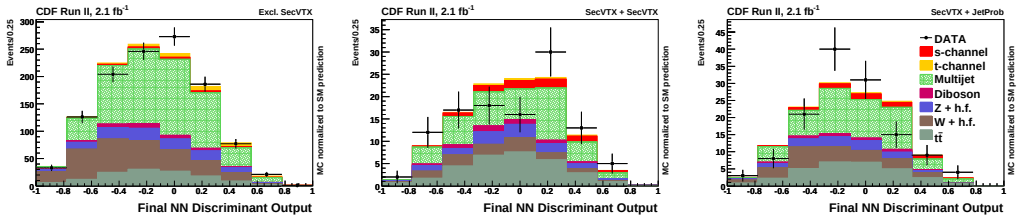
$$\sigma_{s+t}^{exp} = 2.7^{+2.3}_{-2.1} \text{ pb}. \tag{15.1}$$

[1]Because we are interested in processes decaying to $b\bar{b}$, we do consider WW as a background in the "diboson" search.
[2]This is the first analysis lead by the author.

Looking at the first 2.1 fb^{-1} in this signature, we measured a single top production cross-section of

$$\sigma_{s+t}^{obs} = 5.0^{+2.6}_{-2.3} \text{ pb.} \tag{15.2}$$

with is a 50% precision. The probability that the background model B fluctuated equal or up to the observed value in the data, i.e. the observed p-value, is 1.6×10^{-2} (2.1σ). The probability that B fluctuated equal or up to the median of signal plus background (S+B), i.e. the expected p-value, is 7.9×10^{-2} (1.4σ). We also measured the V_{tb} element of the CKM matrix. Using an unconstrained flat prior, we find $|V_{tb}| = 1.24^{+0.34}_{-0.29} \pm 0.07$(theory). Assuming a flat prior between 0 and 1, we measure $V_{tb} > 0.36$ at 95% confidence level. The details of the original analysis are presented in ref. [31, 161], and also in appendix D.



(a) Inclusive SecVtx (b) SecVtx+SecVtx (c) SecVtx+JetProb

Figure 15.1.: Final NN discriminant output distributions using 2.1 fb^{-1} of CDF data.

This analysis was combined with five other analyses of the lepton+jets sample [31, 167]. The expected p-value of the combination is over 5.9 standard deviations (σ), and the observed on right at 5σ. The cross-section measured is

$$\sigma_{s+t} = 2.3^{+0.6}_{-0.5} \text{ pb.} \tag{15.3}$$

The combination yields $|V_{tb}| = 0.91 \pm 0.11$ (exp.) ± 0.07 (th.) using an unconstrained flat prior, and $V_{tb} > 0.71$ at 95% C.L. using a flat prior between 0 and 1. At two dimensional fit was also performed, yielding $\sigma_t = 0.8 \pm 0.4$ pb and $\sigma_s = 1.8^{+0.7}_{-0.5}$ pb (cf.

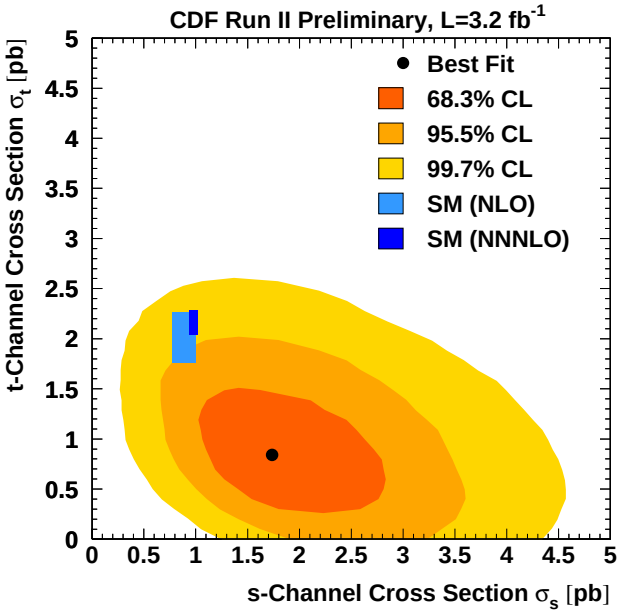


Figure 15.2.: Two-dimensional fit of the s- and t-channel single top cross-section for the CDF combination. Source: [167]

figure 15.2). These results were also combined to those of the DØ collaboration [207] (cf. figure 15.3).

The single top analysis is being update and the improvements were not ready for their inclusion in this dissertation. However, we expect similar performance, and only a slight improvement in sensitivity, due to the lack of separability between the single top signal and the background. The main reason for this is that we cannot fully reconstruct the top quark.

15.2 The $t\bar{t}$ Cross-section

The second cross-check using this channel was performed[3] in 2010 by measuring the $t\bar{t}$ production cross-section using 5.7 fb^{-1} of CDF data [208]. We use a dedicated network ($NN^{t\bar{t}}_{QCD}$, cf. figure 15.4a) to reject the QCD background. Another network

[3]This analysis effort was lead by Qiguang Liu, member of the Purdue Group.

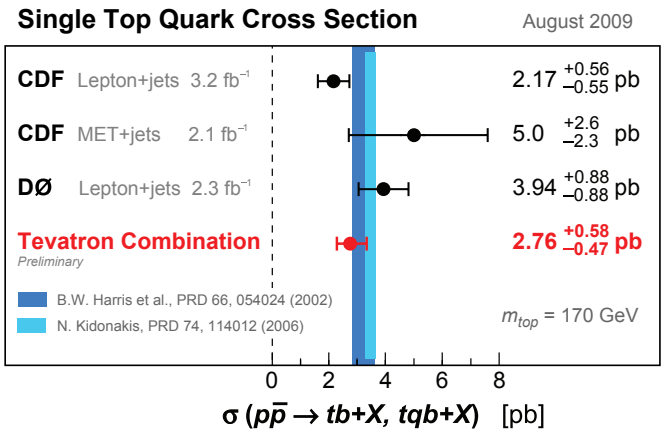


Figure 15.3.: The Tevatron combination of single top cross-section measurements using up to 3.2 fb^{-1} (CDF) and 2.3 fb^{-1} (DØ) of data. Source: [207]

($NN^{t\bar{t}}_{SIG}$, cf. figure 15.4a) separates the $t\bar{t}$ signal from the remaining backgrounds. We measure

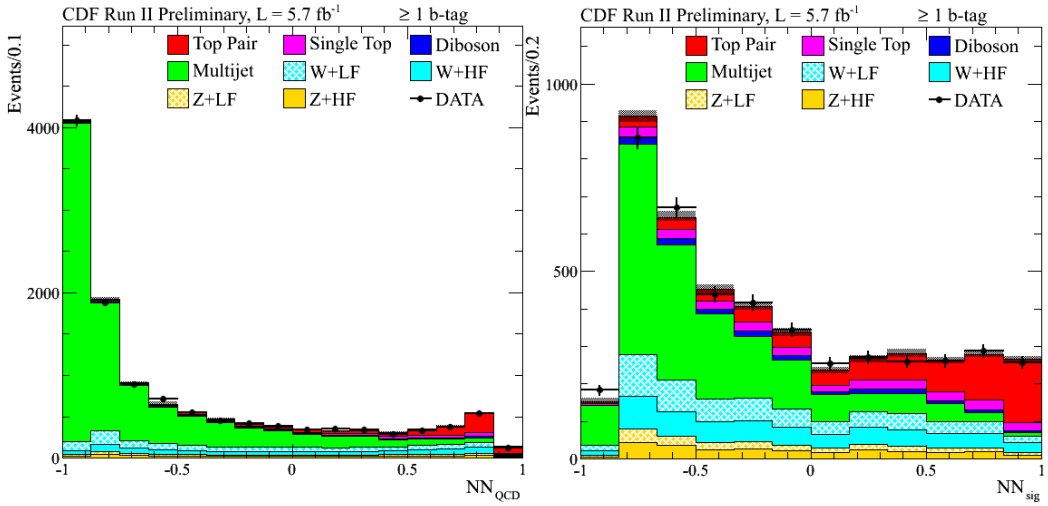
$$\sigma_{t\bar{t}} = 7.11 \pm 0.49\ (\text{stat.}) \pm 0.96\ (\text{syst.}) \pm 0.43\ (\text{lumi.}), \tag{15.4}$$

in good agreement with other measurements performed in more sensitive channels (cf. figure 15.5). This search was not updated, and does not use the relaxed kinematic cuts and the new trigger parameterization. It is nevertheless an important cross-check, probing a well-understood SM process.

15.3 The Diboson Cross-section

The production of WZ and ZZ where $Z \to b\bar{b}$ has the same final state as the VH associated production[4] but with a cross-section 10 to 20 times larger. We cross-check our sensitivity to the latter process using an artificial neural network, NN^{DB}_{SIG}, trained in the signal region described in section 14.2 to separate the WZ and ZZ

[4]All aspects are the same, except for the fact that the Z is a vector boson while the H is scalar. We are not sensitive to the latte effect.



(a) $NN^{t\bar{t}}_{QCD}$ (b) $NN^{t\bar{t}}_{SIG}$

Figure 15.4.: Neural networks used in the measurement of the $t\bar{t}$ cross-section in the $\not\!\!E_T$+b-jets signature: (a) the network used to reject the QCD background, and (b) the network used to discriminate the single top signal from the background.

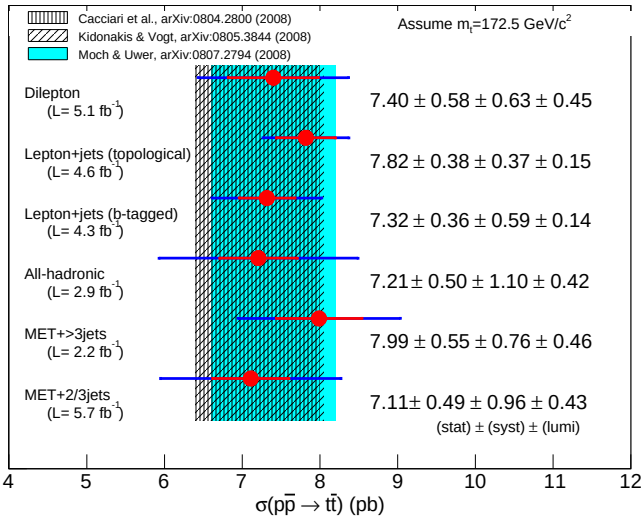
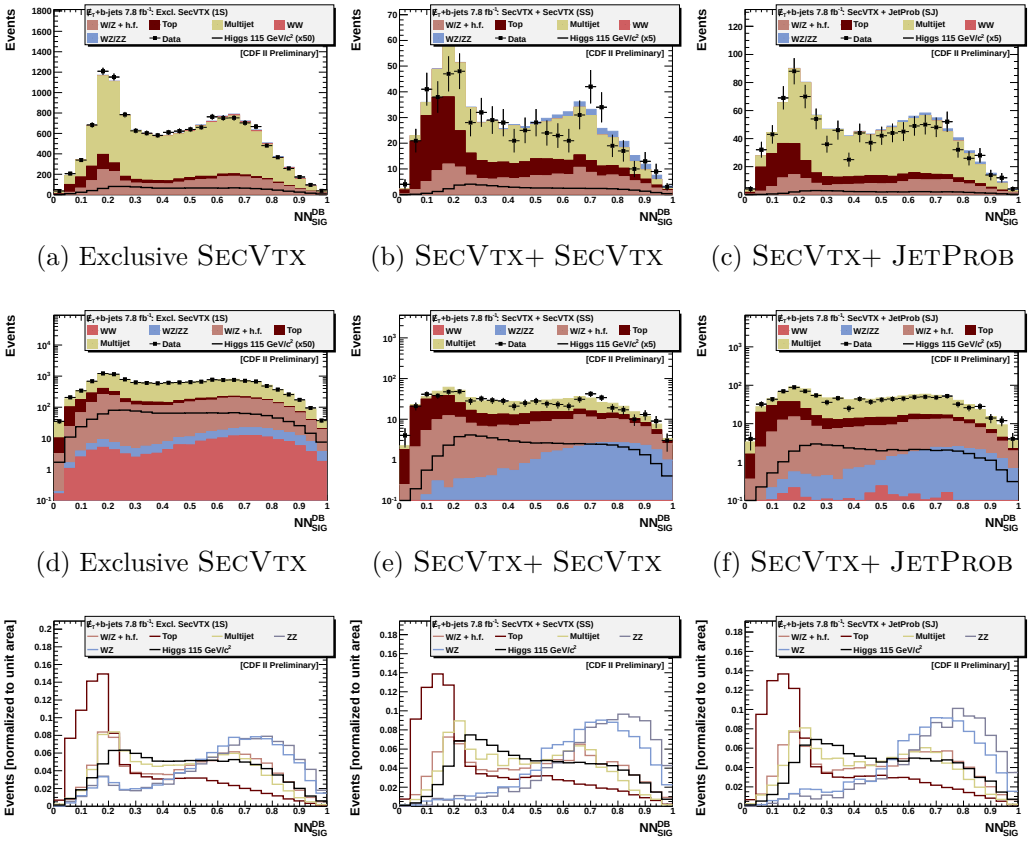


Figure 15.5.: The $t\bar{t}$ cross-section measurements at CDF. The sensitivity of the $\not\!\!E_T$+2/3b-jets is similar to that of the dilepton and all-hardonic channels.

signals from the remaining backgrounds. Figure 15.6 shows the output of NN^{DB}_{SIG}. Fitting separately for WZ and ZZ production, we obtain

$$\sigma_{WZ}/SM = 0.09 \qquad \sigma_{ZZ}/SM = 1.7. \tag{15.5}$$



(a) Exclusive SecVtx (b) SecVtx+ SecVtx (c) SecVtx+ JetProb

(d) Exclusive SecVtx (e) SecVtx+ SecVtx (f) SecVtx+ JetProb

(g) Exclusive SecVtx (h) SecVtx+ SecVtx (i) SecVtx+ JetProb

Figure 15.6.: Distribution of NN^{DB}_{SIG} for the three flavor categories. (a-f) Expected yields. (g-i) Shape comparison.

Figure 15.8a shows the posterior probability density distribution for $(\sigma_{WZ}, \sigma_{ZZ})$ times the SM prediction. This result is compatible with the SM prediction of $(1, 1)$ within uncertainties. The most likely value according to the fit is shifted because of the difficulty to separate WZ from ZZ production in the discriminant, as shown in figure 15.6.

We also extract the cross-section for $\sigma(WZ + ZZ)$ using the same discriminant, and obtain

$$\sigma(WZ + ZZ)/SM = 1.1 \pm 0.6. \tag{15.6}$$

Finally, we probed diboson production using the di-jet invariant mass spectrum $m(j_1, j_2)$. We obtain

$$\sigma_{WZ}/SM = 0.54 \qquad \sigma_{ZZ}/SM = 0.07. \tag{15.7}$$

and

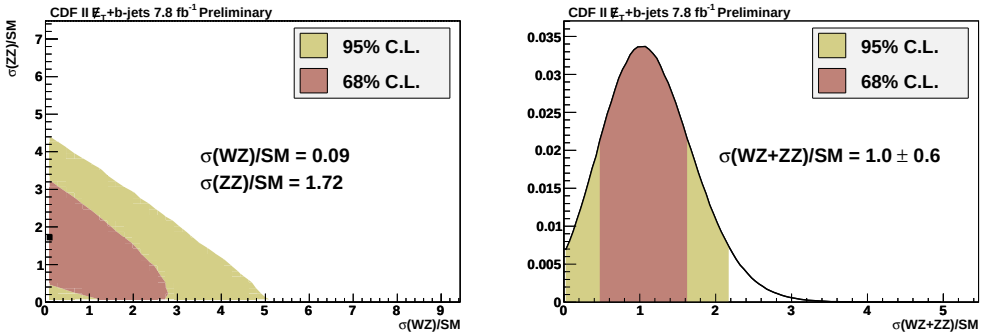
$$\sigma(WZ + ZZ)/SM = 0.26^{+0.66}_{-0.23}. \tag{15.8}$$

As expected, the choice of a network discriminant function improves the sensitivity of the analysis.

Although not as precise as diboson measurements in other signatures [209–212], this result shows that we are able to probe cross-sections that are about one order of magnitude larger that the SM production using our analysis techniques. We note however that this result is not very sensitive; we expect that combining this result with those from other searches at CDF and DØ will improve the sensitivity.

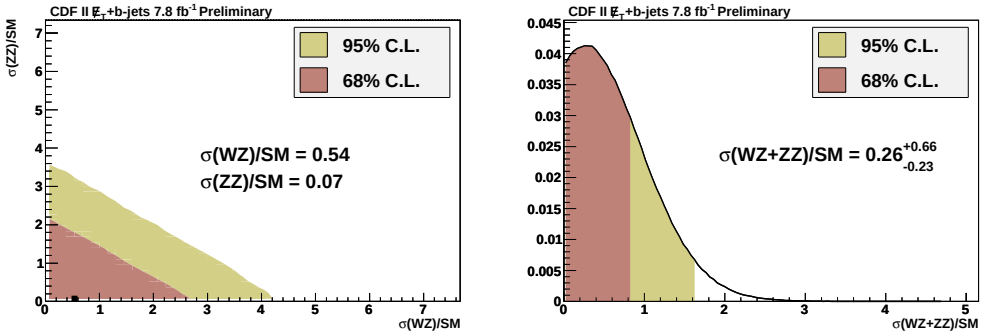
15.4 Summary

In this section, we have presented three cross-checks to our analysis technique. We have shown that our tools are able to measure cross-sections ten times larger than the Higgs signal we are after. This makes us confident in the robustness of our tools.



(a) (b)

Figure 15.7.: Posterior probability density distributions for (a) $(\sigma_{WZ}, \sigma_{ZZ})$ and (b) $\sigma(WZ + ZZ)$ obtained from the NN^{DB}_{SIG} distribution (figure 15.6). The red and gold bands represent the 68% and 95% C.L. intervals respectively. The cross-sections are expressed by their ratio to the SM prediction.



(a) (b)

Figure 15.8.: Posterior probability density distributions for (a) $(\sigma_{WZ}, \sigma_{ZZ})$ and (b) $\sigma(WZ + ZZ)$ obtained from the $m(j_1, j_2)$ distribution. The red and gold bands represent the 68% and 95% C.L. intervals respectively. The cross-sections are expressed by their ratio to the SM prediction.

16. SUMMARY AND FUTURE PROSPECTS

In this dissertation, we presented an experimental test of the standard model of particle physics. The predictions of the SM have been verified with great precision [7] but the Higgs boson still eludes us. Its mass is not predicted by the theory, but can be obtained by fits to precision electroweak data, which favor $m_H < 161$ GeV/c^2. We therefore performed a search for the Higgs boson in this low mass range.

We analyzed a sample of $p\bar{p}$ collision events at a center-of-mass energy of 1.96 TeV corresponding to an integrated luminosity of 7.8 fb^{-1} recorded by the Collider Detector at Fermilab. We focused on the $\not\!\!E_T$+b-jets signature, i.e. events with no leptons, large transverse energy imbalance and two high-E_T jets, of which at least one is compatible with originating from the decay of a b-quark. In this sample, we search for the Higgs boson produced in association with a vector boson (W or Z) where the Higgs decays into a $b\bar{b}$ pair – the dominant decay mode for $m_H < 135$ GeV/c^2 – and the Z decays into neutrinos. Because of the limitations of our object reconstruction capabilities, we are also sensitive to events where $W \to \ell\nu$ and, in a much lesser extent, where $Z \to \ell\ell$ and the lepton is not identified.

The events are collected using for the first time the logical combination of three independent triggers each using the $\not\!\!E_T$ information. To model the efficiency associated to this combined trigger, we devised a novel artificial neural network technique. With it, we can parameterize easily the combination of an arbitrary number of triggers. Indeed, the network is very good in approximating the trigger efficiency in a large-dimensional space (9 or 14 variables in our case), and it also takes into account the correlations between the variables. This is an important improvement over conventional fit-based approaches, which become difficult and labor intensive in the case of many triggers.

The backgrounds in this analysis consist of electroweak processes, which involve the production of a neutrino, and QCD multi-jet production, which does not involve neutrinos but can generate the $\not\!\!E_T$+b-jets signature when one of the jets is mis-measured. We reduce these backgrounds first using basic selection criteria, and requiring that at least one of the jets in the event be b-tagged.

The new trigger parameterization allows to significantly relax the kinematic requirements with respect to previous iterations of this analysis, since we do no longer need to select events from the fully-efficient regions of the trigger. The relaxed selection criteria allow for a gain of 40% in acceptance to the Higgs signal.

After the basic selection, the main background is still QCD multi-jet production. We suppress it by one order of magnitude using a neural network specifically trained to separate the QCD background from the signal. This technique was introduced in the search for the single top quark in this signature.

We train another network to separate the signal from the remaining backgrounds. In the absence of evidence of SM Higgs production, we use Bayesian statistical inference to compute 95% confidence level (C.L.) limits on the production cross-section of the SM Higgs in association with a vector boson times the branching fraction to $\not\!\!E_T + b\bar{b}$.

This process is repeated for eleven mass points between $100 < m_H < 150$ GeV/c^2 in 5 GeV/c^2 steps. The expected median expected upper limits vary between 2.25× SM and 20.1× SM in this range. We set observed limits between 1.8 and $30.4 \times \sigma_{SM}(VH) \cdot \mathcal{B}(\not\!\!E_T b\bar{b})$. The improvement with respect to the last iteration of the analysis is of the order of 10%, as shown in figure 16.1, where we compare this analysis to the two results published by CDF for the $\not\!\!E_T$+b-jets signature. These results are among the most sensitive at the Tevatron.

We looked for signals with cross-sections of the other of the pb (i.e. about one order of magnitude larger than the Higgs cross-section) in order to validate the analysis techniques. We studied the single top, top pair and diboson cross-section in this signature, confirming the robustness of our analysis tools.

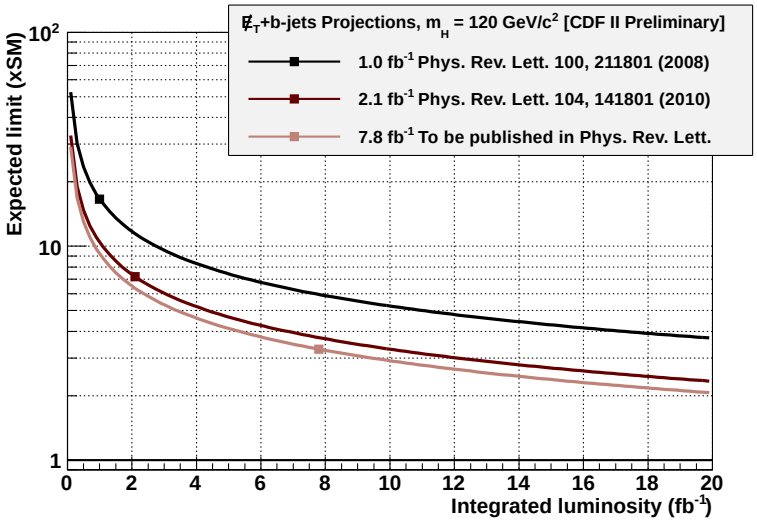


Figure 16.1.: Search for the Higgs in the $\not\!\!E_T$+b-jets at CDF: comparison between the two results published by CDF and this analysis, to be published in PRL. The dots represent the individual analyses, while the lines extrapolate them according to $\mathcal{L}^{-1/2}$. The improvement in sensitivity over the last iteration is of the order of 10%.

Future prospects

The techniques developed for this analysis, namely the use of dedicated neural networks to parameterize the trigger efficiency and to separate the signal from the backgrounds can be widely applied. The main difficulty for improving the signal over background ratio resides in understanding the physics involved to pick appropriate variables for this task. These tools have been used in searches for new physics in the $\not\!\!E_T$+jets signature, as well as in other channels. Pattern recognition represents a major advance in computing in the last two decades. We foresee that its use will be generalized as the complexity of the analyses increases. Such tools however, may not be seen as black boxes, and must be understood before their use.

Another important aspect is background modeling. A good background model is crucial for the quality of an analysis. The data-driven technique we devised, which is an improvement over that of the last iteration of the analysis, is needed for predicting large backgrounds for which it is difficult to generate MC simulations with small

systematic uncertainties; this is particularly true for the large datasets expected from the LHC. As with every technique, it is important to carefully validate it.

The major limitation in this signature consists in the systematic uncertainties associated with W+jets and Z+jets production. We have presented a data-driven approach to evaluate these backgrounds, but additional studies are needed to implement it in this analysis. We remark that this is not the only option to reduce the uncertainties: new event generators which include high order effects could reduce the uncertainties.

We would like to end this discussion with b quark identification. In the last decades, the b quark has played an important role in particle physics, since both the top quark and the Higgs boson (at the mass points favored by the common knowledge at the time of writing) decay into it. Identifying b-jets in a collision is a very powerful way to reduce the backgrounds. The algorithms used in this analysis have an efficiency of the order of 40-50%. An improvement in these figures could lead to a significant gain in sensitivity. We look forward the implementation of the HOBIT tagger in the next iteration of this analysis.

Finally, all the techniques presented in this dissertation will be useful in analyzing data at the LHC, whether to search for the Higgs or new physics scenarii. Whether the Higgs exists as we expect or not, we have exciting times ahead of us in our quest to understand nature.

LIST OF REFERENCES

LIST OF REFERENCES

[1] B. Pontecorvo. Neutrino Experiments and the Problem of Conservation of Leptonic Charge. *Sov.Phys.JETP*, 26:984–988, 1968.

[2] J.W.F. Valle. Neutrino physics overview. *J.Phys.Conf.Ser.*, 53:473–505, 2006.

[3] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957.

[4] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[5] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

[7] K. Nakamura et al. Review of particle physics. *J. Phys.*, G37:075021, 2010.

[8] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory (Frontiers in Physics)*. Perseus Books, 2008.

[9] Leon N. Cooper. Bound electron pairs in a degenerate fermi gas. *Phys. Rev.*, 104:1189–1190, Nov 1956.

[10] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Microscopic theory of superconductivity. *Phys. Rev.*, 106:162–164, Apr 1957.

[11] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Theory of superconductivity. *Phys. Rev.*, 108:1175–1204, Dec 1957.

[12] Paul A.M. Dirac. The Lagrangian in quantum mechanics. *Phys.Z.Sowjetunion*, 3:64–72, 1933.

[13] R. P. Feynman. Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.*, 20:367–387, Apr 1948.

[14] Emmy Noether. Invariant Variation Problems. *Gott.Nachr.*, 1918:235–257, 1918.

[15] E. Fermi. An attempt of a theory of beta radiation. 1. *Z.Phys.*, 88:161–177, 1934.

[16] R. P. Feynman and M. Gell-Mann. Theory of the fermi interaction. *Phys. Rev.*, 109:193–198, Jan 1958.

[17] E. C. G. Sudarshan and R. E. Marshak. Chirality invariance and the universal fermi interaction. *Phys. Rev.*, 109:1860–1862, Mar 1958.

[18] F.J. Hasert et al. Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment. *Phys.Lett.*, B46:138–140, 1973.

[19] F.J. Hasert et al. Observation of Neutrino Like Interactions without Muon or Electron in the Gargamelle Neutrino Experiment. *Nucl.Phys.*, B73:1–22, 1974.

[20] F.J. Hasert et al. Search for elastic muon-neutrino electron scattering. *Physics Letters B*, 46(1):121 – 124, 1973.

[21] S.L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961.

[22] Francis Halzen and Alan Martin. *Quarks & Leptons: An introductory course in modern particle physics*. John Wiley & Sons, New York, USA, 1984.

[23] P. Bagnaia and others (UA2 Collaboration). Evidence for $z^0 \to e^+e^-$ at the cern pp collider. *Physics Letters B*, 129(1-2):130 – 140, 1983.

[24] G. Arnison and others (UA1 Collaboration). Experimental observation of lepton pairs of invariant mass around 95 gev/c2 at the cern sps collider. *Physics Letters B*, 126(5):398 – 410, 1983.

[25] G. Arnison and others (UA1 Collaboration). Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 gev. *Physics Letters B*, 122(1):103 – 116, 1983.

[26] M. Banner and others (UA2 Collaboration). Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the cern pp collider. *Physics Letters B*, 122(5-6):476 – 485, 1983.

[27] Siegfried Bethke. The 2009 World Average of alpha(s). *Eur.Phys.J.*, C64:689–703, 2009.

[28] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.

[29] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.

[30] Tevatron Electroweak Working Group. Combination of CDF and DØ results on the mass of the top quark using up to 5.8 fb^{-1} of data. 2011.

[31] T. Aaltonen et al. First Observation of Electroweak Single Top Quark Production. *Phys.Rev.Lett.*, 103:092002, 2009.

[32] V. M. Abazov et al. Observation of Single Top Quark Production. 2009.

[33] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[34] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963.

[35] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev.*, D2:1285–1292, 1970.

[36] Donald H. Perkins. *Introduction to High Energy Physics*. Cambridge University Press, April 2000.

[37] Ling-Lie Chau and Wai-Yee Keung. Comments on the Parametrization of the Kobayashi-Maskawa Matrix. *Phys. Rev. Lett.*, 53:1802, 1984.

[38] C. Amsler et al. Review of particle physics. *Phys. Lett.*, B667:1, 2008.

[39] R.P. Feynman. The behavior of hadron collisions at extreme energies. *Conf.Proc.*, C690905:237–258, 1969.

[40] Pavel M. Nadolsky, Hung-Liang Lai, Qing-Hong Cao, Joey Huston, Jon Pumplin, et al. Implications of CTEQ global analysis for collider observables. *Phys.Rev.*, D78:013004, 2008.

[41] H. L. Lai et al. Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions. *Eur. Phys. J.*, C12:375–392, 2000.

[42] Hung-Liang Lai et al. New parton distributions for collider physics. *Phys. Rev.*, D82:074024, 2010.

[43] A. Martin et al. Parton distributions: A New global analysis. *Eur. Phys. J.*, C4:463–496, 1998.

[44] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Update of parton distributions at NNLO. *Phys.Lett.*, B652:292–299, 2007.

[45] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.J.*, C63:189–285, 2009.

[46] R.S. Thorne, A.D. Martin, W.J. Stirling, and G. Watt. Status of MRST/MSTW PDF sets. 2009.

[47] R.S. Thorne, A.D. Martin, and W.J. Stirling. MRST parton distributions: Status 2006. pages 81–84, 2006.

[48] James M. Zanotti. Hadronic structure from lattice QCD. *PoS*, LC2008:051, 2008.

[49] A. Brignole, J.R. Espinosa, M. QuirÃ³s, and F. Zwirner. Aspects of the electroweak phase transition in the minimal supersymmetric standard model. *Physics Letters B*, 324(2):181 – 191, 1994.

[50] Thomas Hambye and Kurt Riesselmann. Matching conditions and Higgs mass upper bounds revisited. *Phys.Rev.*, D55:7255–7262, 1997.

[51] Thomas Hambye and Kurt Riesselmann. SM Higgs mass bounds from theory. 1997.

[52] William J. Marciano. Precision electroweak measurements and the Higgs mass. *eConf*, C040802:L009, 2004.

[53] The lep electroweak working group.

[54] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys. Lett.*, B565:61–75, 2003.

[55] Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb-1 of Data. 2011. * Temporary entry *.

[56] Tom Junk, Eric James, Ben Kilminster, Craig Group, and Weiming Yao. Cdf/pub/exotic/public/10609. Combination of CDF's Searches for the Standard Model Higgs boson with up to 8.2 fb^{-1} of data.

[57] The CMS and ATLAS collaborations. Combined Standard Model Higgs boson searches with up to 2.3 fb^{-1} of pp collision data at $\sqrt{s} = 7$ TeV at the LHC. CMS PAS HIG-11-023, ATLAS-CONF-2011-157.

[58] E. Komatsu et al. Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *Astrophys.J.Suppl.*, 180:330–376, 2009.

[59] Fermilab mission: http://www.fnal.gov/pub/about/whatis/mission.html.

[60] Fermilab Beams Division. The numi rookie book. http://www-bdnew.fnal.gov/operations/rookie_books/NuMI_v1.pdf, September 2006.

[61] Fermilab history: http://www.fnal.gov/pub/about/whatis/history.html.

[62] J. D. Cockcroft and E. T. S. Walton. Experiments with High Velocity Positive Ions. II. The Disintegration of Elements by High Velocity Protons. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character (1905-1934)*, 137(831):229–242, July 1932.

[63] Fermilab Beams Division. The linac rookie book. http://www-bdnew.fnal.gov/operations/rookie_books/LINAC_v2.pdf, July 2006.

[64] Fermilab Beams Division. The booster rookie book. http://www-bdnew.fnal.gov/operations/rookie_books/Booster_V4.1.pdf, April 2009.

[65] Fermilab Beams Division. The miniboone mini-rookie book. http://www-bdnew.fnal.gov/operations/rookie_books/MiniBooNE_v1.pdf, September 2006.

[66] Fermilab Beams Division. The main injector rookie book. http://www-bdnew.fnal.gov/operations/rookie_books/Main_Injector_v1.1.pdf, April 2010.

[67] K. Seiya, T. Berenc, J. Dey, B. Chase, C. Rivetta, et al. Status of slip stacking at Fermilab Main Injector. *Conf.Proc.*, C0505161:347, 2005.

[68] http://www.fnal.gov/pub/news04/update_archive/update_11-10.html. Technical report.

[69] Fermilab Beams Division. The antiproton source rookie book. http://www-bdnew.fnal.gov/operations/rookie_books/Pbar_v2.2.pdf, June 2011.

[70] John Marriner. Stochastic cooling overview. *Nucl.Instrum.Meth.*, A532:11–18, 2004.

[71] S. van der Meer. Stochastic cooling and the accumulation of antiprotons. *Rev. Mod. Phys.*, 57:689–697, Jul 1985.

[72] Fermilab Beams Division. The recycler rookie book. http://www-bdnew.fnal.gov/operations/rookie_books/Recycler_RB_v1.42.pdf, August 2010.

[73] Fermilab Beams Division. The tevatron rookie book. http://www-bdnew.fnal.gov/operations/rookie_books/Tevatron_v2.3.pdf, July 2009.

[74] R. Blair et al. The CDF-II detector: Technical design report. FERMILAB-PUB-96-390-E.

[75] P. Azzi. The CDF silicon detector upgrade. *Nucl. Instrum. Meth.*, A419:532–537, 1998.

[76] A. Sill. CDF Run II silicon tracking projects. *Nucl. Instrum. Meth.*, A447:1–8, 2000.

[77] Christopher S. Hill. Operational experience and performance of the CDFII silicon detector. *Nucl. Instrum. Meth.*, A530:1–6, 2004.

[78] Anthony A. Affolder et al. Intermediate silicon layers detector for the CDF experiment. *Nucl. Instrum. Meth.*, A453:84–88, 2000.

[79] Bill Ashmanskas et al. The CDF silicon vertex trigger. *Nucl.Instrum.Meth.*, A518:532–536, 2004.

[80] Anthony Allen Affolder et al. CDF central outer tracker. *Nucl. Instrum. Meth.*, A526:249–299, 2004.

[81] L. Balka et al. The CDF Central Electromagnetic Calorimeter. *Nucl. Instrum. Meth.*, A267:272, 1988.

[82] S. Kuhlmann et al. The CDF calorimetry upgrade for run two *b*. *Nucl. Instrum. Meth.*, A518:39–41, 2004.

[83] G. Apollinari, Konstantin Goulianos, P. Melese, and M. Lindgren. Shower maximum detector for the CDF plug upgrade calorimeter. *Nucl. Instrum. Meth.*, A412:515–526, 1998.

[84] S. Bertolucci et al. The CDF Central and Endwall Hadron Calorimeter. *Nucl. Instrum. Meth.*, A267:301, 1988.

[85] G. Ascoli et al. CDF Central Muon Detector. *Nucl. Instrum. Meth.*, A268:33, 1988.

[86] C. M. Ginsburg. CDF Run 2 muon system. *Eur. Phys. J.*, C33:s1002–s1004, 2004.

[87] D. Acosta et al. A time-of-flight detector in CDF-II. *Nucl. Instrum. Meth.*, A518:605–608, 2004.

[88] D. Acosta et al. The CDF Cherenkov luminosity monitor. *Nucl. Instrum. Meth.*, A461:540–544, 2001.

[89] D. Acosta et al. The performance of the CDF luminosity monitor. *Nucl. Instrum. Meth.*, A494:57–62, 2002.

[90] S. Holm et al. System architecture and hardware design of the CDF XFT online track processor. *IEEE Trans.Nucl.Sci.*, 47:895–902, 2000.

[91] A. Abulencia et al. The CDF II eXtremely fast tracker upgrade. *Nucl.Instrum.Meth.*, A572:358–360, 2007.

[92] Robert Downing, Nathan Eddy, Lee Holloway, Mike Kasten, Hyunsoo Kim, et al. Track extrapolation and distribution for the CDF-II trigger system. *Nucl.Instrum.Meth.*, A570:36–50, 2007.

[93] Jahred A. Adelman et al. The Silicon Vertex Trigger upgrade at CDF. *Nucl. Instrum. Meth.*, A572:361–364, 2007.

[94] S. Amerio, A. Annovi, M. Bettini, M. Bucciantonio, P. Catastini, et al. The GigaFitter: Performance at CDF and perspectives for future applications. *J.Phys.Conf.Ser.*, 219:022001, 2010. No Electronic Copy.

[95] J. Antos, M. Babik, D. Benjamin, S. Cabrera, A.W. Chan, et al. Data production models for the CDF experiment. 2006.

[96] R. Colombo et al. The CDF computing and analysis system: First experience. pages 15–19, 2001.

[97] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[98] P. Murat. Stntuple: Root-based microdst analysis system for cdf. In *CHEP 2011 Proceedings*. Science Press, 2011.

[99] F.D. Snider. Tracking at CDF: algorithms and experience from Run I and Run II. *Nucl.Instrum.Meth.*, A566:133–141, 2006.

[100] T.K. Nelson. The CDF-II silicon tracking system. *Nucl.Instrum.Meth.*, A485:1–5, 2002.

[101] Christopher Paul Hays, Y. Huang, A.V. Kotwal, H.K. Gerberich, S. Menzemer, et al. Inside-out tracking at CDF. *Nucl.Instrum.Meth.*, A538:249–254, 2005.

[102] D. Acosta et al. Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using lepton + jets events with secondary vertex b-tagging. *Phys. Rev. D*, 71(5):052003, 2005.

[103] Adrian Buzatu. Search for the Standard Model Higgs boson produced in association with a W Boson in the isolated-track charged-lepton channel using the Collider Detector at Fermilab. 2011. PhD. Thesis (advisor: Andreas Warburton).

[104] Pierluigi Totaro. Search for standard model Higgs boson in di-tau final state at the Tevatron. *PoS*, ICHEP2010:080, 2010.

[105] Elisabetta Pianori. Searches For The Standard Model Higgs Boson in final states with tau leptons at the Tevatron. 2011.

[106] Gerald C. Blazey, Jay R. Dittmann, Stephen D. Ellis, V. Daniel Elvira, K. Frame, S. Grinstein, Robert Hirosky, R. Piegaia, H. Schellman, R. Snihur, V. Sorin, and Dieter Zeppenfeld. Run ii jet physics: Proceedings of the run ii qcd and weak boson physics workshop. 2000. cite arxiv:hep-ex/0005012 Comment: These proceedings include a recommendation for a standard jet cone and a standard jet KT algorithm for Run II. The algorithms are suitable for both CDF and D0 jet reconstruction as well as theoretical calculations. 32 Pages, 38 Figures.

[107] F. Abe et al. The Topology of three jet events in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev.*, D45:1448–1458, 1992.

[108] A. Bhatti et al. Determination of the jet energy scale at the collider detector at Fermilab. *Nucl. Instrum. Meth.*, A566:375–412, 2006.

[109] Gerald C. Blazey, Jay R. Dittmann, Stephen D. Ellis, V.Daniel Elvira, K. Frame, et al. Run II jet physics. pages 47–77, 2000.

[110] CDF Jet Energy and Resolution Group. Generic jet corrections for run ii. CDF/PHYS/JET/7358.

[111] Torbjorn Sjostrand, Leif Lonnblad, and Stephen Mrenna. PYTHIA 6.2: Physics and manual. 2001.

[112] G. Abbiendi et al. Search for neutral Higgs bosons in e+ e- collisions at S**(1/2) approximately equals 189-GeV. *Eur.Phys.J.*, C12:567–586, 2000.

[113] C. Adloff et al. Diffraction dissociation in photoproduction at HERA. *Z. Phys.*, C74:221–236, 1997.

[114] Artur Apresyan. *Search for the Standard Model Higgs boson in the Missing Transverse Energy and b-jet signature in proton-antiproton collisions at a center of mass energy of 1.96 TeV*. PhD thesis, Purdue University, 2009. FERMILAB-THESIS-2009-09.

[115] T. Aaltonen et al. Measurement of the $t\bar{t}$ Production Cross Section in 2 fb^{-1} of $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using Lepton Plus Jets Events with Soft Muon b-Tagging. *Phys.Rev.*, D79:052007, 2009.

[116] D. Acosta et al. Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using lepton plus jets events with semileptonic B decays to muons. *Phys.Rev.*, D72:032002, 2005.

[117] T. Affolder et al. Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys.Rev.*, D64:032002, 2001.

[118] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 0101:010, 2001.

[119] A. Abulencia et al. Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ *tev* using *lepton + jets* events with jet probability *b*-tagging. *Phys. Rev. D*, 74(7):072006, Oct 2006.

[120] Svenja Richter. Search for electroweak single top-quark production with the CDF II experiment. 2007. Ph.D. Thesis (Advisors: Feindt, M., Wagner, W.).

[121] Miguel Vidal Marono. Search for Third Generation Squarks in the Missing Transverse Energy plus Jet Sample at CDF Run II. 2010. Ph.D. Thesis (Advisor: Oscar Gonzalez).

[122] J. Freeman, J.D. Lewis, W. Ketchum, S. Poprocki, A. Pronko, et al. An Artificial neural network based *b* jet identification algorithm at the CDF Experiment. 2011.

[123] Paolo Mastrandrea. Study of the heavy flavour fractions in z+jets events from proton-antiproton collisions at energy = 1.96 TeV with the CDF II detector at the Tevatron collider. 2008. Ph.D. Thesis (advisors: Angelo Scribano, Carlo Dionisi).

[124] V.M. Abazov et al. b-Jet Identification in the D0 Experiment. *Nucl.Instrum.Meth.*, A620:490–517, 2010.

[125] John Freeman, Mike Kirby Thomas Junk, Thomas Phillips, Rick Snider, Marco Trovato, Jesus Vizan, and Wei-Ming Yao. Hobit: Higgs optimized *b*-identification tagger. CDF/DOC/CDF/CDFR/10488.

[126] John C. Collins, Davison E. Soper, and George Sterman. Soft gluons and factorization. *Nuclear Physics B*, 308(4):833 – 856, 1988.

[127] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv.Ser.Direct.High Energy Phys.*, 5:1–91, 1988.

[128] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys.Rept.*, 97:31–145, 1983.

[129] Torbjorn and Sjostrand. Jet fragmentation of multiparton configurations in a string framework. *Nuclear Physics B*, 248(2):469 – 502, 1984.

[130] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 0307:001, 2003.

[131] Michelangelo L. Mangano, Mauro Moretti, and Roberto Pittau. Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b} + n$ jets as a case study. *Nucl.Phys.*, B632:343–362, 2002.

[132] D. Amati and G. Veneziano. Preconfinement as a Property of Perturbative QCD. *Phys.Lett.*, B83:87, 1979.

[133] Fabio Maltoni and Tim Stelzer. MadEvent: Automatic event generation with MadGraph. *JHEP*, 0302:027, 2003.

[134] Johan Alwall et al. MadGraph/MadEvent v4: The New Web Generation. *JHEP*, 09:028, 2007.

[135] T. Stelzer and W.F. Long. Automatic generation of tree level helicity amplitudes. *Comput.Phys.Commun.*, 81:357–371, 1994.

[136] S. Catani, F. Krauss, R. Kuhn, and B.R. Webber. QCD matrix elements + parton showers. *JHEP*, 0111:063, 2001.

[137] Michel Herquet and Fabio Maltoni. MadGraph/MadEvent : A multipurpose event generator. *Nucl.Phys.Proc.Suppl.*, 179-180:211–217, 2008.

[138] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 0411:040, 2004.

[139] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 0711:070, 2007. * Temporary entry *.

[140] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 0206:029, 2002.

[141] Stefan Hoche, Frank Krauss, Marek Schonherr, and Frank Siegert. Automating the POWHEG method in Sherpa. *JHEP*, 1104:024, 2011.

[142] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Mcfm: A monte carlo for femtobarn processes at hadron colliders - users guide. http://mcfm.fnal.gov/mcfm.pdf.

[143] John M. Campbell and R. Keith Ellis. Radiative corrections to Z b anti-b production. *Phys. Rev.*, D62:114012, 2000.

[144] Rene Brun, Federico Carminati, and Simone Giani. GEANT Detector Description and Simulation Tool. 1994. CERN Program Library Long Writeup.

[145] R. Veenhof. GARFIELD, recent developments. *Nucl.Instrum.Meth.*, A419:726–730, 1998.

[146] Guenter Grindhammer, M. Rudowicz, and S. Peters. THE FAST SIMULATION OF ELECTROMAGNETIC AND HADRONIC SHOWERS. *Nucl.Instrum.Meth.*, A290:469, 1990.

[147] S. Agostinelli et al. GEANT4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.

[148] Viktor Veszpremi. *Search for the standard model Higgs boson in proton- antiproton collisions at a center-of-mass energy of 1.96 TeV*. PhD thesis, Purdue University, 2007. FERMILAB-THESIS-2007-30.

[149] D.Acosta, A.Kovalev, S.Lammel, T.Munar, C.Rott, D.Tsybychev, and S.M.Wang. Clean up of the missing transverse energy datasets. CDF/DOC/EXOTIC/CDFR/7410, 2004.

[150] Julien Baglio and Abdelhak Djouadi. Predictions for Higgs production at the Tevatron and the associated uncertainties. *JHEP*, 1010:064, 2010.

[151] Gregory Mahlon and Stephen J. Parke. Improved spin basis for angular correlation studies in single top quark production at the Tevatron. *Phys.Rev.*, D55:7249–7254, 1997.

[152] Gregory Mahlon and Stephen J. Parke. Single top quark production at the LHC: Understanding spin. *Phys.Lett.*, B476:323–330, 2000.

[153] Douglas O. Carlson and C.P. Yuan. Probing new physics from the single top production. 1995.

[154] Christopher T. Hill and Stephen J. Parke. Top production: Sensitivity to new physics. *Phys.Rev.*, D49:4454–4462, 1994.

[155] Nikolaos Kidonakis. Single top production at the Tevatron: Threshold resummation and finite-order soft gluon corrections. *Phys. Rev.*, D74:114012, 2006.

[156] Nikolaos Kidonakis. Single top quark production cross section at hadron colliders. *PoS*, DIS2010:196, 2010.

[157] B.W. Harris, Eric Laenen, L. Phaf, Z. Sullivan, and S. Weinzierl. The Fully differential single top quark cross-section in next to leading order QCD. *Phys.Rev.*, D66:054024, 2002.

[158] Nikolaos Kidonakis. Higher-order corrections to top-antitop pair and single top quark production. 2009.

[159] E. Boos et al. Method for simulating electroweak top-quark production events in the NLO approximation: SingleTop event generator. *Phys. Atom. Nucl.*, 69:1317–1329, 2006.

[160] Zack Sullivan. Understanding single-top-quark production and jets at hadron colliders. *Phys. Rev.*, D70:114012, 2004.

[161] T. Aaltonen et al. Search for single top quark production in pbar p collisions at sqrts=1.96 TeV in the missing transverse energy plus jets topology. *Phys.Rev.*, D81:072003, 2010.

[162] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.

[163] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126:298, 1977.

[164] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.

[165] Jan Lueck. Observation of Electroweak Single Top-Quark Production with the CDF II Experiment. 2009. Ph.D. Thesis (Advisor: Thomas Muller, Wolfgang Wagner).

[166] Zhenbin Wu, Jan Lueck, and Dominic Hirschbuehl. Powheg signal samples used in single top analysis. CDF/ANAL/TOP/CDFR/10325.

[167] T. Aaltonen et al. Observation of Single Top Quark Production and Measurement of |Vtb| with CDF. *Phys.Rev.*, D82:112005, 2010.

[168] John M. Campbell, R.Keith Ellis, and Francesco Tramontano. Single top production and decay at next-to-leading order. *Phys.Rev.*, D70:094012, 2004.

[169] S. Moch and P. Uwer. Heavy-quark pair production at two loops in QCD. *Nucl.Phys.Proc.Suppl.*, 183:75–80, 2008.

[170] Combination of CDF and D0 Results on the Mass of the Top Quark. 2009.

[171] Matteo Cacciari, Stefano Frixione, Michelangelo L. Mangano, Paolo Nason, and Giovanni Ridolfi. Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC. *JHEP*, 09:127, 2008.

[172] Nikolaos Kidonakis and Ramona Vogt. The Theoretical top quark cross section at the Tevatron and the LHC. *Phys. Rev.*, D78:074005, 2008.

[173] T. Aaltonen et al. Top-quark mass measurement using events with missing transverse energy and jets at CDF. 2011. Long author list - awaiting processing.

[174] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA, 1995.

[175] Mr. Bayes and Mr. Price. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter to john canton, a. m. f. r. s. *Philosophical Transactions*, 53:370–418, 1763.

[176] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Netw.*, 2(5):359–366, 1989.

[177] Yoshifusa Ito. Approximation capability of layered neural networks with sigmoid units on two layers. *Neural Comput.*, 6:1233–1243, November 1994.

[178] C. Lee Giles and Tom Maxwell. Learning, invariance, and generalization in high-order neural networks. *Appl. Opt.*, 26(23):4972–4978, Dec 1987.

[179] Jonathan Richard Shewchuk. An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. August 1994.

[180] Donald Goldfarb. A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, 24(109):23–26, January 1970.

[181] R. Fletcher. A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322, 1970.

[182] D. F. Shanno. Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24(111):647–656, 1970.

[183] C. G. BROYDEN. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, 1970.

[184] Andreas Hocker et al. TMVA: Toolkit for multivariate data analysis. *PoS*, ACAT:040, 2007.

[185] Jon Wilson, Homer Wolfe, Jake Connors, Richard Hughes, and Brian Winer. Search for the Higgs Boson Produced in Association with Top Quarks Using 7.5 fb $^{-1}$. CDF/PUB/EXOTIC/PUBLIC/10574, 2011.

[186] Kenneth O. Stanley and Risto Miikkulainen. Evolving neural networks through augmenting topologies. *Evol. Comput.*, 10:99–127, June 2002.

[187] P. Speckmayer, A. Hocker, J. Stelzer, and H. Voss. The toolkit for multivariate data analysis, TMVA 4. *J.Phys.Conf.Ser.*, 219:032057, 2010.

[188] O. Gonzalez. The met triggers after the l2-cal upgrade. CDF/DOC/TRIGGER/GROUP/9030, February 2008.

[189] O.Gonzalez and M.Vidal. Report for the official dataset to recover events in the missing-et sample. CDF/ANAL/EXOTIC/CDFR/9231, 2008.

[190] T. Aaltonen et al. A Search for the Higgs Boson Using Neural Networks in Events with Missing Energy and b-quark Jets in p anti-p Collisions at s**(1/2) = 1.96-TeV. *Phys.Rev.Lett.*, 104:141801, 2010.

[191] O.Gonzalez and M.Vidal. Trigger efficiency of the met+jets sample for the 2 fb^{-1} dataset. CDF/PHYS/MISSING_ET/CDFR/9355, 2008.

[192] O.Gonzalez and M.Vidal. Getting ready for the 4 fb-1 dataset: Trigger efficiency of the met+jets samples. CDF/PHYS/MISSING_ET/CDFR/9745, 2009.

[193] J. Antos et al. Data processing model for the CDF experiment. *IEEE Trans. Nucl. Sci.*, 53:2897–2906, 2006.

[194] Gunter Zech. Frequentist and Bayesian confidence limits. *Eur.Phys.J.direct*, C4:12, 2002.

[195] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl.Instrum.Meth.*, A434:435–443, 1999.

[196] T. Junk. Sensitivity, exclusion and discovery with small signals, large backgrounds, and large systematic uncertainties. CDF/DOC/STATISTICS/PUBLIC/8128, October 2007.

[197] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[198] W.K. Hastings. Monte carlo samping methods using markov chains and their applications. *Biometrika*, pages 97–109, 1970.

[199] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur.Phys.J.*, C71:1554, 2011.

[200] M. Cacciari et al. The t anti-t cross-section at 1.8-TeV and 1.96-TeV: A study of the systematics due to parton densities and scale dependence. *JHEP*, 04:068, 2004.

[201] A. Abulencia et al. Measurement of the b jet cross-section in events with a Z boson in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96- TeV. *Phys. Rev.*, D74:032008, 2006.

[202] John M. Campbell and R. Keith Ellis. An update on vector boson pair production at hadron colliders. *Phys. Rev.*, D60:113006, 1999.

[203] Torbjorn Sjostrand. A Model for Initial State Parton Showers. *Phys. Lett.*, B157:321, 1985.

[204] Mats Bengtsson, Torbjorn Sjostrand, and Maria van Zijl. INITIAL STATE RADIATION EFFECTS ON W AND JET PRODUCTION. *Z. Phys.*, C32:67, 1986.

[205] Marcela S. Carena et al. Report of the Tevatron Higgs working group, Part 6. 2000.

[206] Brandon Scott Parks. Search for the Higgs Boson in the $ZH \to \nu\nu b\bar{b}$ Channel at CDF Run II. 2008. Ph.D. Thesis (advisors: Brian L. Winer, Richard Hughes).

[207] Tevatron Electroweak Working Group. Combination of CDF and D0 Measurements of the Single Top Production Cross Section. 2009.

[208] Daniela Bortoletto, Qiuguang Liu, Fabrizio Margaroli, and Karolos Potamianos. Measuring the Top Pair Background to the Higgs Search in the MET+jets Channel with L=5.7 fb^{-1}. CDF/ANAL/TOP/PUBLIC/10237.

[209] T. Aaltonen et al. Measurement of the $WW + WZ$ Production Cross Section Using a Matrix Element Technique in Lepton + Jets Events. *Phys.Rev.*, D82:112001, 2010.

[210] T. Aaltonen et al. Measurement of the WW+WZ Production Cross Section Using the Lepton+Jets Final State at CDF II. *Phys.Rev.Lett.*, 104:101801, 2010.

[211] T. Aaltonen et al. Search for WW and WZ production in lepton plus jets final state at CDF. *Phys.Rev.*, D79:112011, 2009.

[212] T. Aaltonen et al. Search for WW and WZ resonances decaying to electron, missing E_T, and two jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys.Rev.Lett.*, 104:241801, 2010.

[213] Johan Alwall et al. Is V(tb) = 1? *Eur. Phys. J.*, C49:791–801, 2007.

[214] Timothy M. P. Tait and C. P. Yuan. Single top quark production as a window to physics beyond the standard model. *Phys. Rev.*, D63:014018, 2001.

[215] V. M. Abazov et al. Evidence for production of single top quarks and first direct measurement of $|v_{tb}|$. *Phys. Rev. Lett.*, 98:181802, 2007.

[216] V.M. Abazov et al. Evidence for production of single top quarks. *Phys. Rev.*, D 78:012005, 2008.

[217] T. Aaltonen et al. Measurement of the Single Top Quark Production Cross Section at CDF. *Phys. Rev. Lett.*, 101:252001, 2008.

[218] D. W. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley, and B. W. Suter. The multilayer perceptron as an approximation to a bayes optimal discriminant function. *Neural Networks, IEEE Transactions on*, 1(4):296–298, 1990.

SEARCH FOR THE HIGGS BOSON AND RARE STANDARD MODEL PROCESSES IN THE $\not\!\!E_T$+B-JETS SIGNATURE AT THE COLLIDER DETECTOR AT FERMILAB

VOLUME 2

APPENDICES

A. THE $\not\!\!E_T$ AND AUXILIARY TRIGGERS

This appendix presents details on the various triggers used in this analysis. We first describe the $\not\!\!E_T$ triggers used to collect the events, and proceed with the auxiliary triggers used for trigger studies or event recovery.

A.1 $\not\!\!E_T$ triggers

The first MET35+JETS trigger path, MET35_&_TWO_JETS was implemented in July 2002. It required

L1 $\not\!\!E_T > 25$ GeV;

L2 Two clusters with $E_T > 10$ GeV;

L3 $\not\!\!E_T > 25$ GeV.

However, because of the increase in instantaneous luminosity, the L2 trigger rates were increasing. Therefore, this trigger had to be prescaled[1]. In March 2005, when the instantaneous luminosity was routinely above $\cdot 10^{32}$ cm^{-2} s^{-1}, the trigger path was revised, leading to the MET_&_CJET_&_JET path, which required

L1 $\not\!\!E_T > 25$ GeV;

L2 Two clusters with $E_T > 10$ GeV, one of which is central ($|\eta < 1.1$);

L3 $\not\!\!E_T > 35$ GeV.

Because of continuous rise in instantaneous luminosity, this trigger was disabled when the instantaneous luminosity was above $1.9 \cdot 10^{32}$ cm^{-2} s^{-1}, and from April 2007 on, it operated with a dynamic prescale[2].

[1] A trigger with a prescale P records only a fraction $1/P$ of events.
[2] A dynamic prescale system has a feedback loop that adjusts the rate of the presecale according to the total trigger rate.

The MET+DIJET path was introduced at the end of 2007 to cope with the high luminosity conditions; it is not prescaled. It requires

L1 $\not\!\!E_T > 15$ GeV and one cluster with $ET_{>}10$ GeV;

L2 $\not\!\!E_T > 28$ GeV and two clusters with $E_T > 10$ GeV;

L3 $\not\!\!E_T > 30$ GeV.

As mentioned in chapter 10, we also use the MET45 trigger path, which has the following requirements

L1 $\not\!\!E_T > 25$ GeV;

L2 $\not\!\!E_T > 35$ GeV;

L3 $\not\!\!E_T > 45$ GeV.

The L3 requirement was later relaxed to 40 GeV, but we keep referring to it as MET45 for simplicity.

Two additional $\not\!\!E_T$ triggers were introduced, but are not directly used for this analysis. The first of these is the MET+BJET path, which requires

L1 $\not\!\!E_T > 15$ GeV, one cluster with $\not\!\!E_T > 5$ GeV, and two tracks with $p_T > 2$ GeV/c;

L2 Two displaced tracks with $p_T > 2$ GeV/c and a minimum angular separation of $\Delta\varphi > 2°$;

L3 $\not\!\!E_T > 20$ GeV and two displaced tracks with $p_T > 2$ GeV/c.

Due to its high rate, this trigger was turned of when the instantaneous luminosity was above $1.5 \cdot 10^{32}$ cm^{-2} s^{-1}. The second of these paths is MET25, which requires

L1 $\not\!\!E_T > 25$ GeV;

L2 A prescale between, which was increased from 100 to 400 with time;

L3 No requirement.

A.2 Auxiliary Triggers

The trigger efficiency studies where performed using the trigger paths described in this section.

The MUON sample of chapter 10 was collected using the MUON_CMUP18 path, which requires

L1 One muon candidate track with $p_T > 6$ GeV/c and a matched track with $p_T > 4$ GeV/c;

L2 One muon candidate track with $p_T > 15$ GeV/c;

L3 One muon candidate track with CMU and CMP hits and $p_T > 18$ GeV/c;

while the JET50 sample is made from events passing the JET50 trigger path, namely

L1 One cluster with $E_T > 5$ GeV, and prescale $P = 50$;

L2 One cluster with $E_T > 50$ GeV, and prescal $P = 2$;

L3 One jet with $E_T > 50$ GeV.

The parameterization described in section 10.5 also used events collected by the JET20 trigger path, requiring

L1 One cluster with $E_T > 5$ GeV, and prescale $P = 50$;

L2 One cluster with $E_T > 15$ GeV, and prescal $P = 25$;

L3 One jet with $E_T > 20$ GeV.

As mentioned in section 10.2, a hardware problem at L1 caused events with a single tower above 127 GeV to be discarded by the trigger, although it should have been recorded. We recover these events using the JET-100 sample, collected by the JET100 trigger:

L1 One cluster with $E_T > 20$ GeV;

L2 One cluster with $E_T > 90$ GeV;

L3 One jet with $E_T > 100$ GeV.

The efficiency for this trigger is 100% for events with $E_T > 127$ GeV at L1.

B. MORE ON THE TRIGGER PARAMETERIZATION

B.1 First Parameterization of the Logical Combination of $\not{E}_T$ Triggers

The method described in section 10.6.2 works as expected and compares well with traditional approaches: we have checked that a network is able to reproduce a known behavior with much less effort (it took one training, versus a complex study to determine regions in which the trigger efficiency could be parameterized in simple terms). In what follows we intend to cope with two limitations of the previous approaches: we strive to model regions with lower efficiency (yet not negligible), and to allow the inclusion of additional triggers.

We now proceed with the validation of a first NN-based parameterization of the trigger efficiency of the logical combination (OR) of the three $\not{E}_T$ trigger paths: MET45 $\oplus$ MET35+JETS $\oplus$ MET+DIJET. To address our second intention, we relax all cuts except the requirement of a central jet[1] and that the two leading jets be separated by $\Delta R \geq 0.8$. All the other cuts are kept as in section 10.6.1. We trained a network with the following configuration 9:11:10:1.

Validation of the Network

When checking this network for consistency, as shown in figures A1 to A4, we remark that the selection is fully efficient for high values of the $\not{E}_T$, irrespective of any other variable; this is caused by the presence of the MET45 path. This as the important side-effect of allowing us to assume a fully-efficient trigger for events with $\not{E}_T \geq 70$, which reduces any issues of the training in regions with lower statistics. But because we want to be un-biased in our comparison with the previous case (in

[1]Relaxing this cut requires additional checks beyond those related to the effect on the trigger.

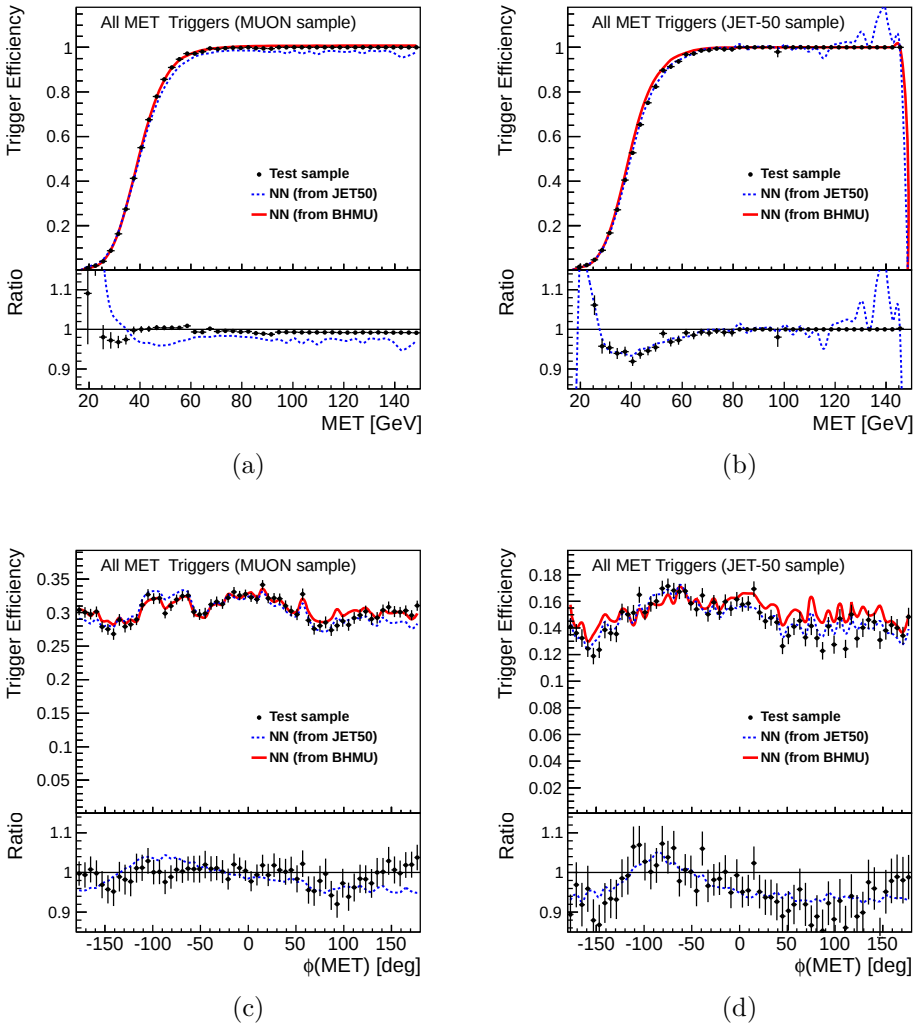


Figure A1.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $\not\!\!E_T$, and (c,d) $\phi(\not\!\!E_T)$.

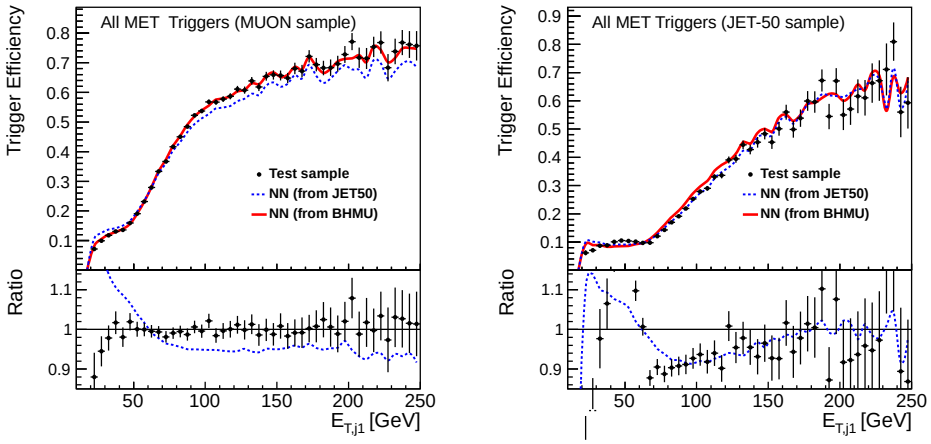


Figure A2.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $E_T^{j_1}$, and (c,d) $E_T^{j_2}$.

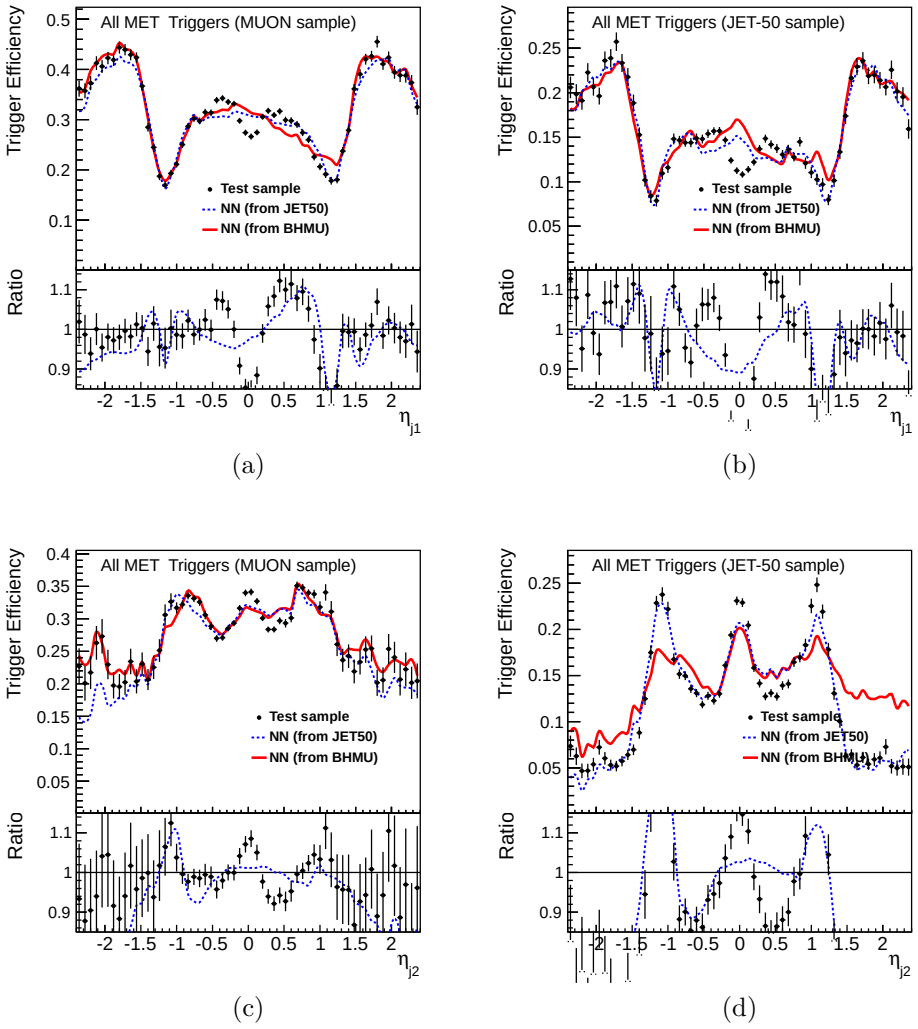


Figure A3.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) η^{j_1}, and (c,d) η^{j_2}.

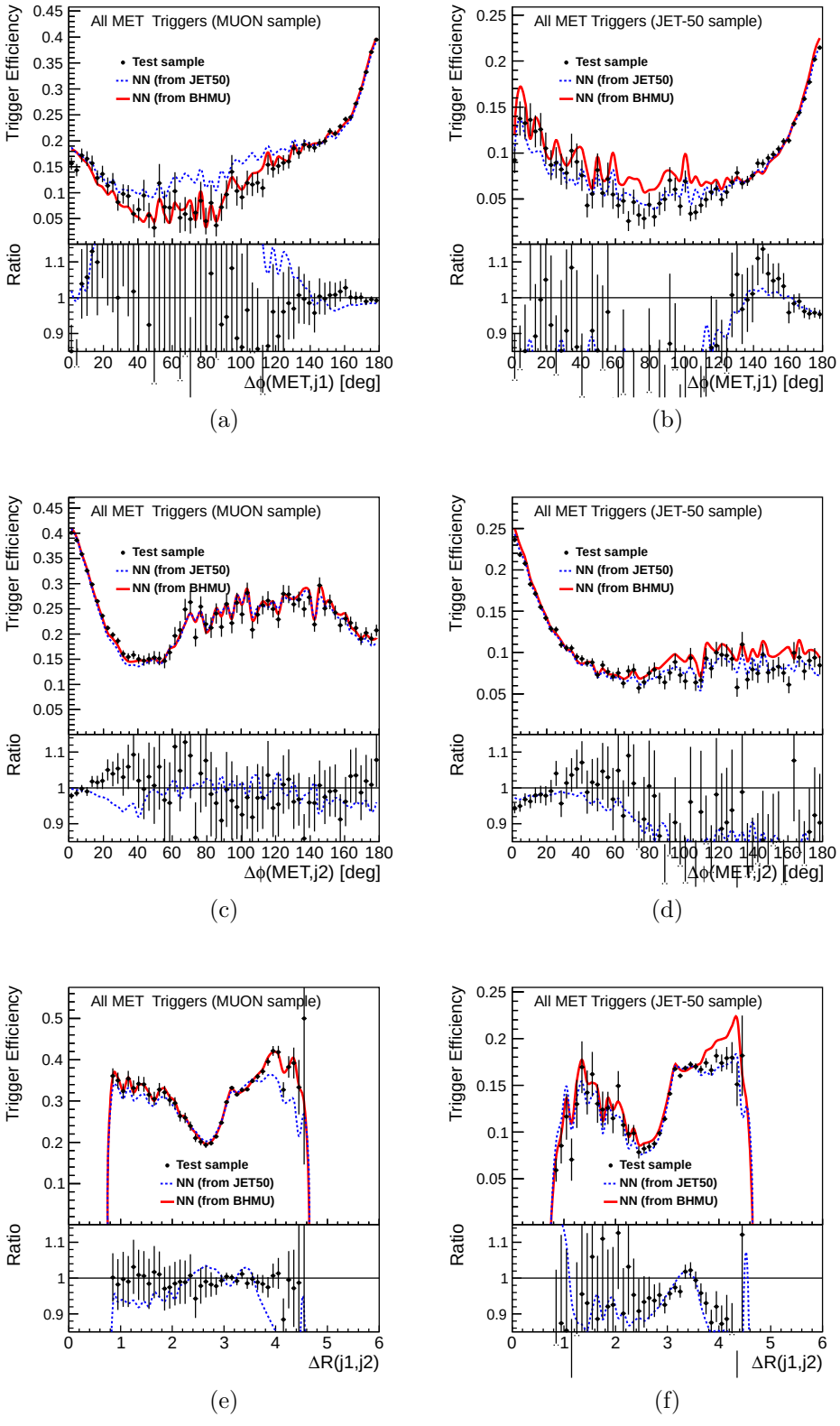


Figure A4.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) $\Delta\varphi(\not{E}_T, j_1)$; (c,d) $\Delta\varphi(\not{E}_T, j_2)$; and (e.f) $\Delta R(j_1, j_2)$.

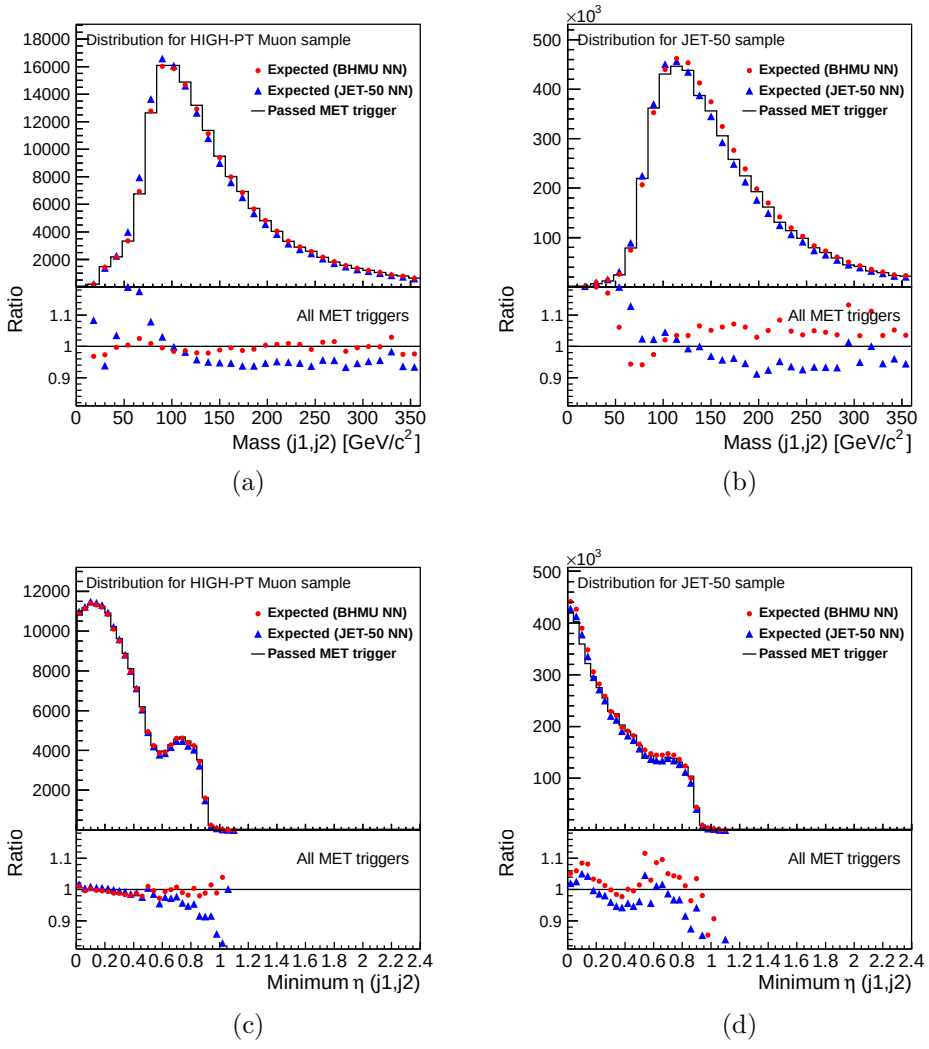


Figure A5.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $m(j_1, j_2)$, and (c,d) $min(\eta^{j_1}, \eta^{j_2})$.

order to estimate the effect of the addition of the MET45 trigger path), we do not use this assumption in this section.

However, we confirm that it is satisfied by the central parameterization, but we couldn't for the alternate case due to the low statistics in the high-$\not\!\!E_T$ tail of the JET-50 sample. Indeed, even though we have pre-selected events to reduce the contribution of events with a low $\not\!\!E_T$ in order to keep the training time low (and since we are not interested in these events anyway), the network has become more complicated and we were forced to reduce the amount of events used in the training to avoid running for weeks. This lead us to train a dedicated network in the region of interest.

We also note a large difference in the response of the two parameterization in the JET-50 sample. This is expected because this sample is biased due to the presence of a trigger jet. This is the motivation behind the use of the MUON sample to derive the central parameterization, and why the uncertainty should be obtained in that (unbiased) sample.

We observe additional discrepancies in the central pseudo-rapidity of the leading jet region. We investigated this effect, but couldn't get it systematically correct. We noted however that it seems dependent on the dimensionality of the training process. Since the addition of more nodes in the hidden layers of the network helps, we decide to increase the number of hidden nodes in the final parameterization. Lacking a conclusive study, we ignore this effect, noting that this variable is not directly related to the cuts we want to emulate. It is likely that this problem is due to differences between events with high and low $\not\!\!E_T$ something which is not expected to appear in the region of high-efficiency, for high-$\not\!\!E_T$ events have a 100% efficiency irrespective of the other variables.

We also checked how several distributions of interest are affected by both parameterizations. These results are shown in figures A6 to A9. In this comparison we have included the efficiency as a function of the invariant mass of the two jets and the efficiency as a function of the pseudo-rapidity of the most central jet (figure A14).

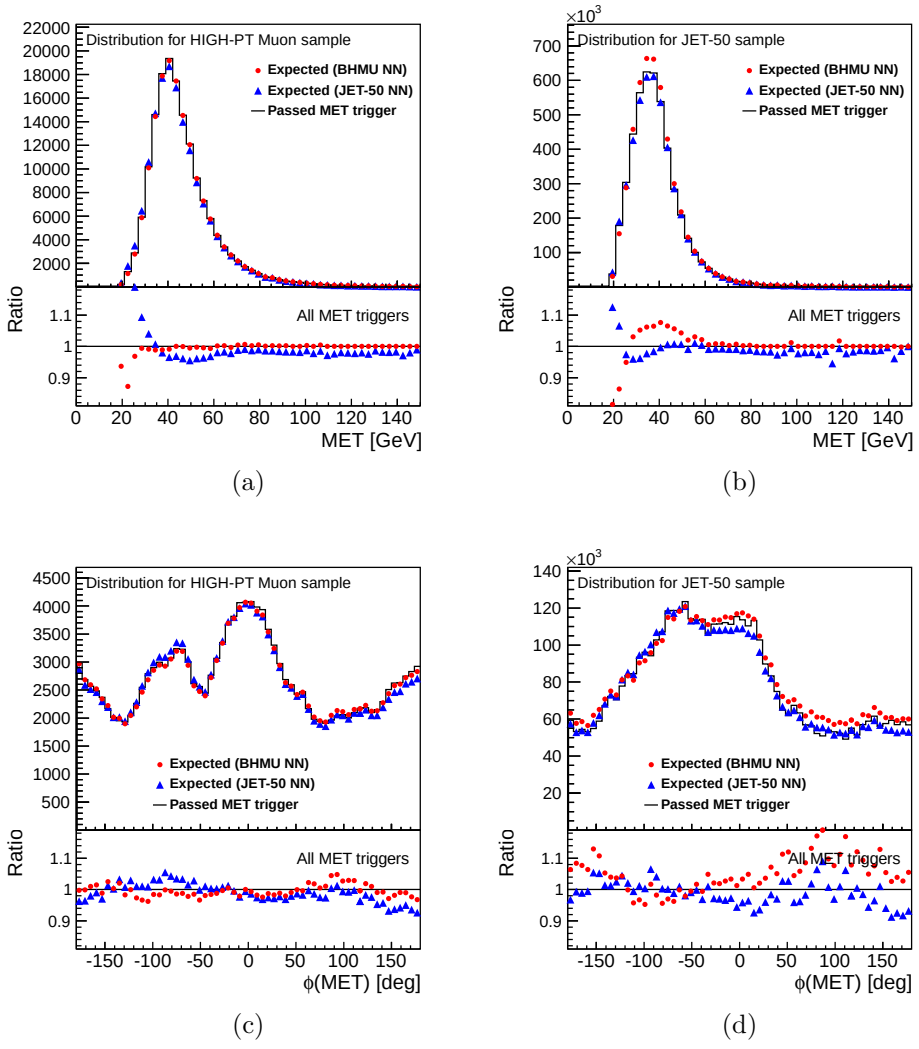


Figure A6.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $\not\!\!E_T$, and (c,d) $\phi(\not\!\!E_T)$.

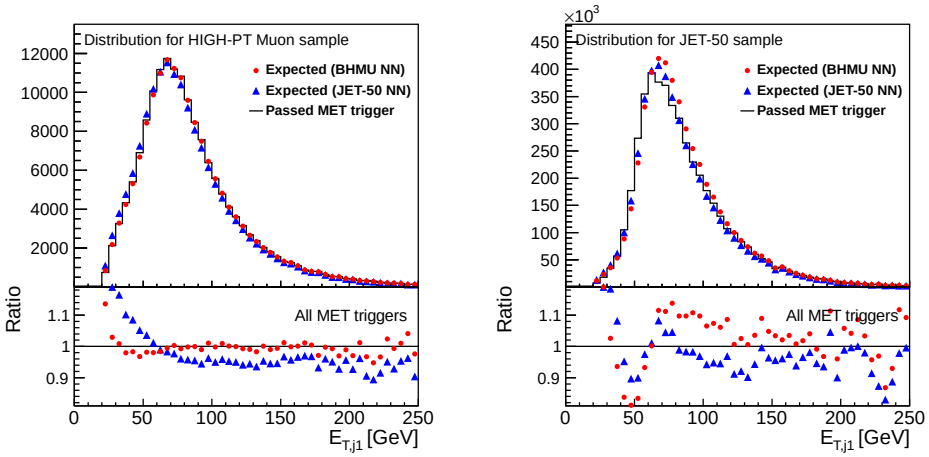


Figure A7.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $E_T^{j_1}$, and (c,d) $E_T^{j_2}$.

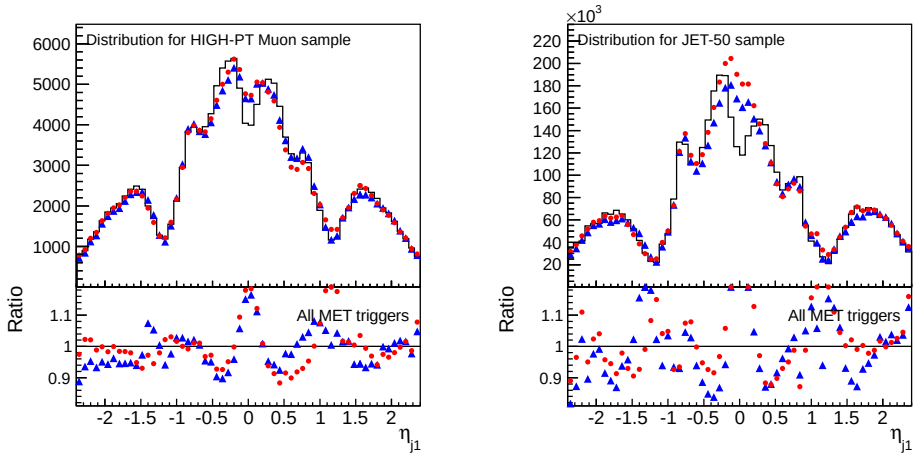


Figure A8.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) η^{j_1}, and (c,d) η^{j_2}.

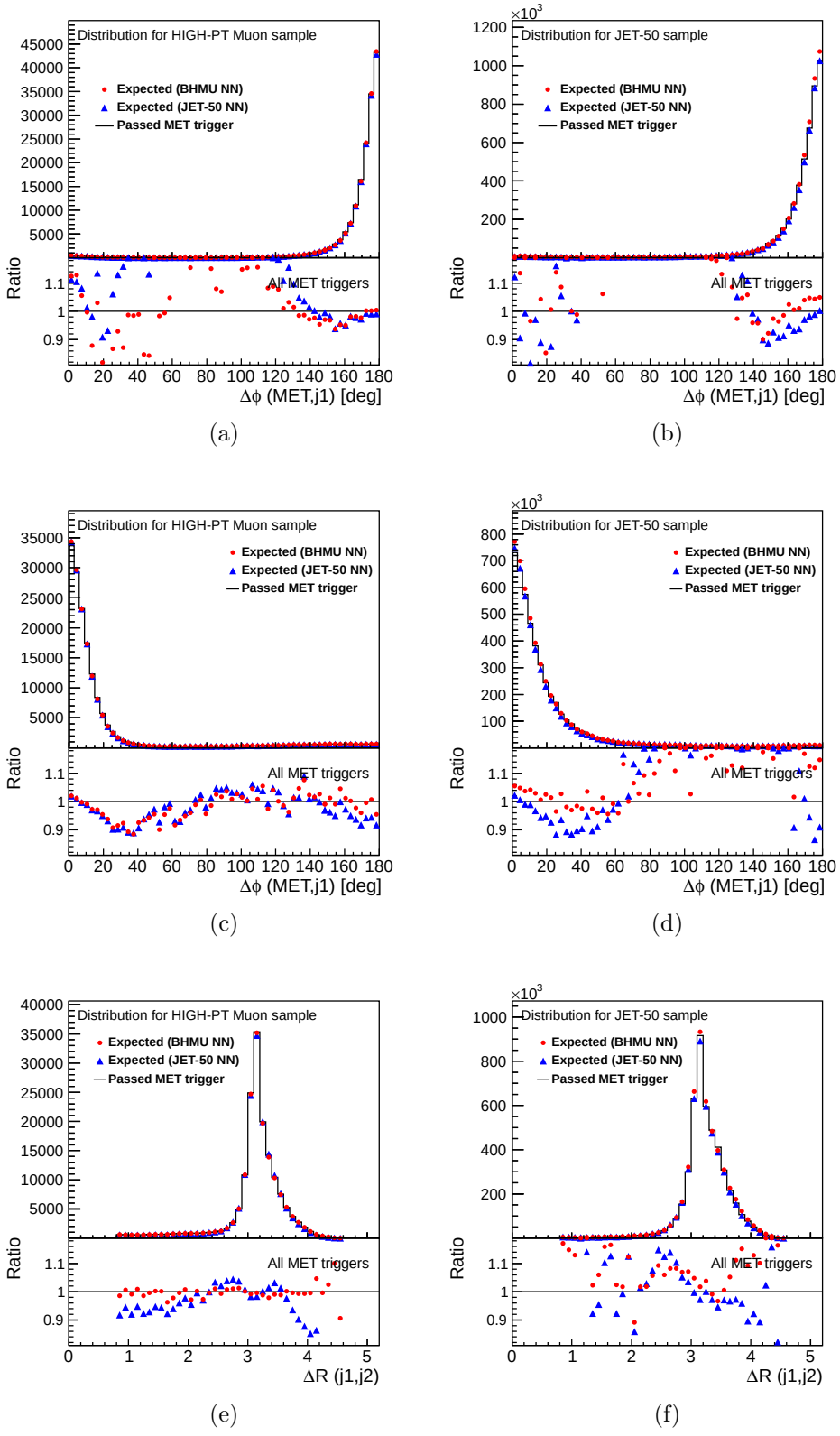


Figure A9.: Trigger efficiency for the MET+CJET+JET path as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) $\Delta\varphi(\not{E}_T, j_1)$; (c,d) $\Delta\varphi(\not{E}_T, j_2)$; and (e.f) $\Delta R(j_1, j_2)$.

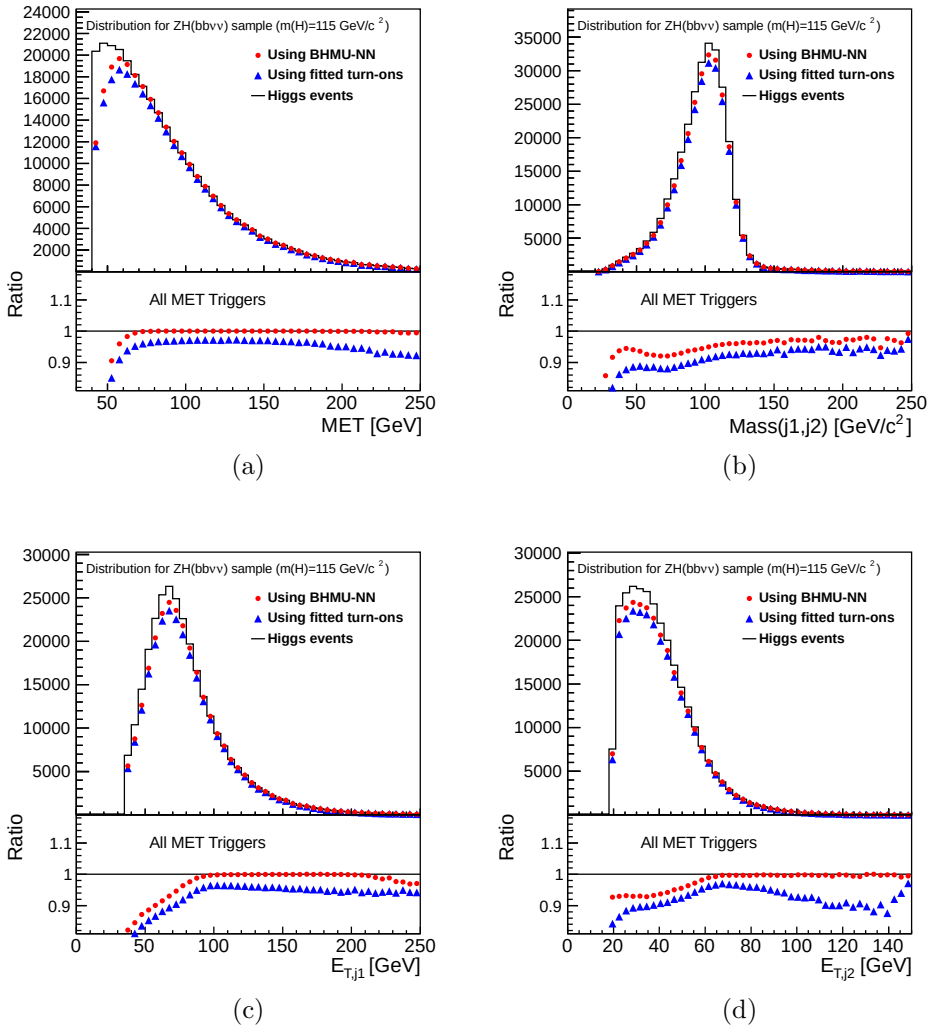


Figure A10.: Distribution of several kinematic variables for ZH events. Comparison between the fit-based parameterization and the central NN one. The distribution without considering any trigger effect is shown as a reference. All three MET trigger paths are considered. Variables: (a) $\not\!\!E_T$; (b) $m(j_1, j_2)$; (c) $E_T^{j_1}$; and (d) $E_T^{j_2}$.

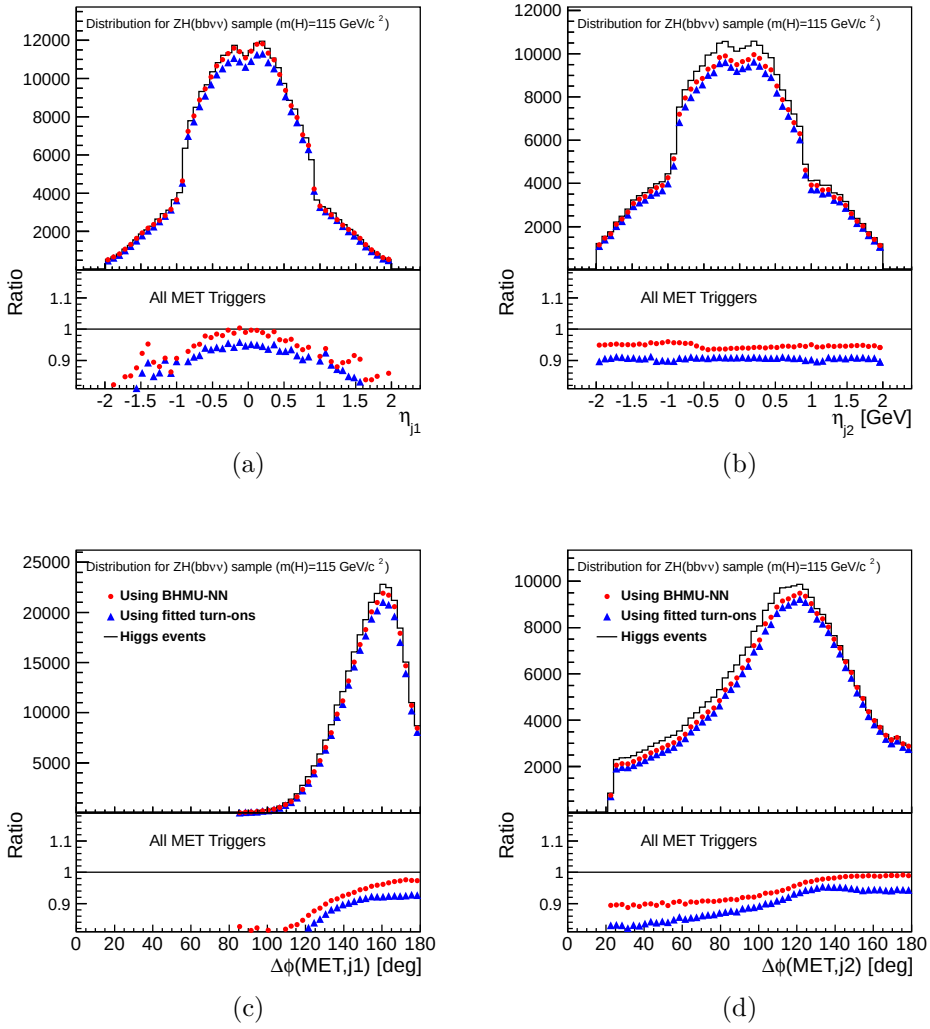


Figure A11.: Distribution of several kinematic variables for ZH events. Comparison between the fit-based parameterization and the central NN one. The distribution without considering any trigger effect is shown as a reference. All three MET trigger paths are considered. Variables: (a) η^{j_1}; (b) η^{j_2}; (c) $E_T^{j_1}$; and (d) $E_T^{j_2}$.

These distributions prove the parameterization is again doing quite well in predicting the kinematics. However, there is room for improvement by increasing statistics, or, equivalently, training a network dedicated to a particular region of interest), and also by increasing the number of nodes in the hidden layers, to try to account for the discrepancies in some regions.

Performance Studies

We use the previous parameterization from section 10.6.5 to compare the improvement due to the addition of the MET45 path. Figures A10 and A11 show the outcome of this comparison, performed with non-relaxed cuts. The improvement is clearly visible, already with this tight cut selection; we measure it to be 4% in total yield. Except for the expected differences due to the addition of a new trigger, we do not see anything unexpected.

We thus proceed with relaxing the cut on $\not\!\!E_T$ to 35 GeV (from 40 before) and that of the E_T of the leading jet to 25 GeV (from 35 before). The results are shown in figures A12 to A13.

We expect some benefit from the latter cut since we use the MET45 trigger, which is agnostic of the jet energies. However, the yields are not increased by much, indicating that the kinematic region opened by relaxing the cuts has small cross-section and efficiency.

Tough, one advantage of relaxing the cuts is the reduction of the uncertainty. The fit-based parameterization didn't handle the correlation between variables, and thus the estimation of the uncertainty might be underestimated, especially since the new kinematic region was not part of the region used to derive the fit-based functions. We observe that the uncertainty of the current parameterization is not very different from the past one, which implies that the NN-driven parameterization didn't increase the systematic uncertainty, which was the main worry about a selection with relaxed cuts.

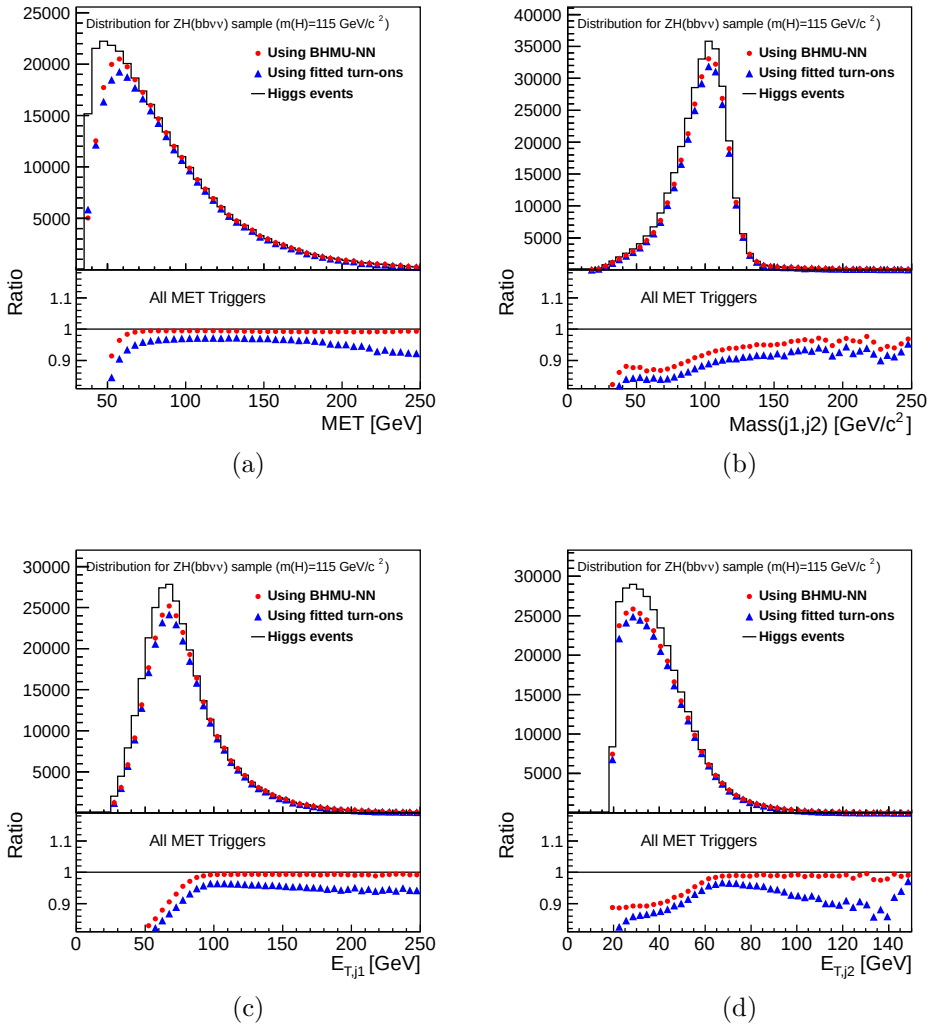


Figure A12.: Distribution of several kinematic variables for ZH events. Comparison between the fit-based parameterization and the central NN one. The distribution without considering any trigger effect is shown as a reference. All three MET trigger paths are considered. Variables: (a) $\not\!\!E_T$; (b) $m(j_1, j_2)$; (c) $E_T^{j_1}$; and (d) $E_T^{j_2}$.

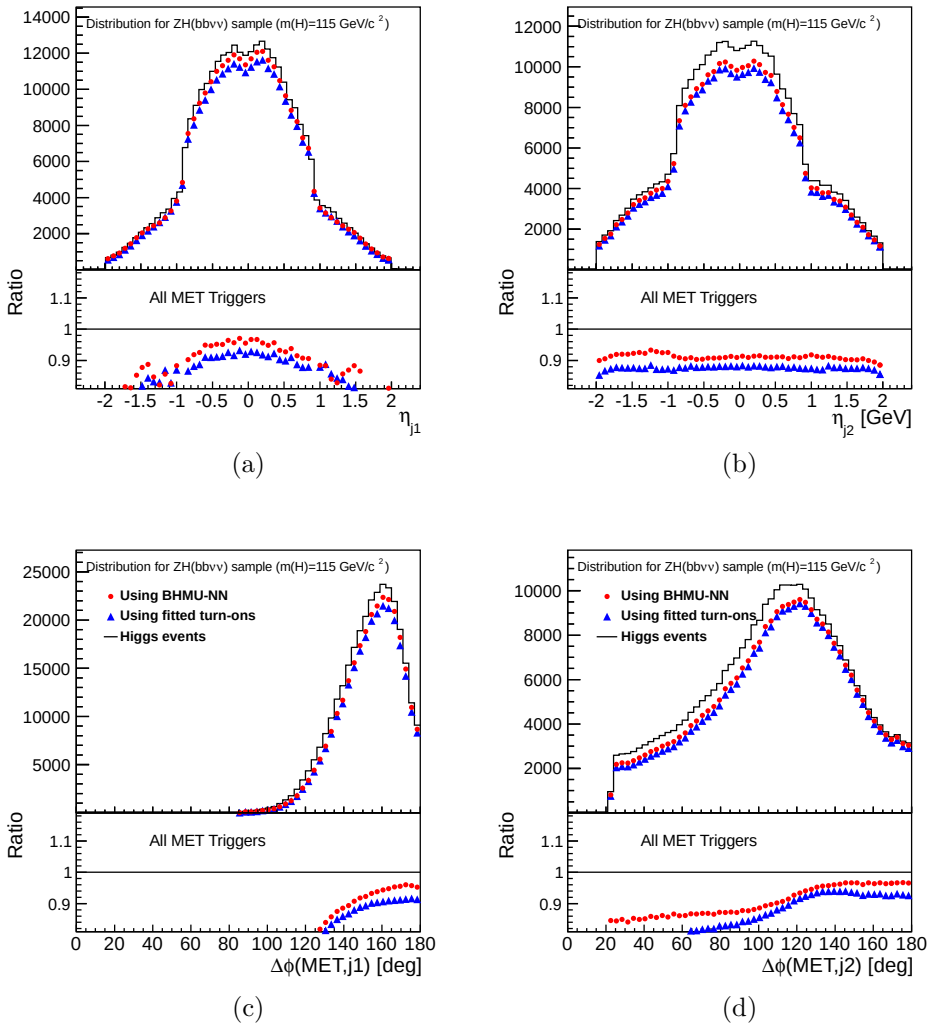


Figure A13.: Distribution of several kinematic variables for ZH events. Comparison between the fit-based parameterization and the central NN one. The distribution without considering any trigger effect is shown as a reference. All three MET trigger paths are considered. Variables: (a) η^{j_1}; (b) η^{j_2}; (c) $E_T^{j_1}$; and (d) $E_T^{j_2}$.

B.2 Final Neural Network Parameterization of the Efficiency

The final parameterization retained for this analysis uses the findings from the previous sections. On top of slight changes to the structure of the network (additional nodes in the hidden layers), we decided to train a network specifically for high-$\not\!\!E_T$ events in order to improve the precision. Additionally, we introduce two 3-jet parameterizations, extending those for 2-jet events.

B.2.1 Extended 2-jet Parameterization

With respect to the selections from the previous section, we applied the following modifications: (a) we relax completely the requirements on the first jet, thus considering jets with $E_T \geq 20$ GeV and $|\eta| < 2.4$; (b) we remove the requirement of a central jet, and (c) we remove the requirement on the ΔR between the leading two jets. We use these requirements even if they are not a perfect match to those in our analysis to derive the most general parameterization.

We train a network with a 9:15:14:1 configuration. The results of this training are presented in figures 10.22 to 10.25. We observe reasonable description of the average efficiencies as a function of the input variables. Some regions are not perfect due the the limited precision, which affects some parts of phase space that are particularly hard to model. This is alleviated by using a reasonable selection. It should be noted that these regions will be assigned large systematic uncertainties due to the intrinsic differences between the MUON and JET-50 samples, for they have a lower efficiency.

In order to validate the network, we compute the pulls in a discretized space using 19,440 bins[2], as shown in figure A15. Due to the correlations between the variables, the parameterization is less efficient than in the simple cases. Tough, we can see that the training process does not pose a big problem.

[2]This represents 5 bins in $\not\!\!E_T$, 4 for the E_T of the jets, 3 for $|\eta|$ of the jets, as well as for the $\Delta\varphi$ and ΔR variables. We use one bin in $\phi(\not\!\!E_T)$, for we are not sensitive to this variable.

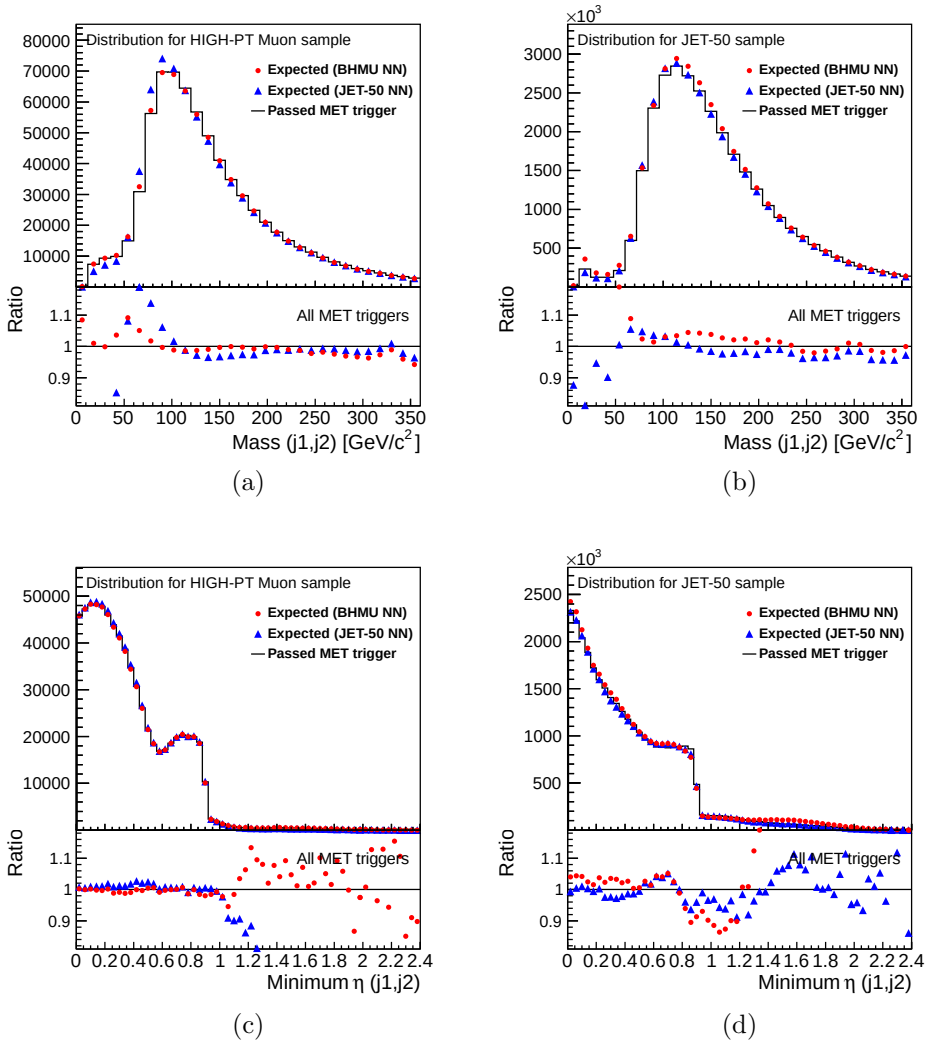
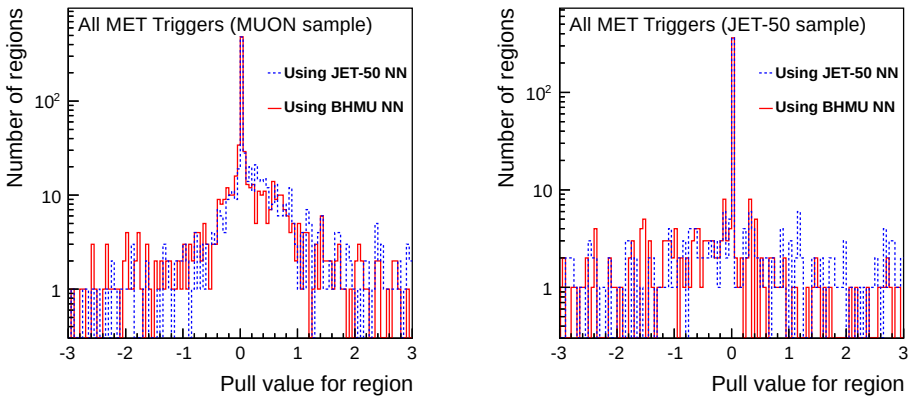
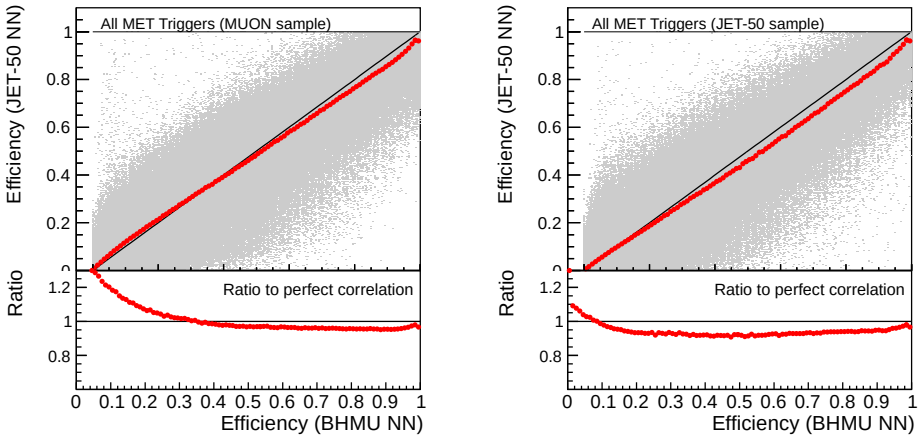


Figure A14.: Trigger efficiency for all the MET trigger paths as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $m(j_1, j_2)$, and (c,d) $min(\eta^{j_1}, \eta^{j_2})$.



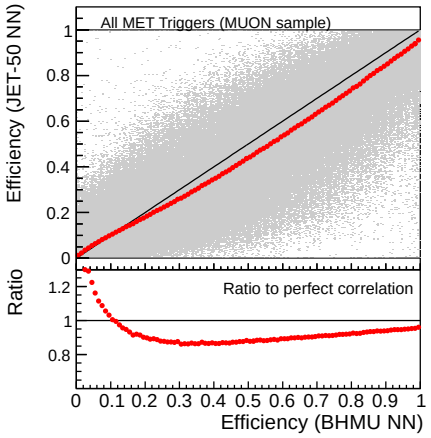
(a) (b)

Figure A15.: Pull (difference relative to the uncertainty) of the observed efficiency of the MET45 path and the predictions from the two final parameterizations in the bins filling the phase space of the two variables used to train the networks ($\not\!\!E_T$ and E_T^{j1}); the pull is plotted in the (a) MUON and (b) JET-50 samples.

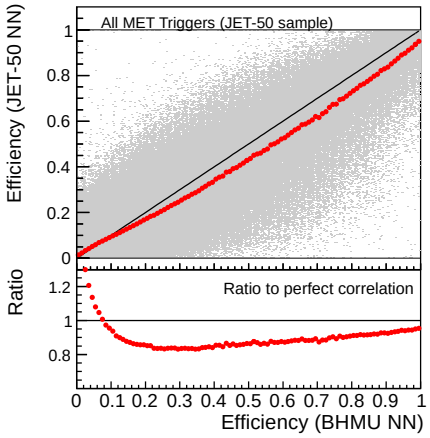


(a) 2-jet events

(b) 2-jet events



(c) 3-jet events



(d) 3-jet events

Figure A16.: Correlation and average distributions of the NN output as trained in the JET-50 sample with respect to that from the NN trained in the MUON sample. Events are taken from (a,c) the MUON and (b,d) the JET-50 sample. The ratio plot gives an idea of the average uncertainty as a function of the central parameterization of the efficiency.

Figure A16 show the distribution of the averaged efficiency expected from the network as a function of the actual efficiency. The general trend is satisfactory, although we cannot conclude much more because of fluctuations.

B.2.2 Parameterization of 3-jet Events

Because the analyses for which this parameterization is intended includes events with three jets in the final state, we perform a specific parameterization of these events. For this purpose, we add five variables to the network: (a) the E_T of the third jet; (b) the pseudo-rapidity of the third jet; (c) the angular separation between the $\not\!\!E_T$ and the third jet in the transverse plane; (d) the separation in $\eta - \phi$ between the leading jet and the third jet $\Delta R(j_1, j_3)$, and (e) $\Delta R(j_2, j_3)$. We train a 14-input network with the following configuration $14 : 20 : 19 : 1$.

The average efficiencies as a function of the input variables are shown in figures A27 to A31. We observed good agreement, as shown in figure A16.

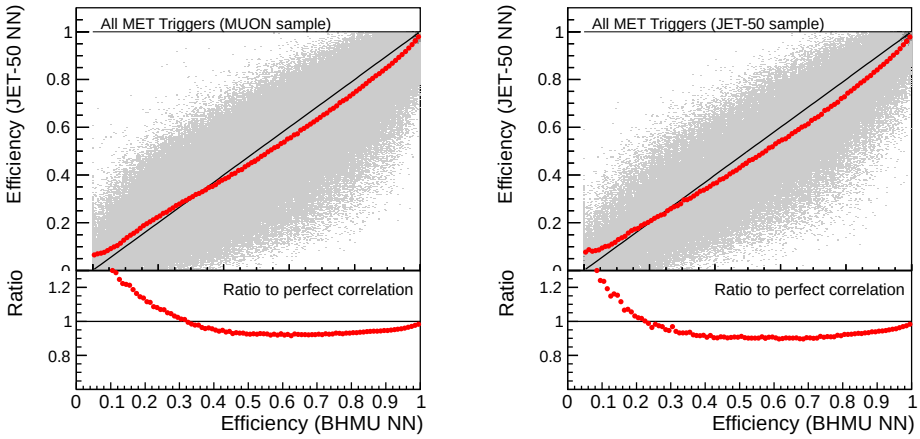
As was the case for the 2-jet selection, the training is not optimal for the high-$\not\!\!E_T$ tail, which is of more interest. This is especially true for the JET-50 sample, for which the efficiency does not saturate at high $\not\!\!E_T$.The problem here is clearly that the sample is depleted of events, since the $\not\!\!E_T$ distribution falls very rapidly.

B.2.3 Parameterization of High-$\not\!\!E_T$ Events

Because of the complexity of the problem, it takes a very long time to train the networks, especially since we want to obtain a good performance. In order to keep the training time reasonable, it is necessary to reduce the amount of events used in the training, reducing the precision in the tails of the distributions, including the high-$\not\!\!E_T$ one, which is the most interesting kinematic region.

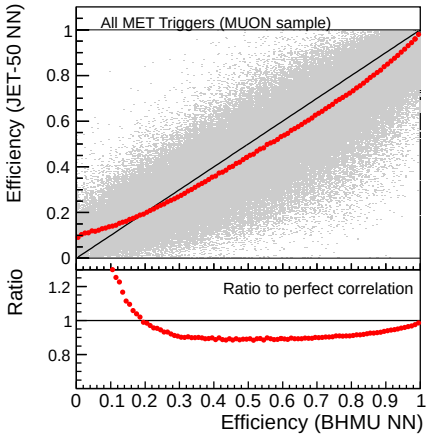
Instead of pre-selecting the events by their $\not\!\!E_T$, we just reduce the sample to the kinematic region we are interested in. It is better to avoid using regions where the efficiency is low, since this introduces larger uncertainties. Therefore, we train networks for events having $\not\!\!E_T > 35$ GeV.

The outcome of these networks are presented in figures A18 to A21, and A27 to A31. They reproduce all the variables reasonably, except in some limited regions

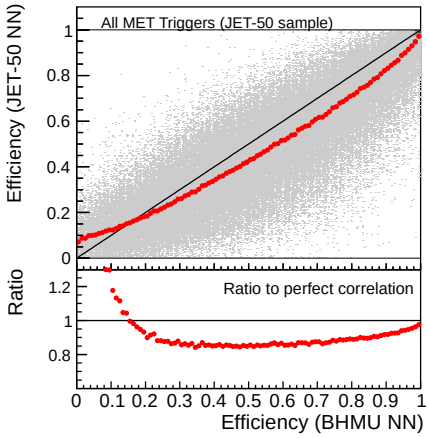


(a) 2-jet events

(b) 2-jet events



(c) 3-jet events



(d) 3-jet events

Figure A17.: Correlation and average distributions of the NN output as trained in the JET-50 sample with respect to that from the NN trained in the MUON sample. Events are taken from (a,c) the MUON and (b,d) the JET-50 sample. The ratio plot gives an idea of the average uncertainty as a function of the central parameterization of the efficiency.

(which are not the most interesting). Compared to the previous parameterizations, there is a clear improvement in the regions of high efficiency, e.g, high-$\not\!\!E_T$.

We show in figure A17 the correlations plots for the 2- and 3-jet networks, and notice that these networks are not that different in practice.

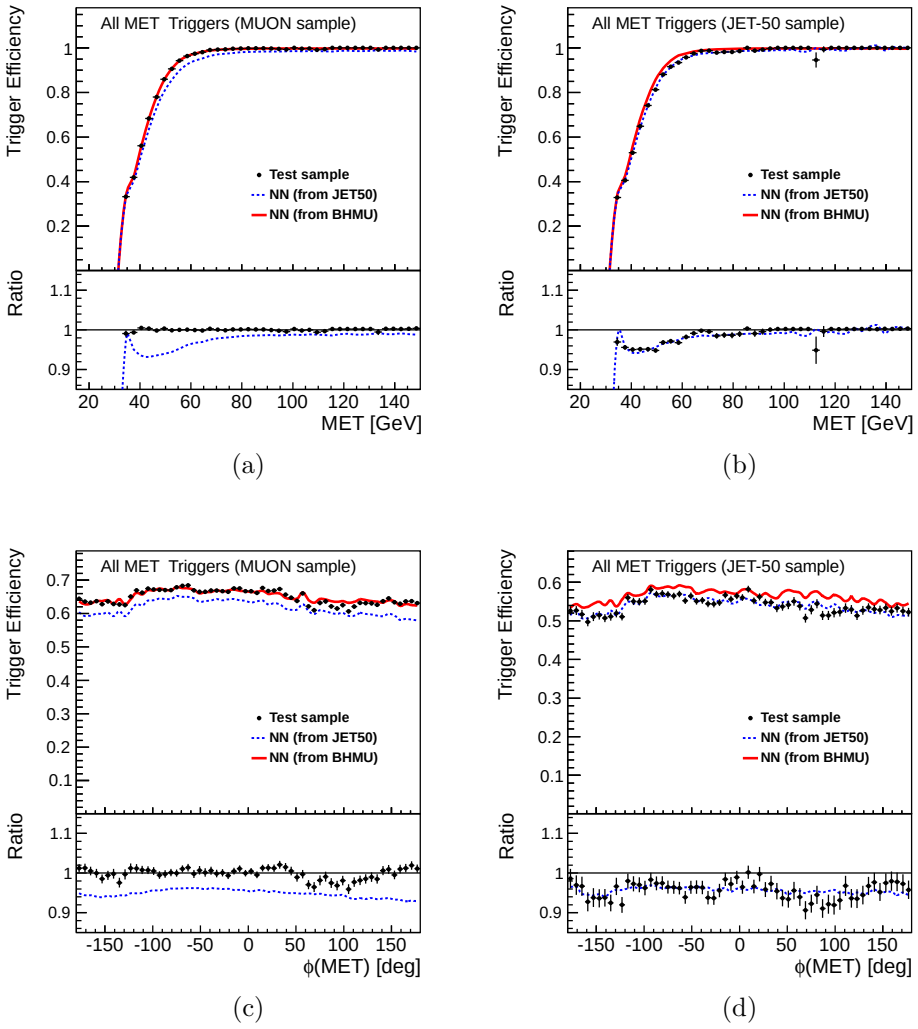


Figure A18.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $\not\!\!E_T$, and (c,d) $\phi(\not\!\!E_T)$.

B.3 Validation of the inputs to the final trigger turn-on paremeterization

B.3.1 Inputs 2-jet neural network

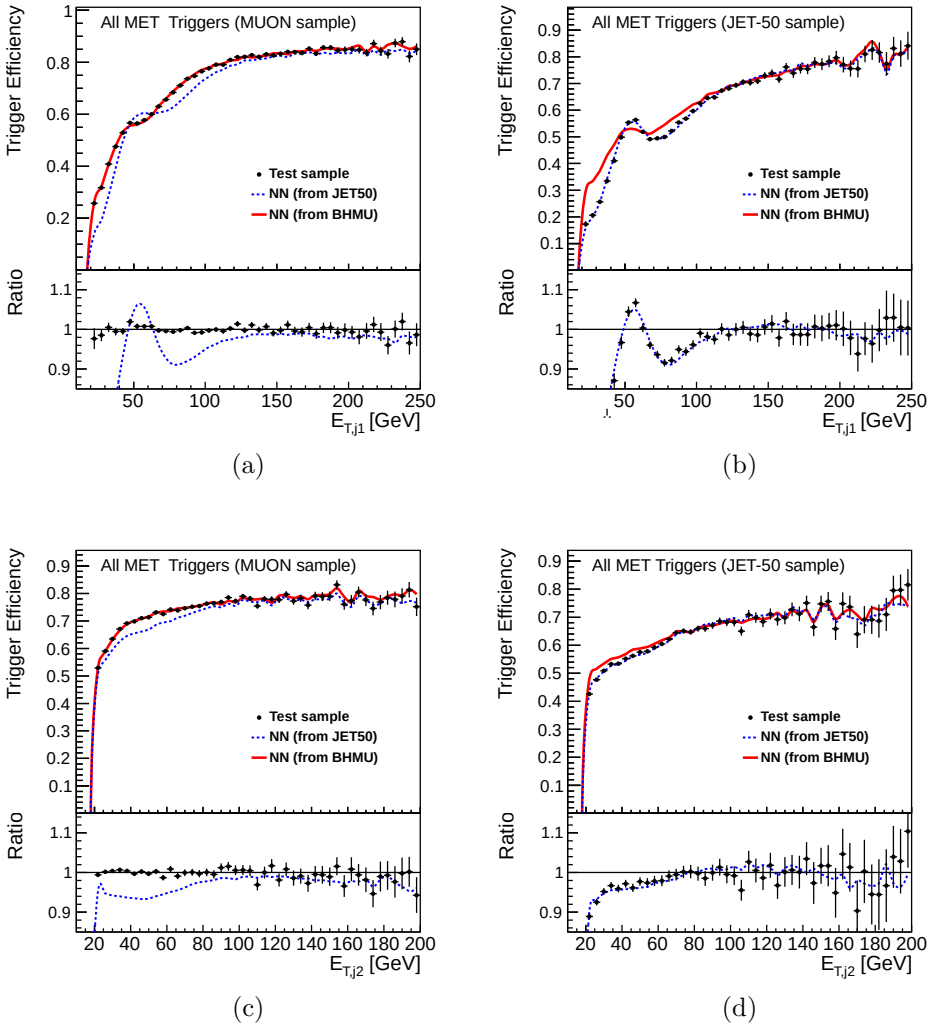


Figure A19.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) E_T^{j1}, and (c,d) E_T^{j2}.

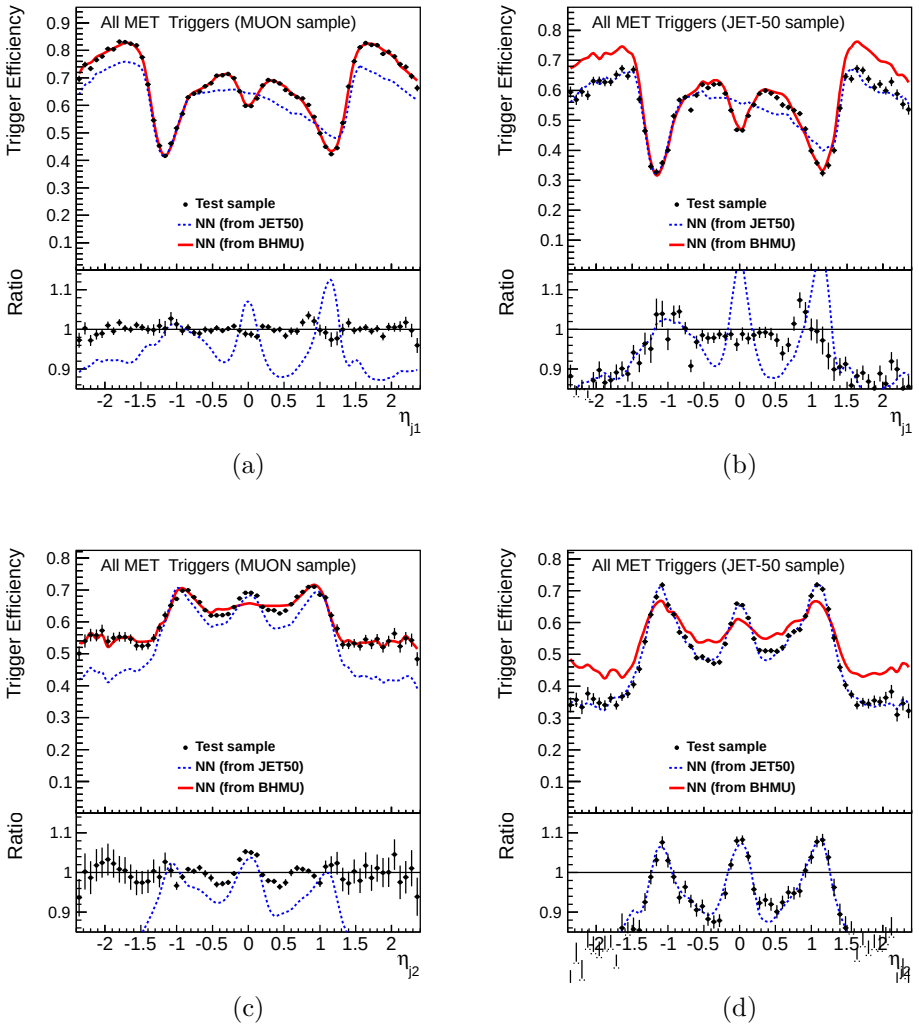


Figure A20.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) η^{j_1}, and (c,d) η^{j_2}.

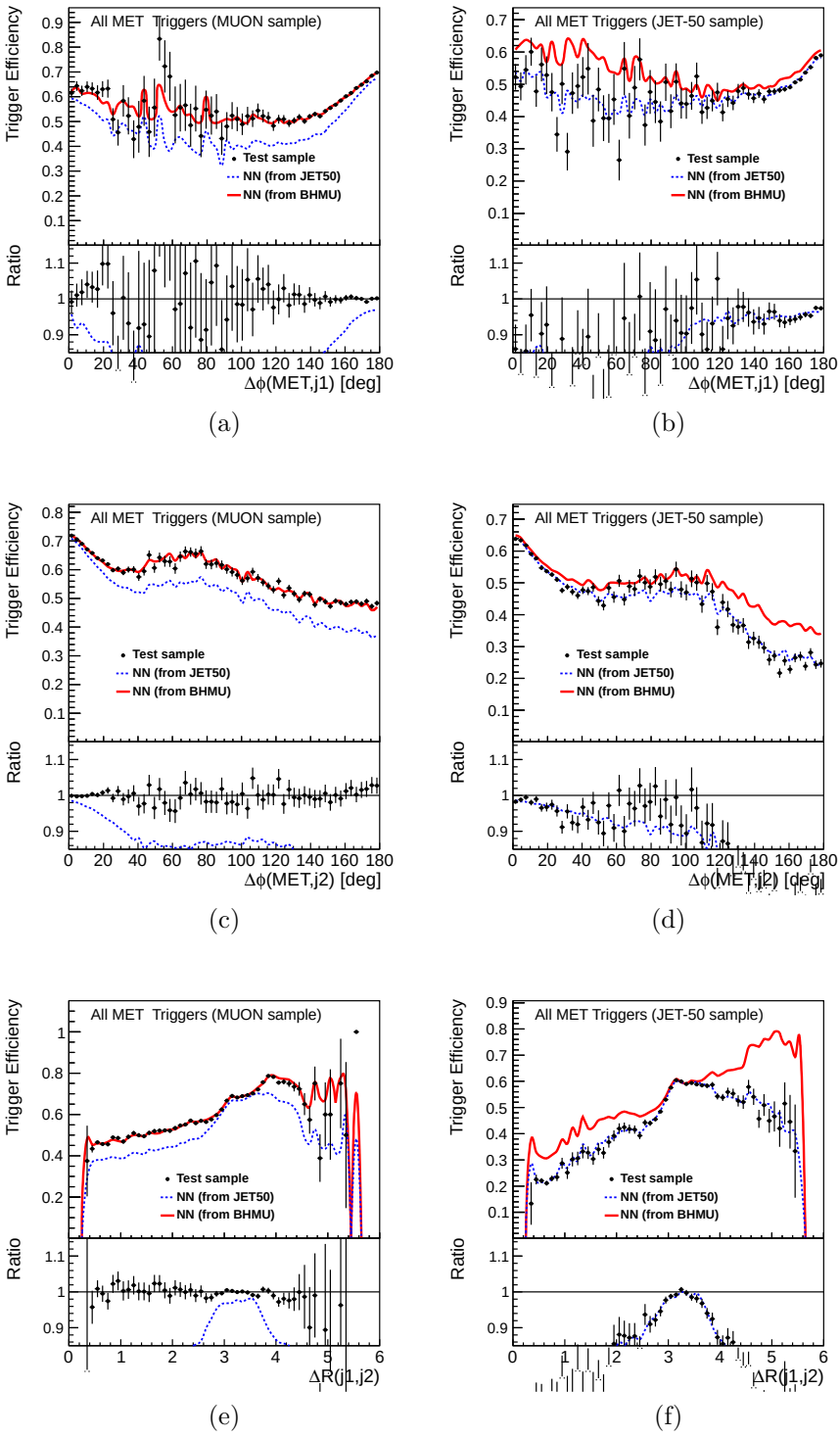


Figure A21.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) $\Delta\varphi(\not\!\!E_T, j_1)$; (c,d) $\Delta\varphi(\not\!\!E_T, j_2)$; and (e.f) $\Delta R(j_1, j_2)$.

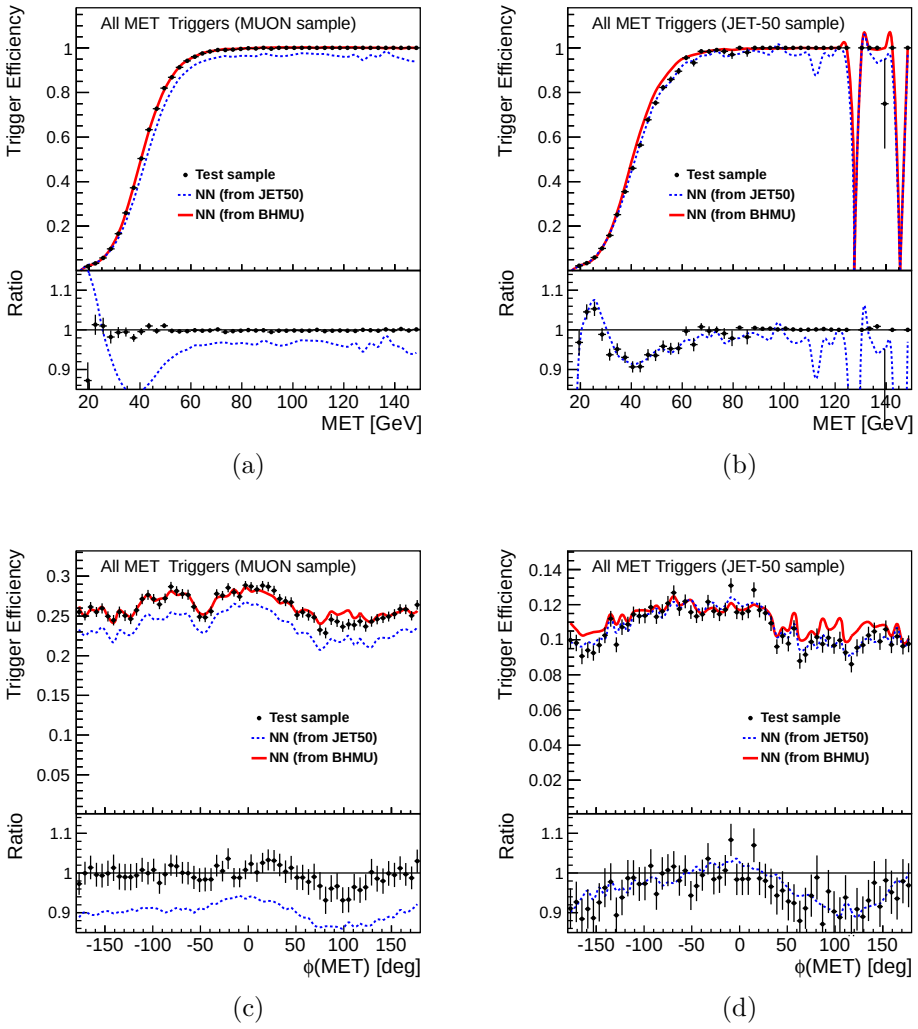


Figure A22.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $\not\!\!E_T$, and (c,d) $\phi(\not\!\!E_T)$.

B.3.2 Inputs 3-jet neural network

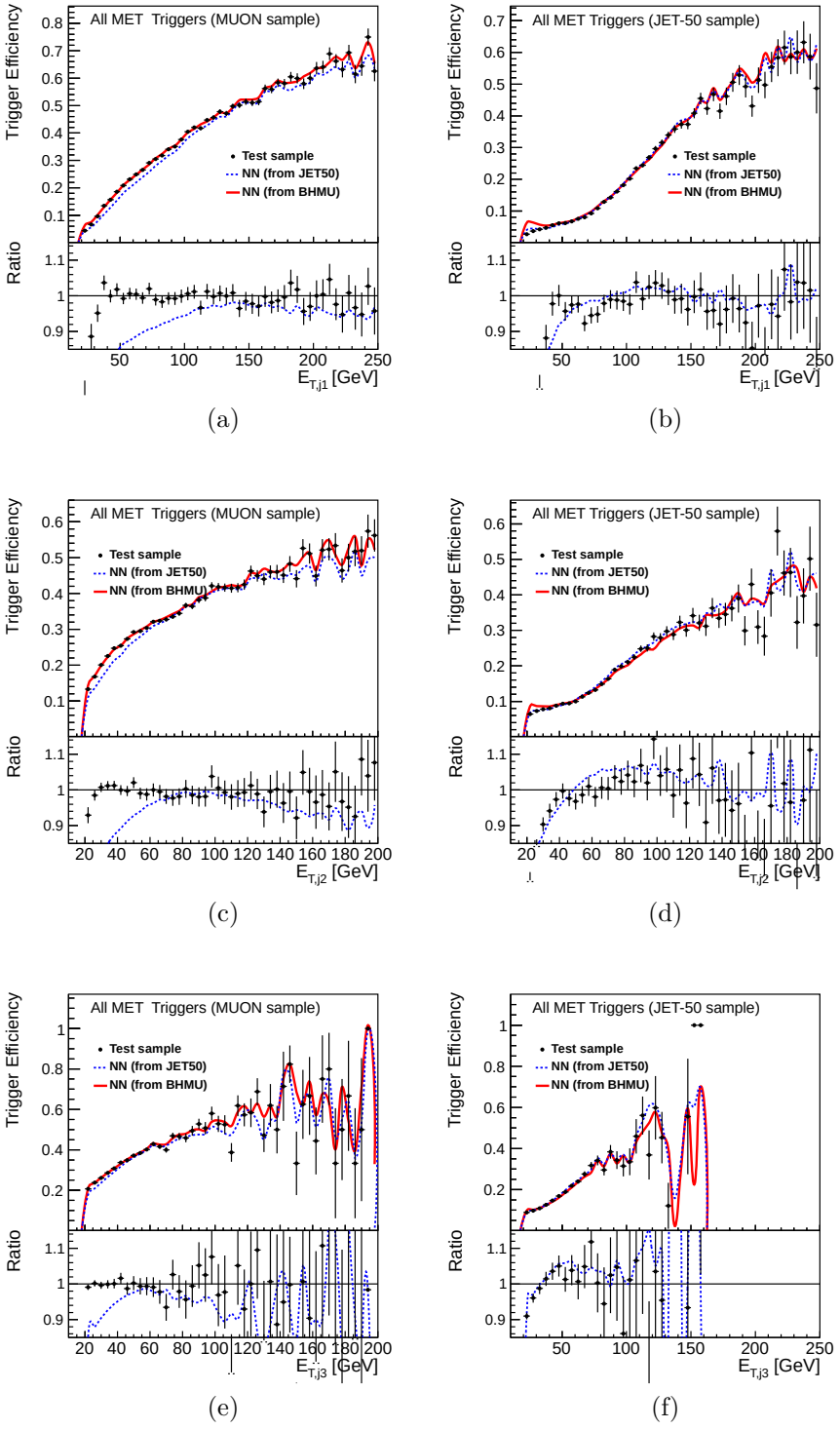


Figure A23.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) E_T^{j1}; (c,d) E_T^{j2}; and (e,f) E_T^{j3}.

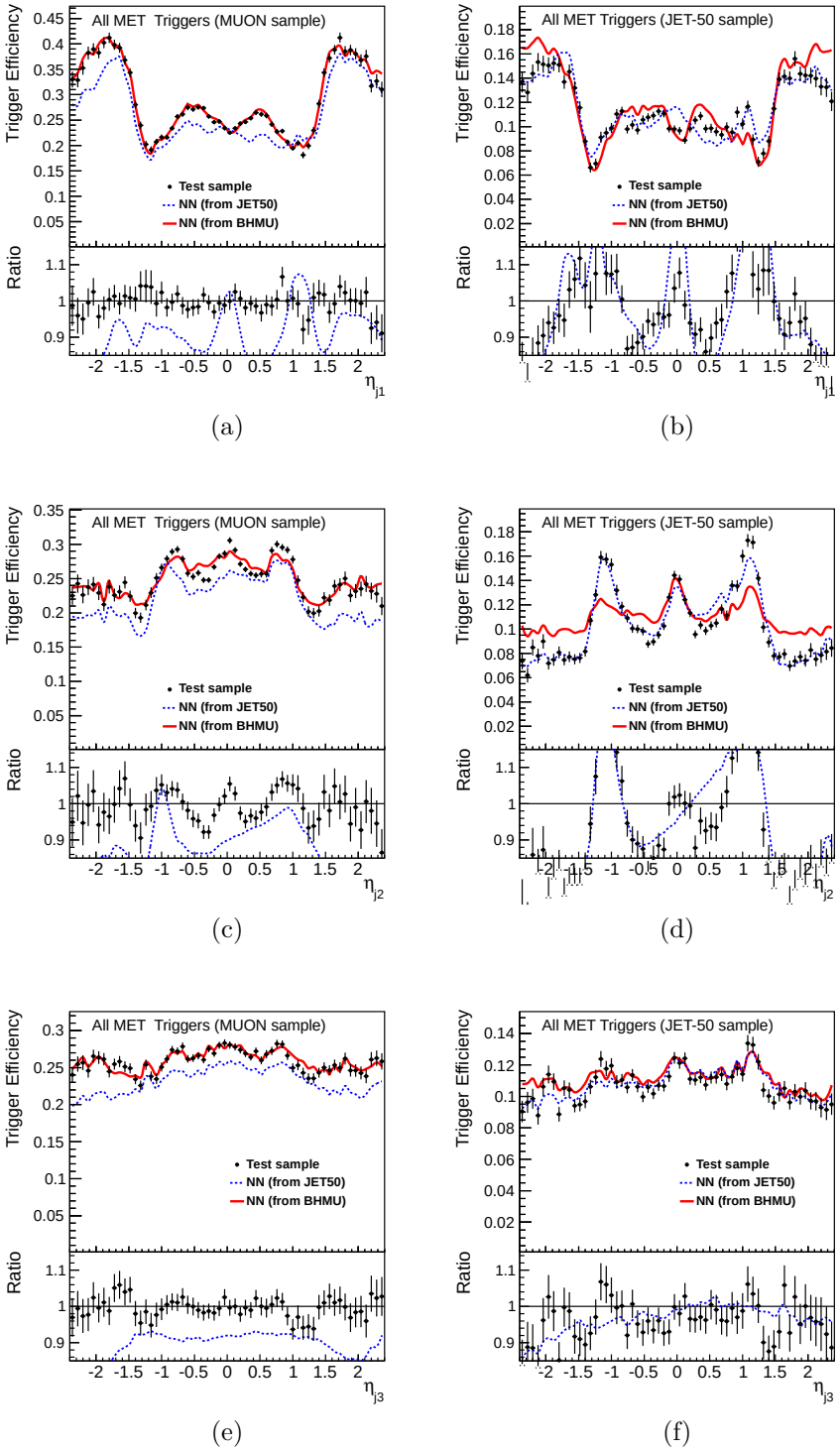


Figure A24.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) η^{j_1}; (c,d) η^{j_2}; and (e,f) η^{j_3}.

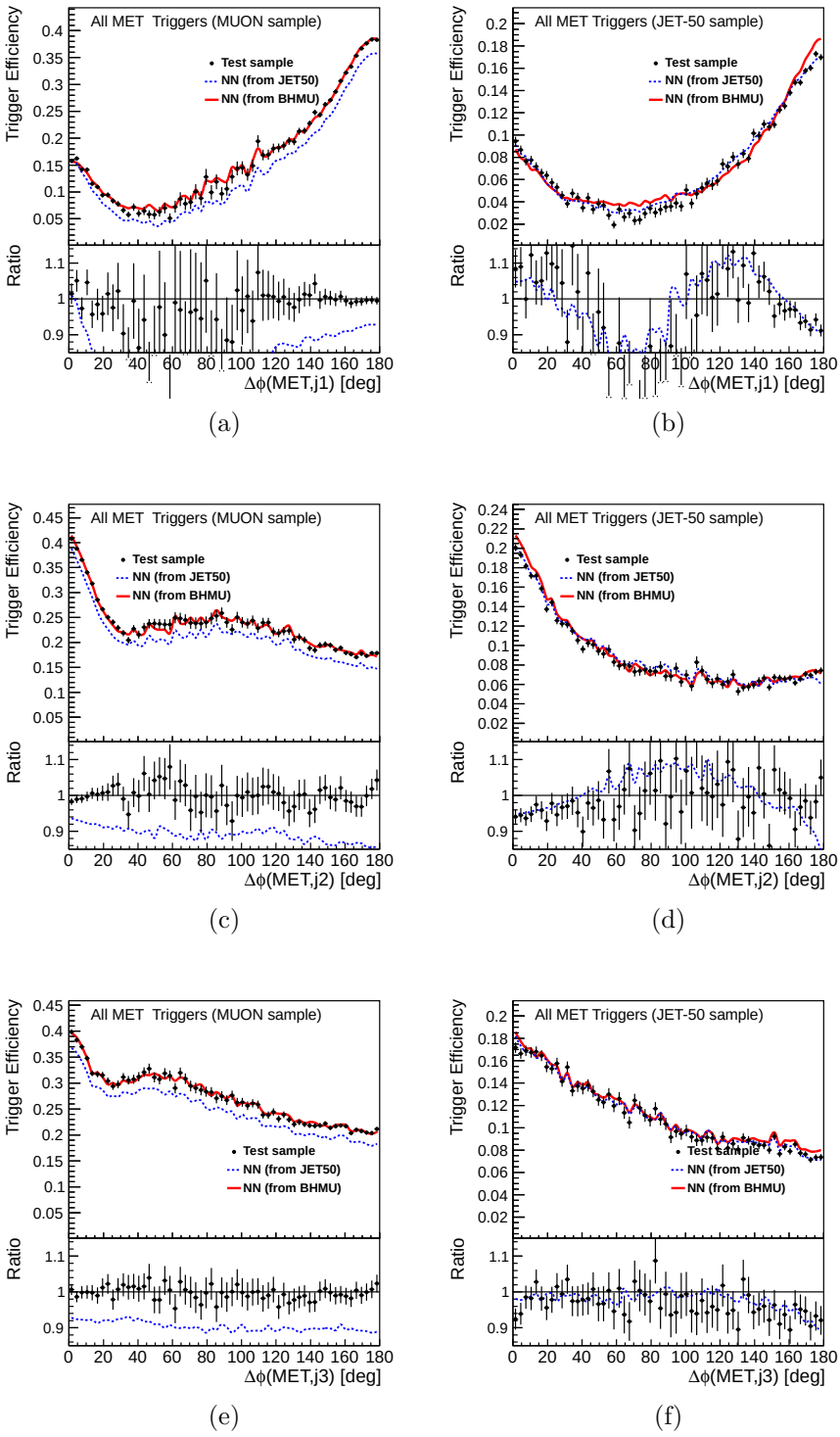


Figure A25.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) $\Delta\varphi(\not\!\!E_T, j_1)$; (c,d) $\Delta\varphi(\not\!\!E_T, j_2)$; and (e,f) $\Delta\varphi(\not\!\!E_T, j_3)$.

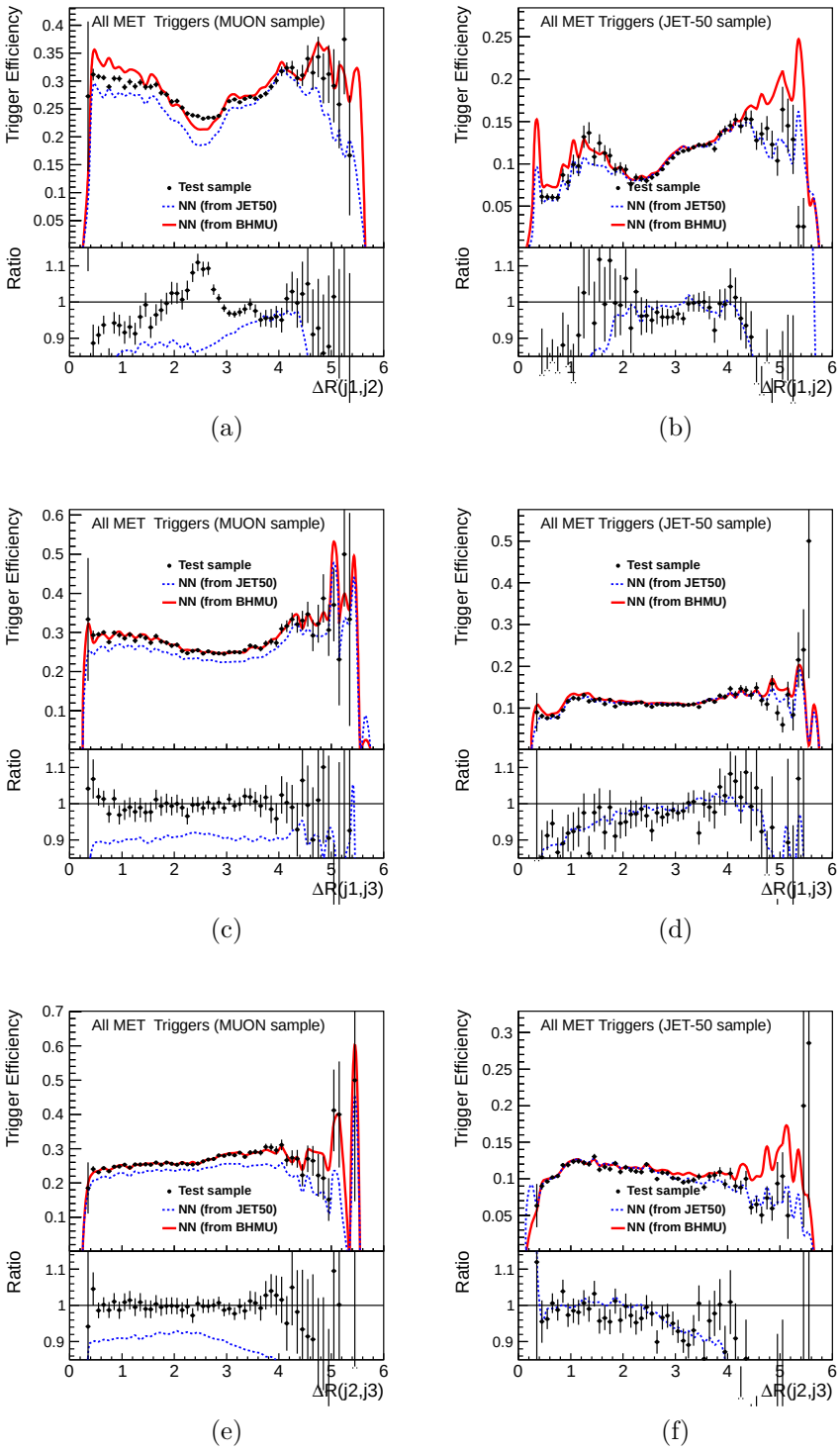


Figure A26.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) $\Delta R(j_1, j_2)$; (c,d) $\Delta R(j_1, j_3)$; and (e,f) $\Delta R(j_2, j_3)$.

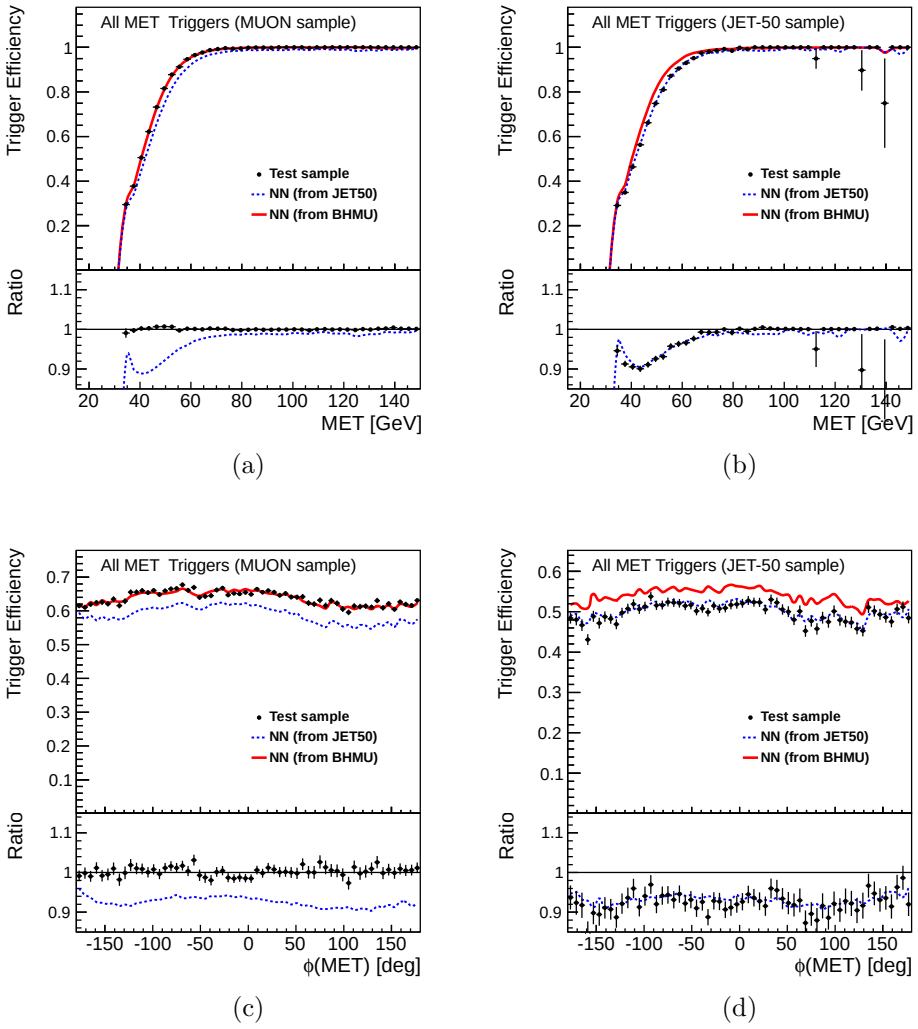


Figure A27.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c) MUON and (b,d) JET-50 samples. Distributions: (a,b) $\not\!\!E_T$, and (c,d) $\phi(\not\!\!E_T)$.

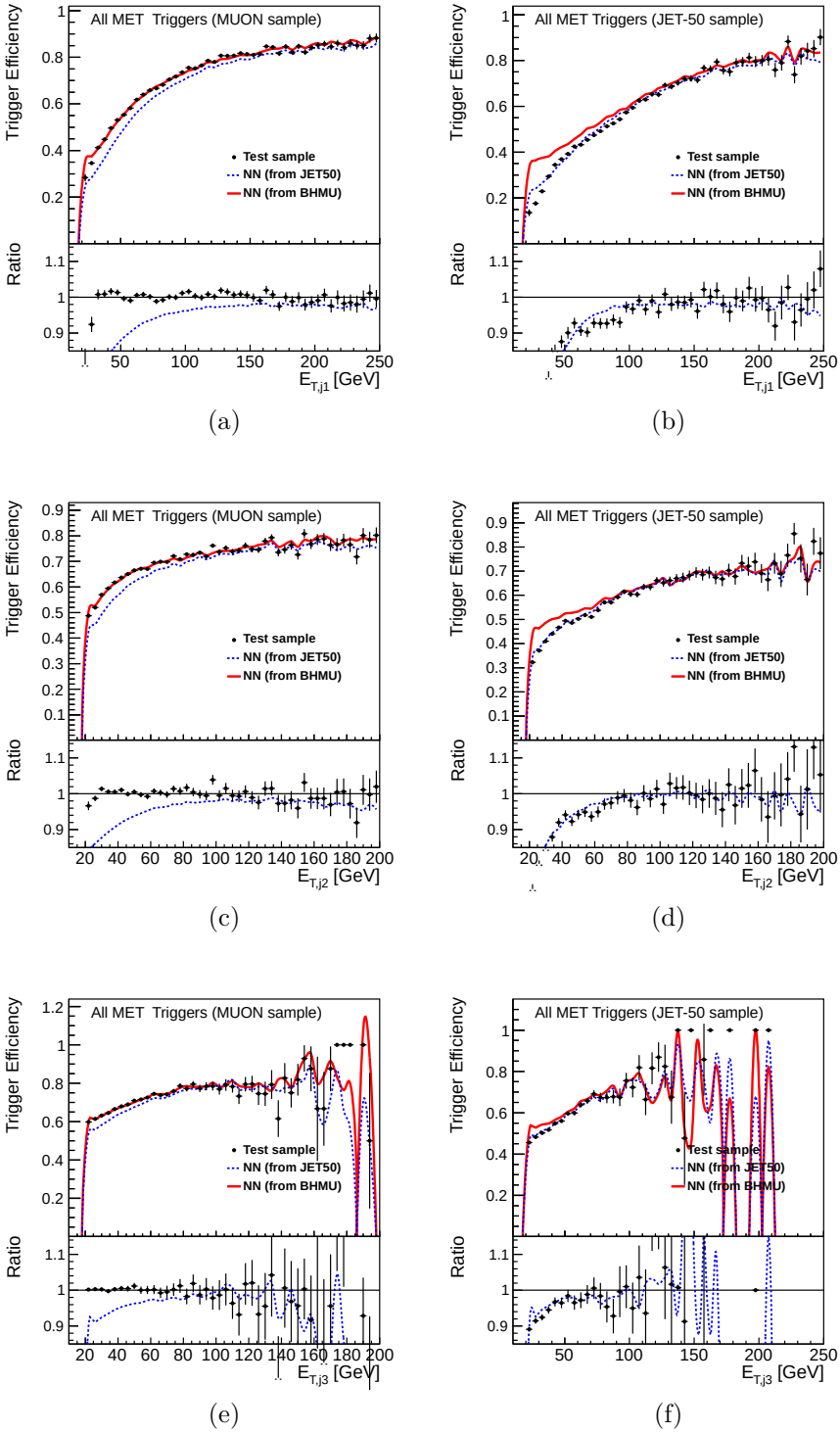


Figure A28.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) E_T^{j1}; (c,d) E_T^{j2}; and (e,f) E_T^{j3}.

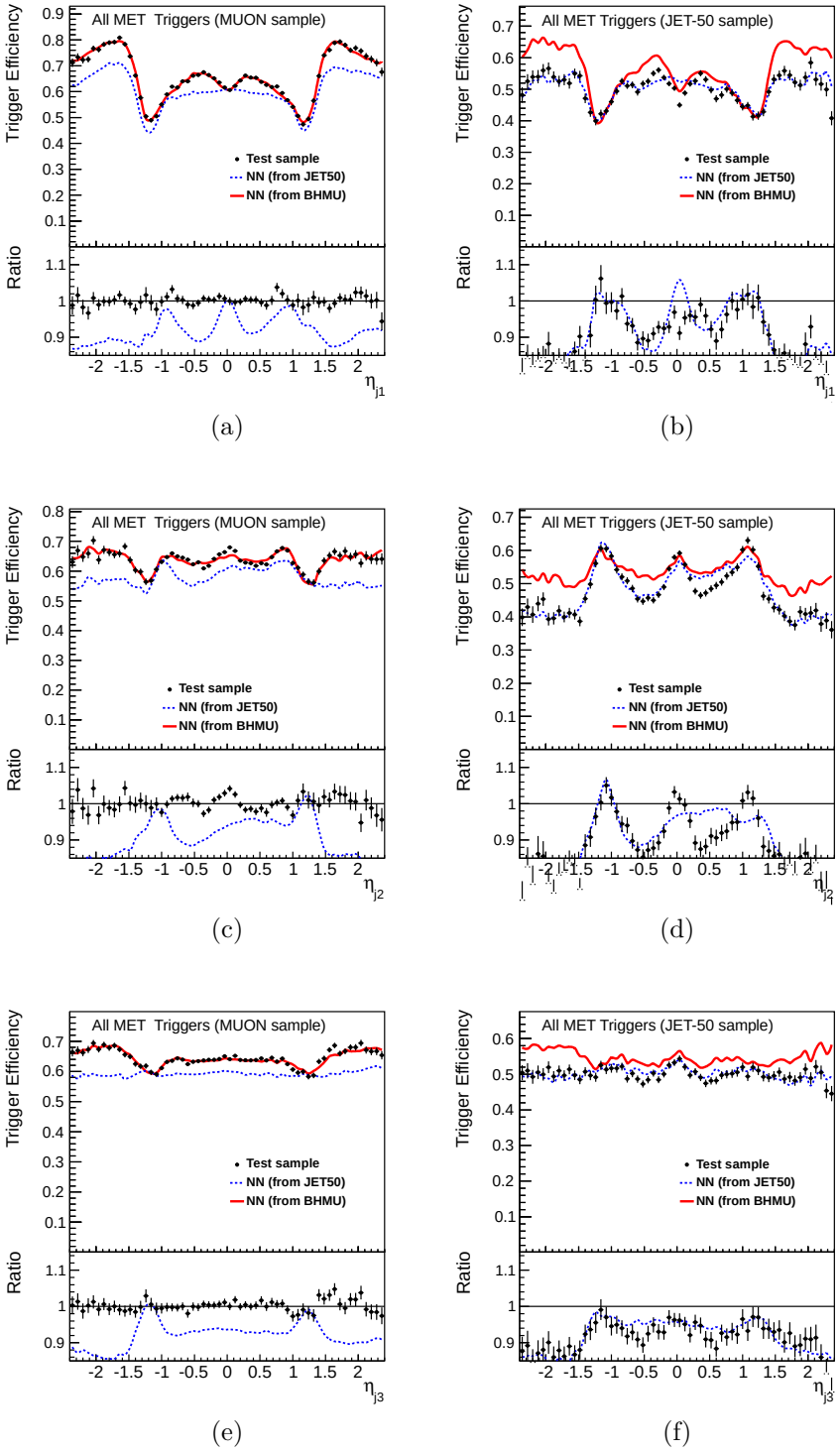


Figure A29.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) η^{j_1}; (c,d) η^{j_2}; and (e,f) η^{j_3}.

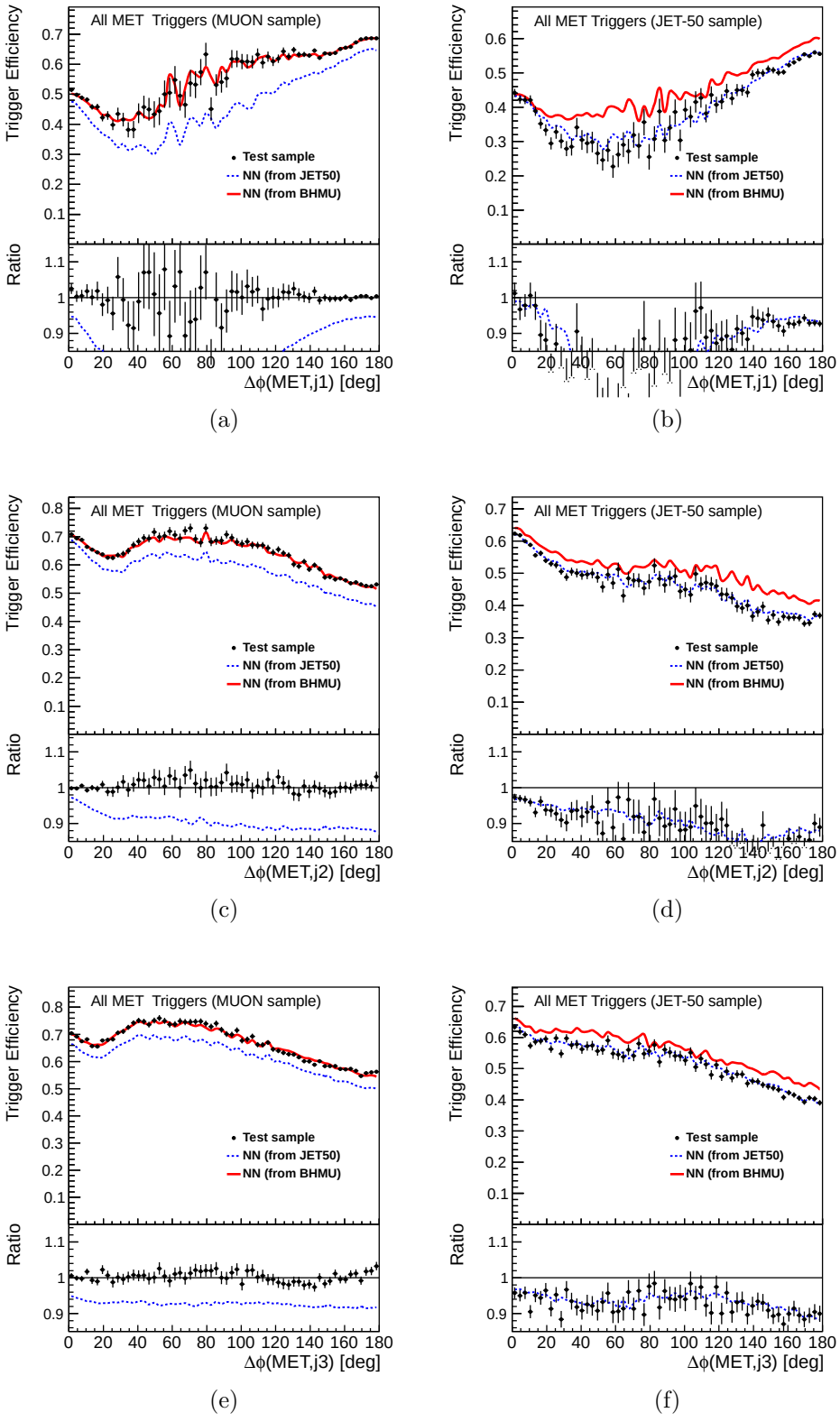


Figure A30.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) $\Delta\varphi(\not\!\!E_T, j_1)$; (c,d) $\Delta\varphi(\not\!\!E_T, j_2)$; and (e,f) $\Delta\varphi(\not\!\!E_T, j_3)$.

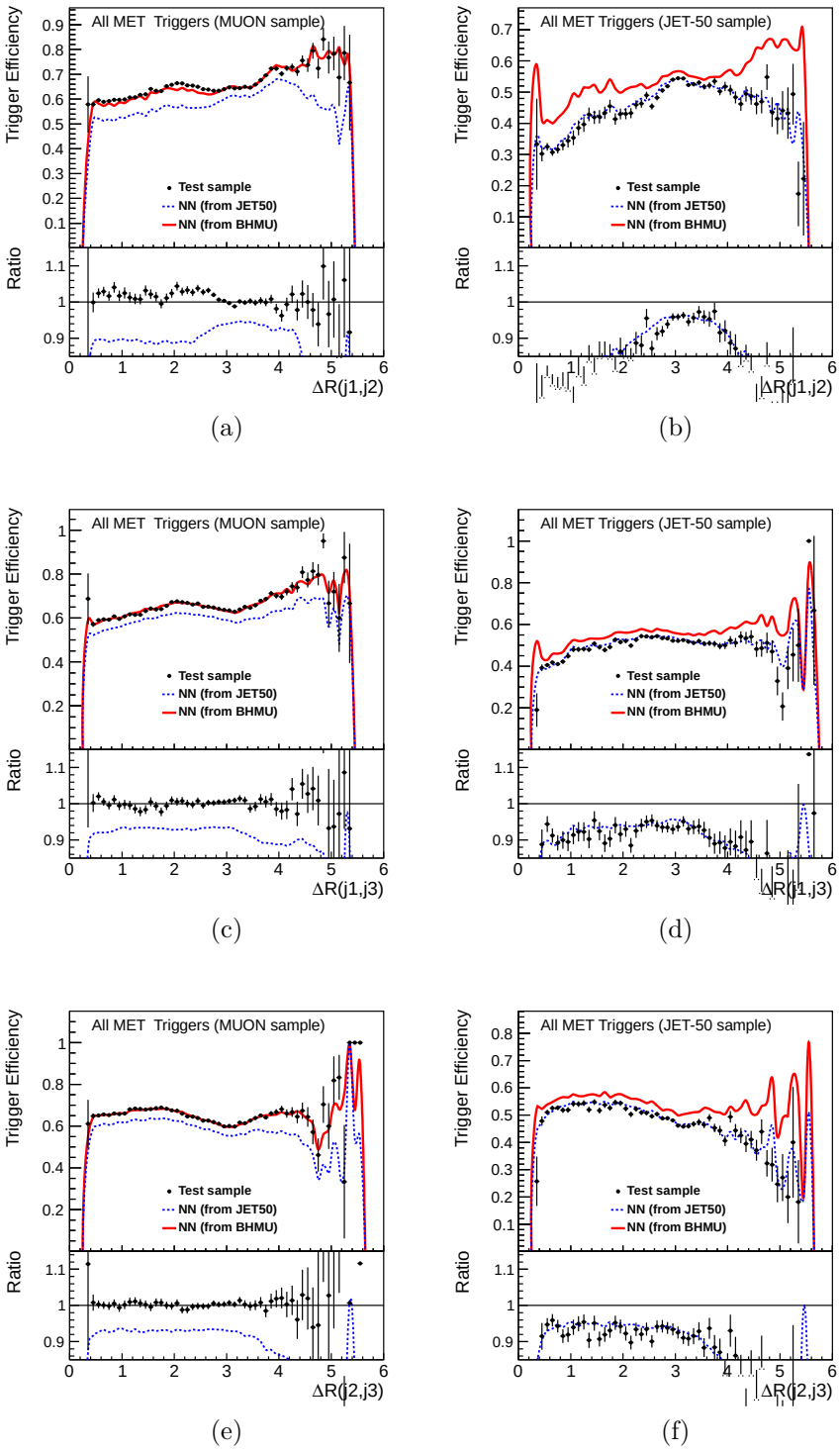


Figure A31.: Trigger efficiency for all the MET paths as a function of several variables as measured and parameterized by a NN function in the (a,c,e) MUON and (b,d,f) JET-50 samples. Distributions: (a,b) $\Delta R(j_1, j_2)$; (c,d) $\Delta R(j_1, j_3)$; and (e,f) $\Delta R(j_2, j_3)$.

C. ADDITIONAL CONTROL REGION PLOTS

In this appendix, we present additional control region plots for the inputs to the NN_{QCD} (cf. section 11.2.6) and NN_{QCD} (cf. section 14.2). We did not reproduce the plots already in the text (figures 11.8, 11.9, 11.10, 11.11, 14.2, and 14.3).

For convenience, we provide again the definitions of the control regions. These consist of events with two or three jets having $\not\!\!E_T > 35$ GeV, $E_T^{j_1} > 25$ GeV, $E_T^{j_2} > 20$ GeV, $E_T^{j_3} > 15$ GeV (when applicable), $|\eta^{j_{(1,2)}}| < 2.0$ (and at least one with $|\eta| < 0.9$), and $\eta^{j_3} < 2.4$. Additionally, we require the following for each region:

TRM $35 < \not\!\!E_T < 70$ GeV, $\Delta\varphi(j_2, \not\!\!E_T) \leq 0.4$, and no identified lepton;

QCD $\not\!\!E_T > 70$ GeV, $\Delta\varphi(j_2, \not\!\!E_T) \leq 0.4$, and no identified lepton;

EWK $\not\!\!E_T > 35$ GeV, $\Delta\varphi(j_1, \not\!\!E_T) > 1.5$, $\Delta\varphi(j_{(2,3)}, \not\!\!E_T) > 0.4$, and at least one identified lepton.

Pre-selection $\not\!\!E_T > 35$ GeV, $\Delta\varphi(j_1, \not\!\!E_T) > 1.5$, $\Delta\varphi(j_{(2,3)}, \not\!\!E_T) > 0.4$, and no identified lepton.

The signal region consists of pre-selection events satisfying $NN_{QCD} > 0.45$. We remind the reader that in the following plots, the Multijet component refers to the sum of QCD HF, QCD LF mis-tags, and EWK LF mis-tags. The former two are derived from data (cf. section 8.7.1) while the latter is derived from mis-tag matrices (cf. sections 6.1.2 and 6.2.2).

C.1 TRM Control Region

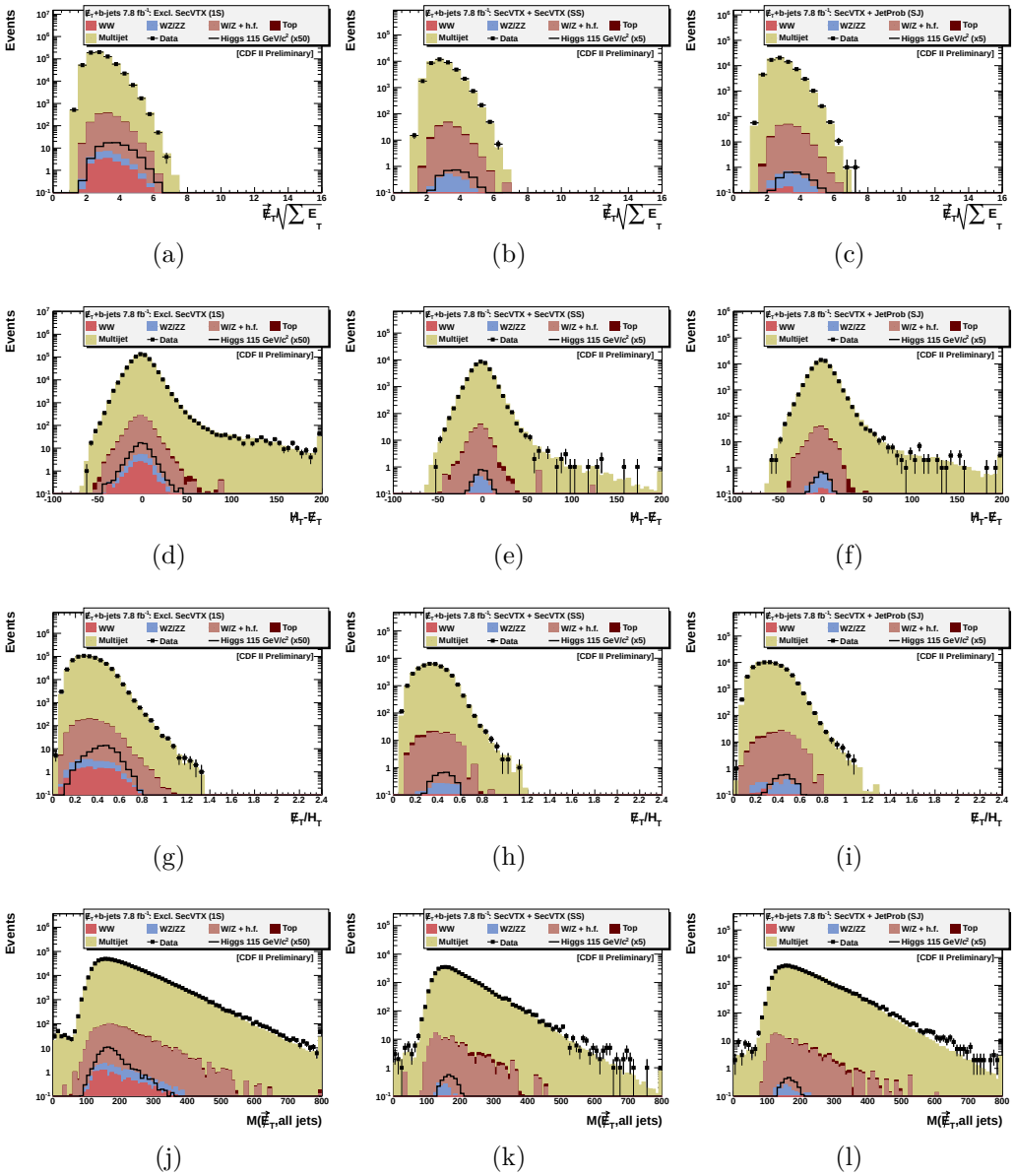


Figure A1.: Distribution of several input variables to NN_{QCD} in the TRM region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

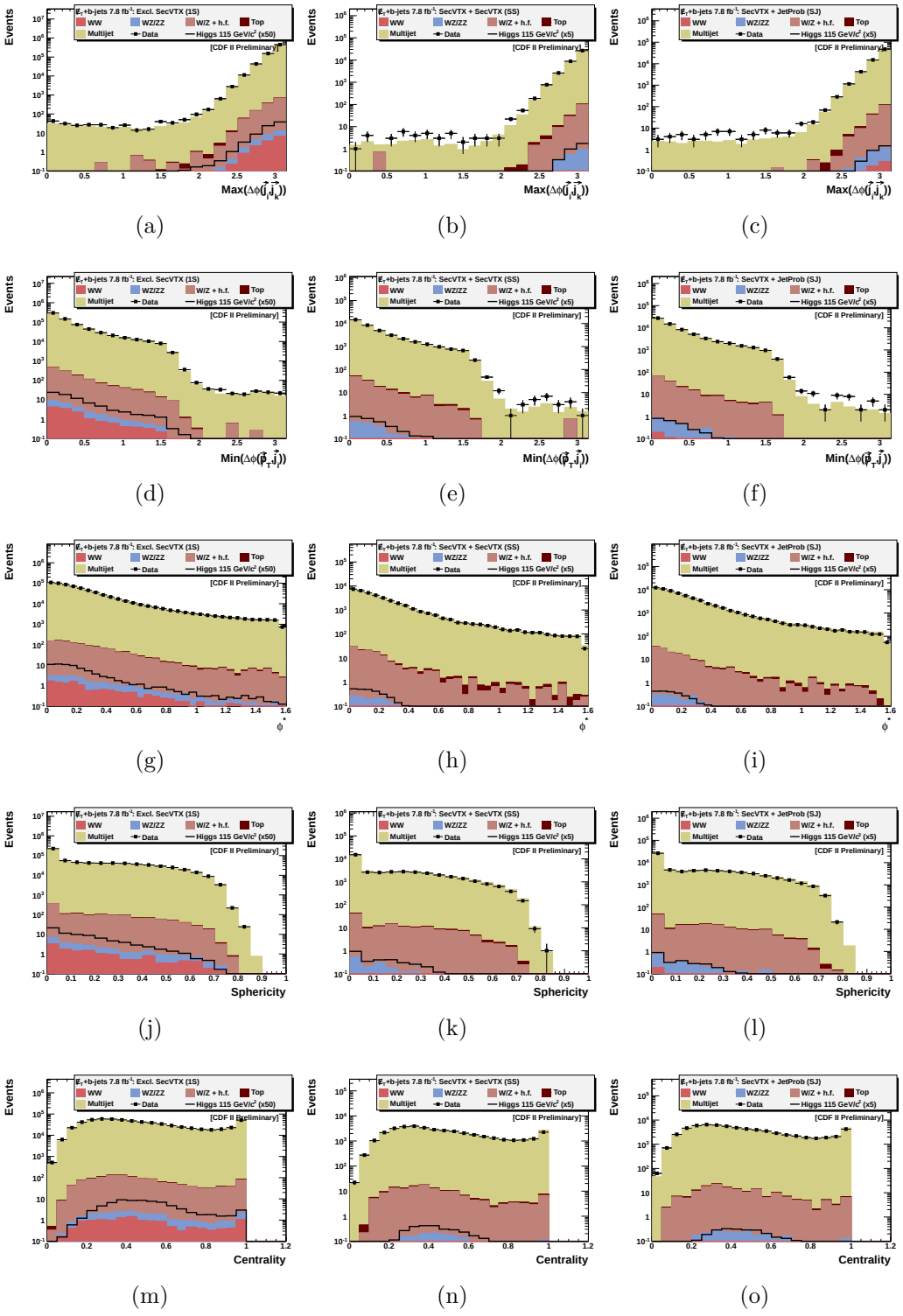


Figure A2.: Distribution of several input variables to NN_{QCD} in the TRM region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

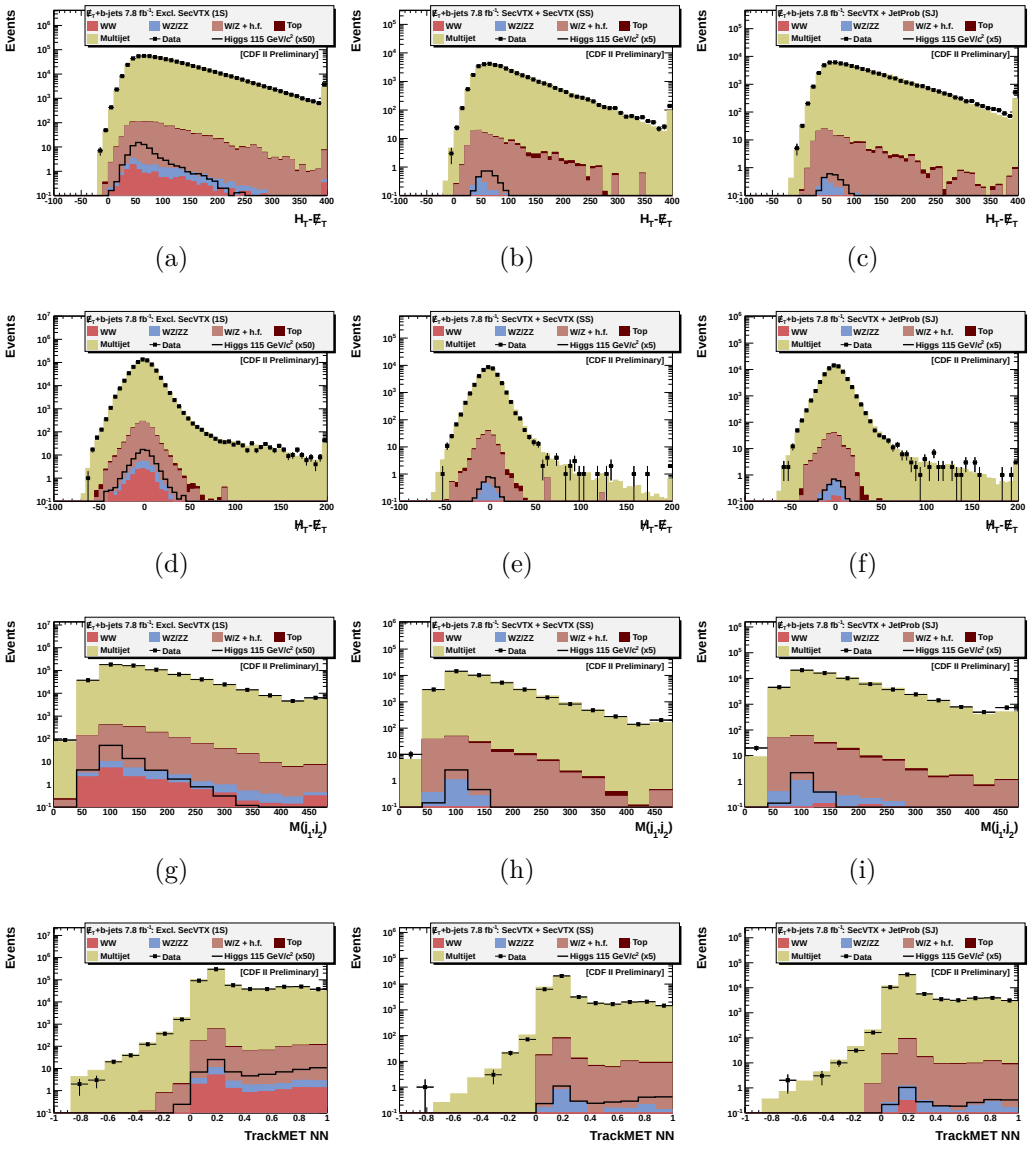


Figure A3.: Distribution of several input variables to NN_{SIG} in the TRM region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

C.2 QCD Control Region

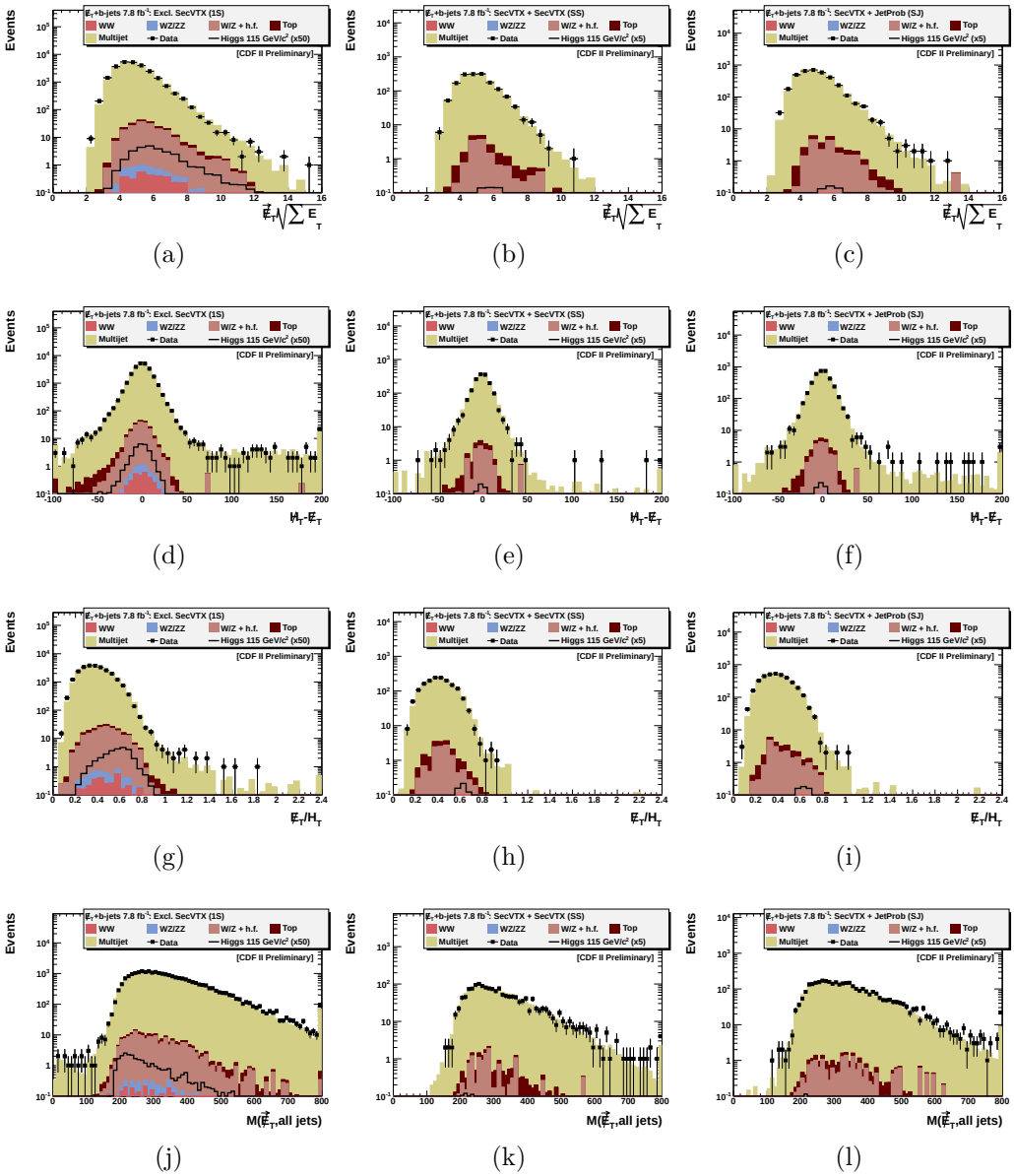


Figure A4.: Distribution of several input variables to NN_{QCD} in the QCD region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

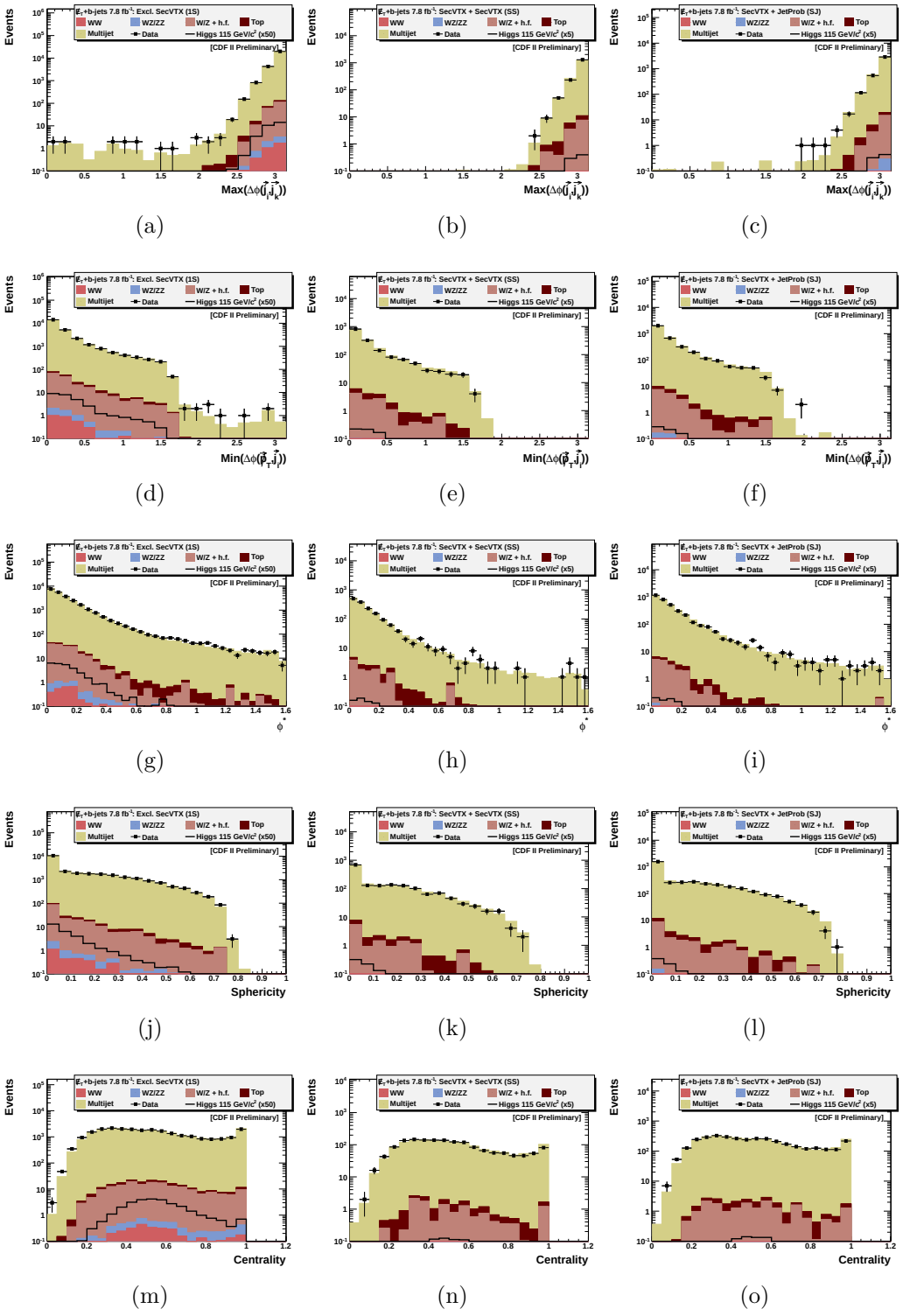


Figure A5.: Distribution of several input variables to NN_{QCD} in the QCD region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

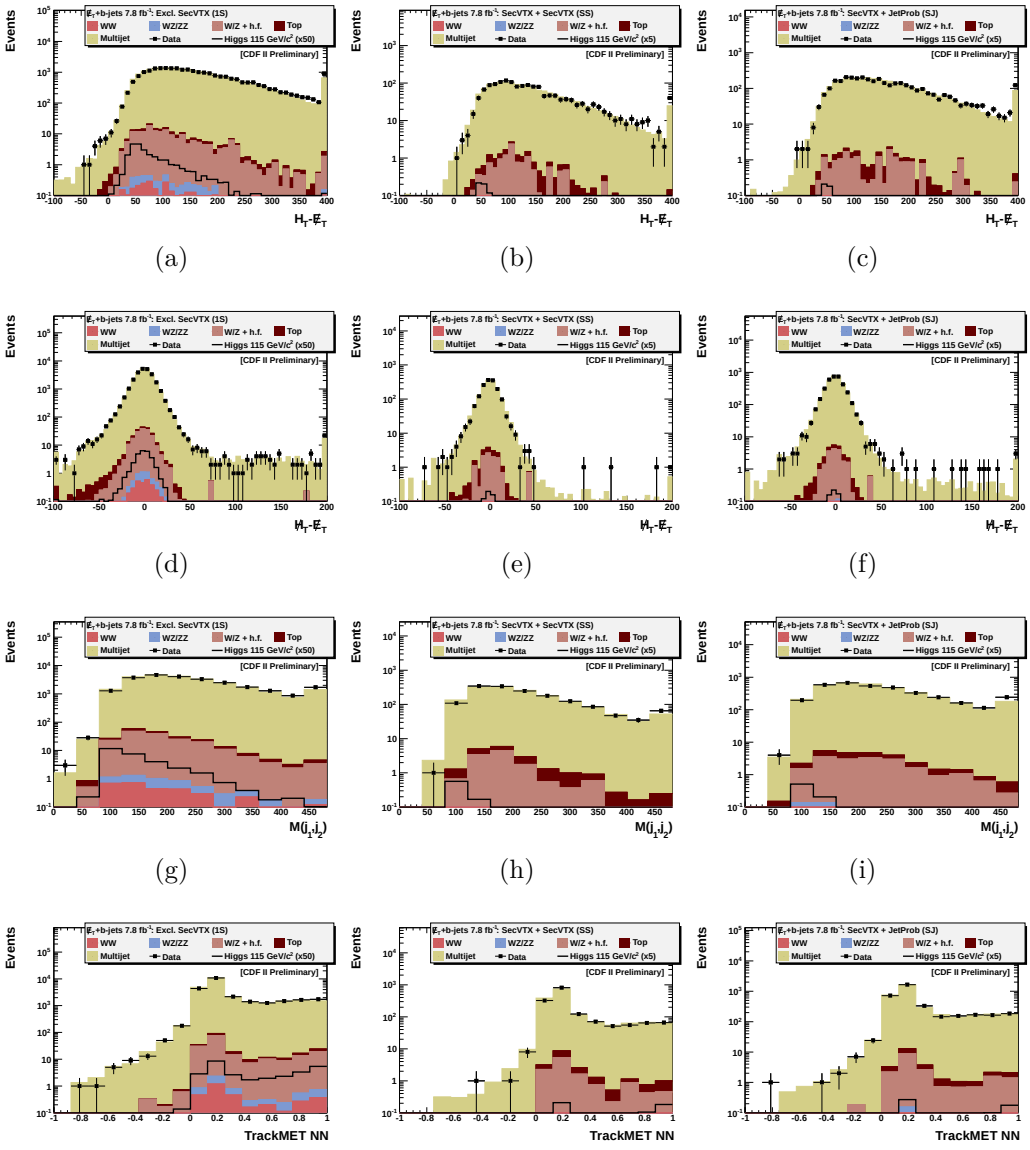


Figure A6.: Distribution of several input variables to NN_{SIG} in the QCD region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

C.3 EWK Control Region

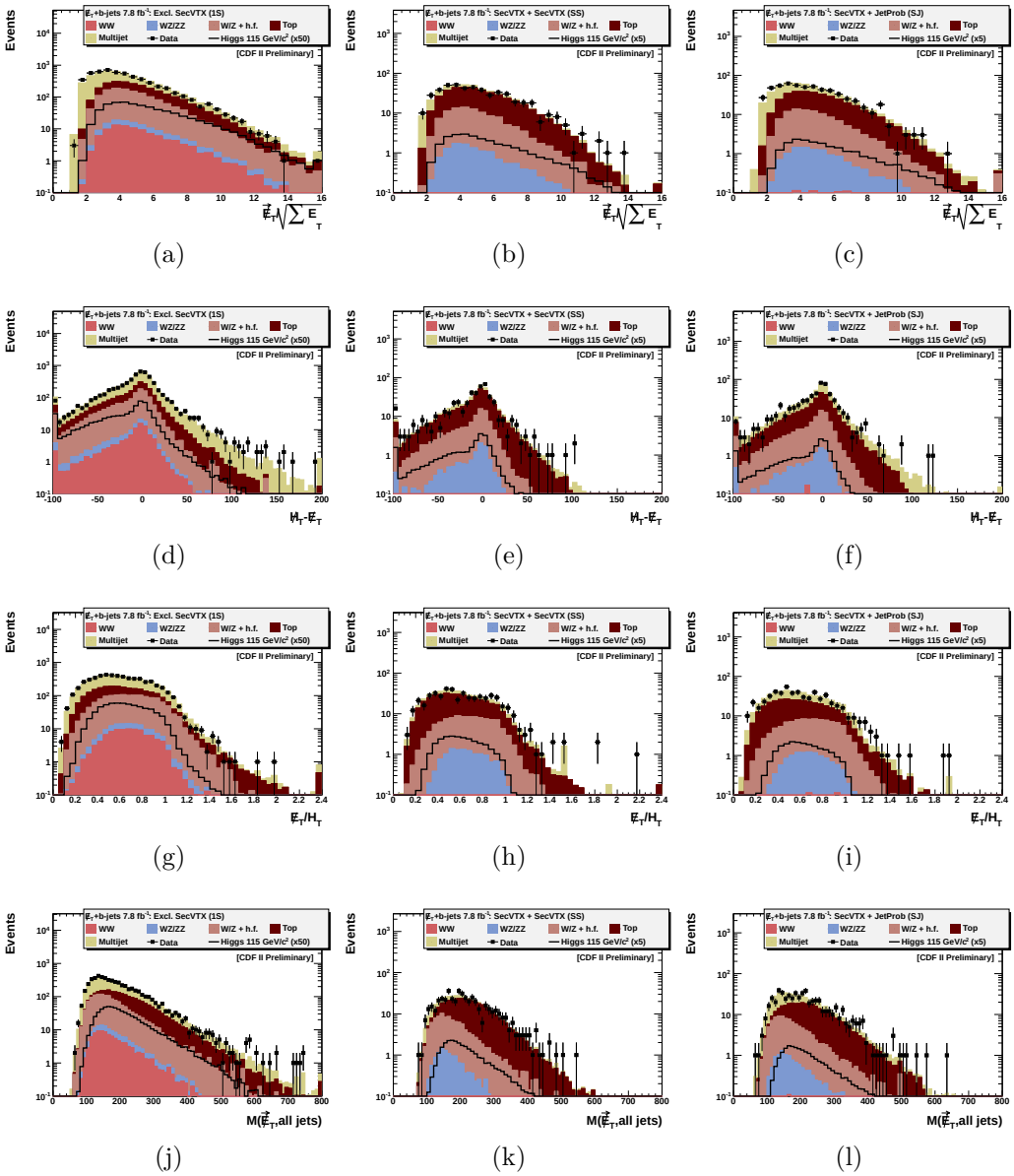


Figure A7.: Distribution of several input variables to NN_{QCD} in the EWK region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

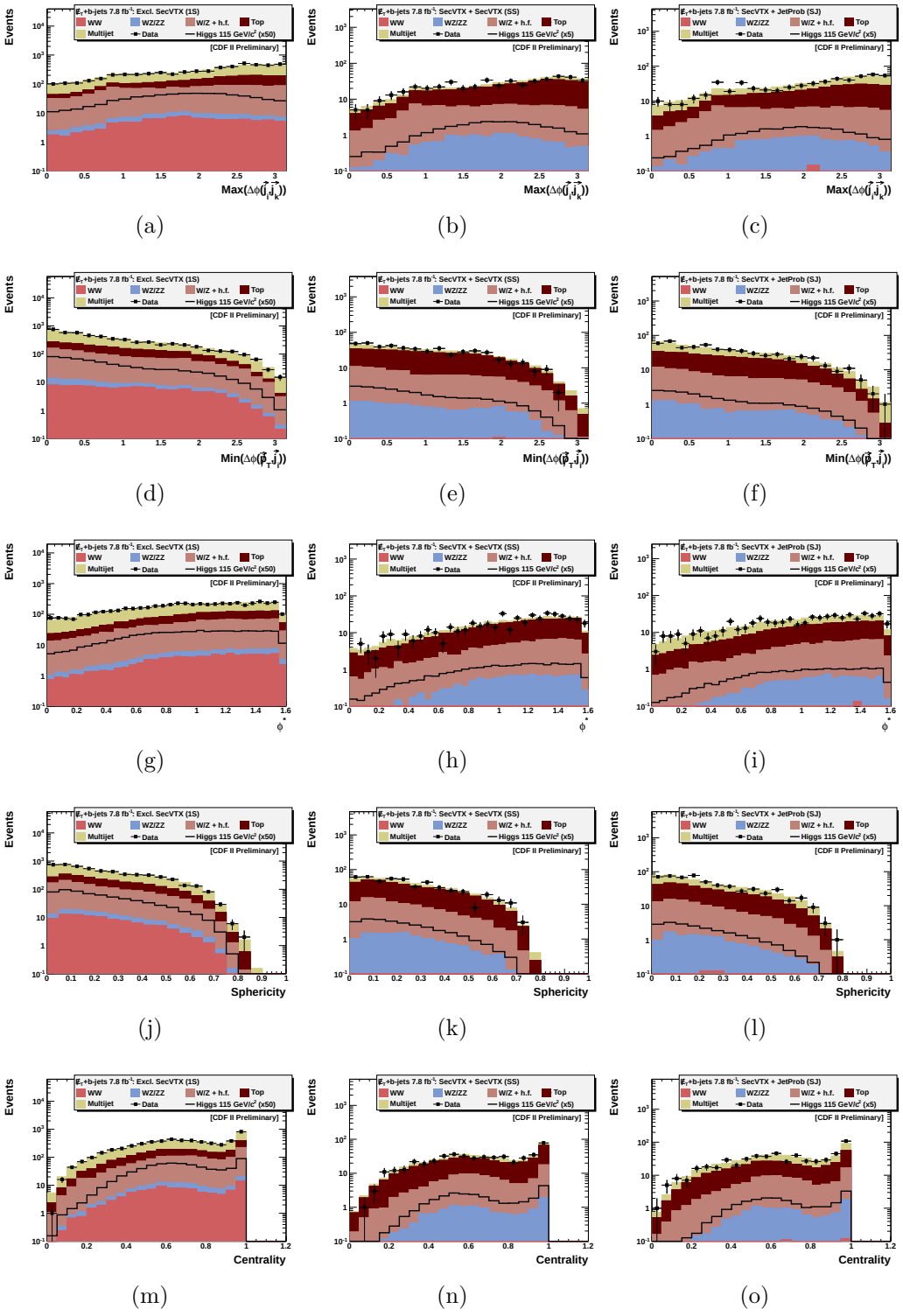


Figure A8.: Distribution of several input variables to NN_{QCD} in the EWK region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

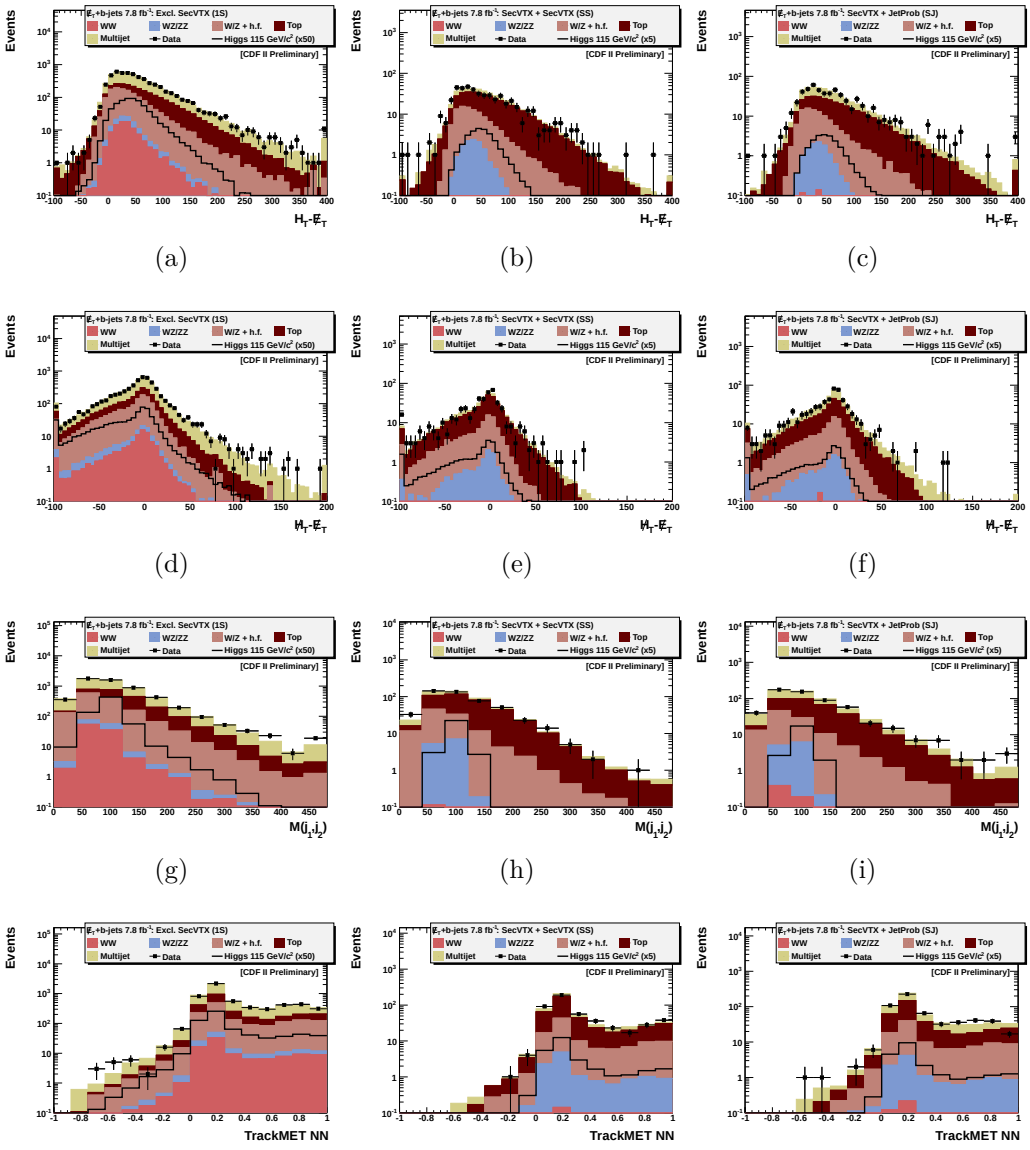


Figure A9.: Distribution of several input variables to NN_{SIG} in the EWK region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

C.4 Pre-selection

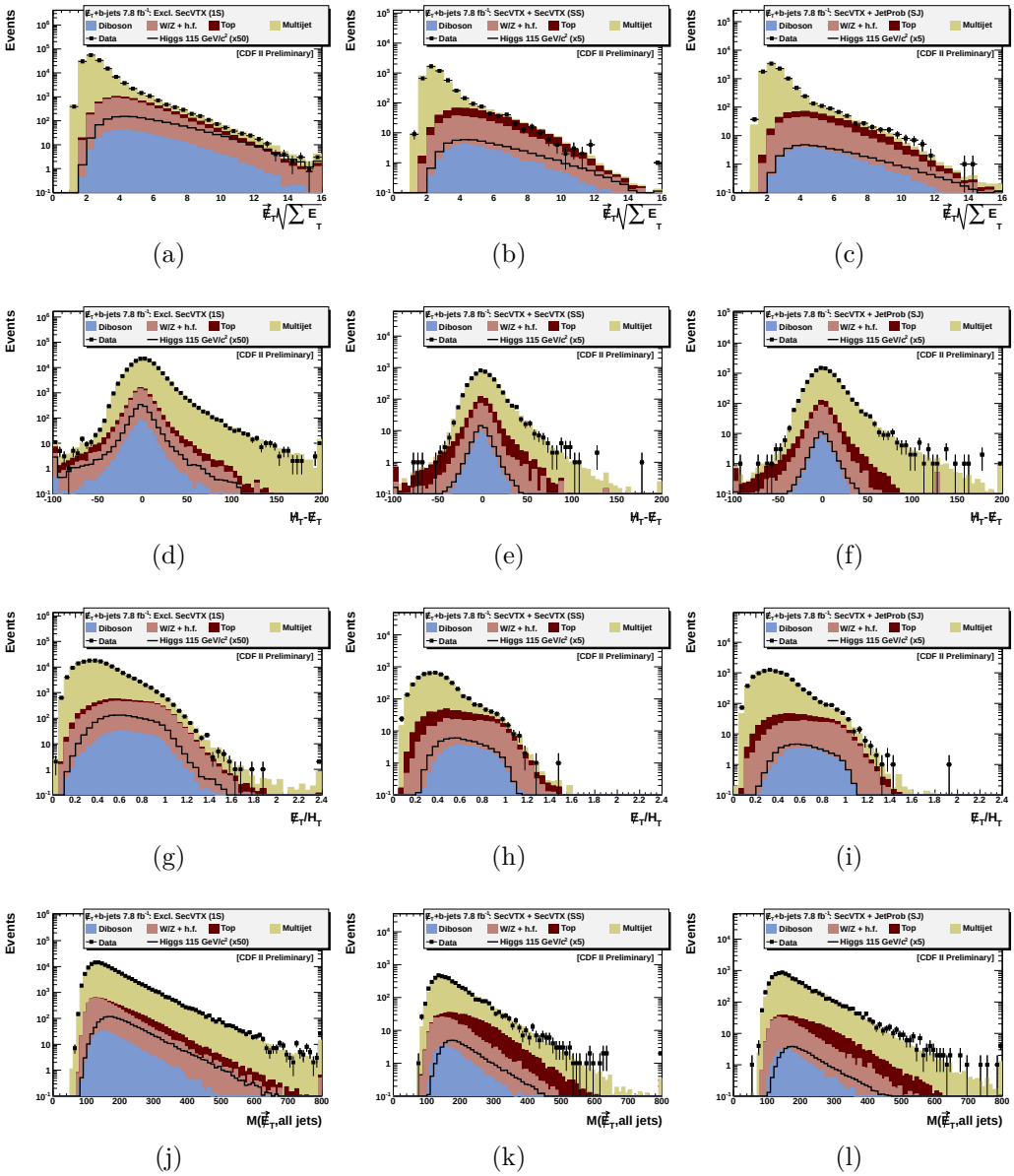


Figure A10.: Distribution of several input variables to NN_{QCD} in the pre-selection region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

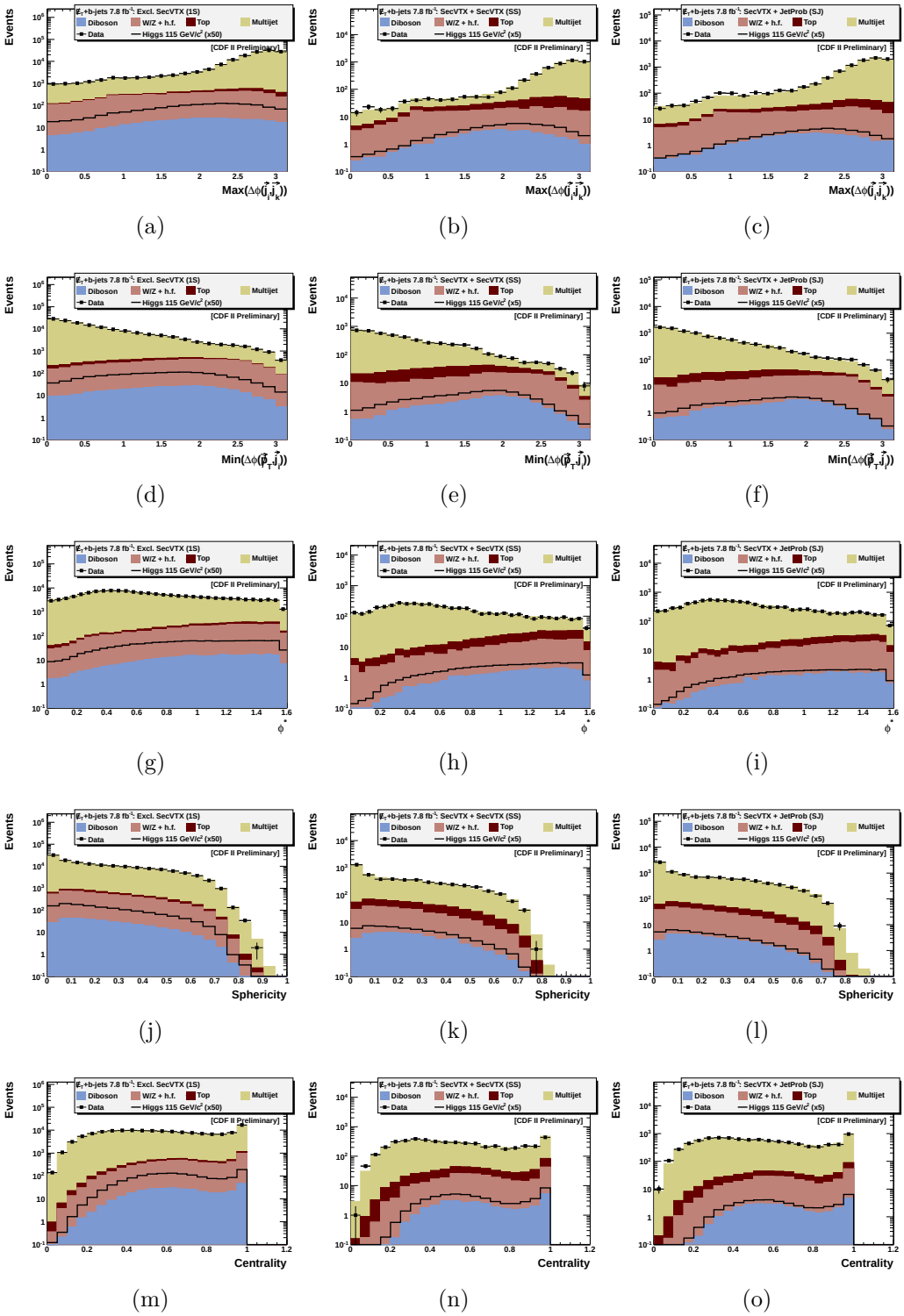
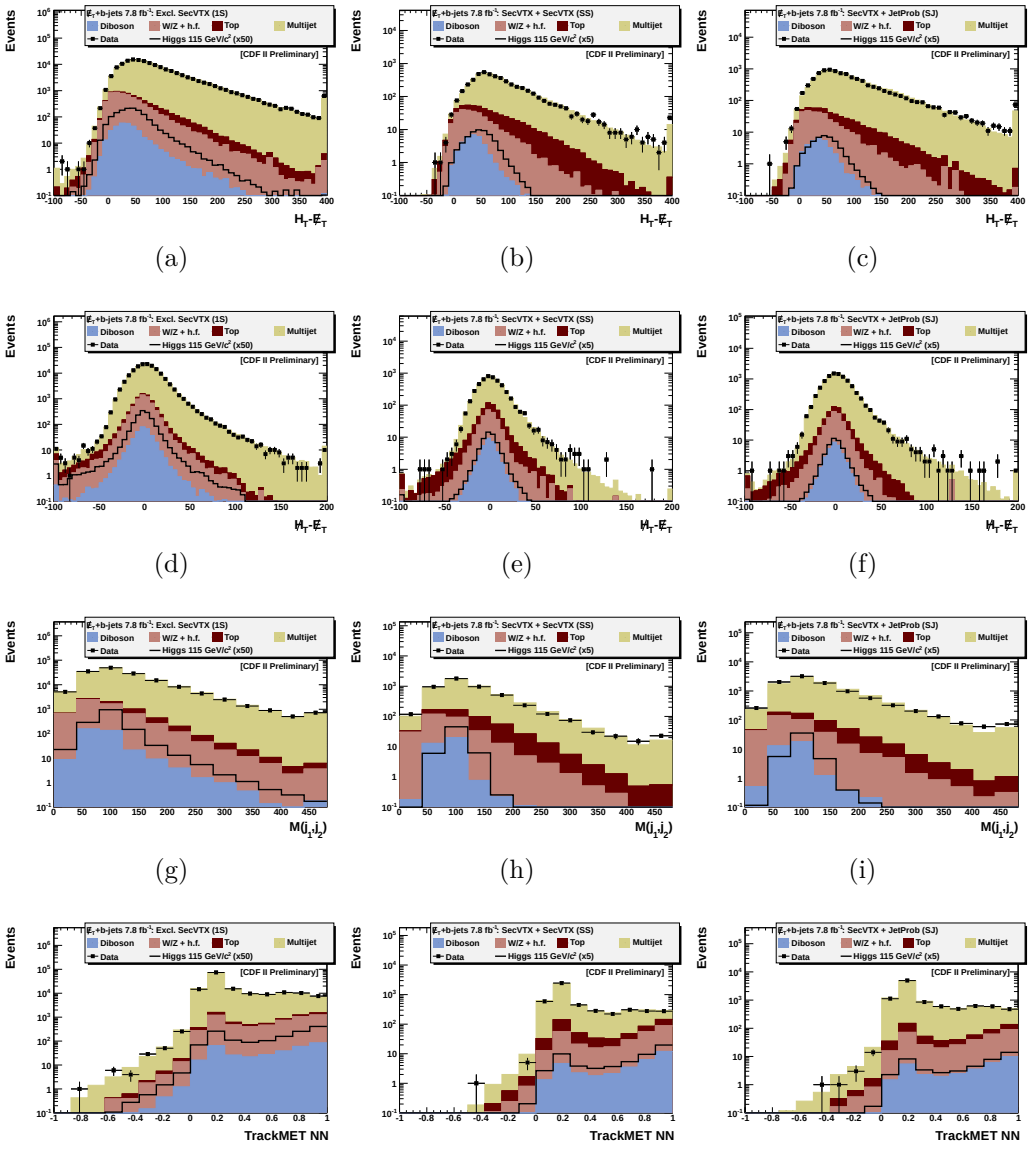


Figure A11.: Distribution of several input variables to NN_{QCD} in the pre-selection region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A12.: Distribution of several input variables to NN_{SIG} in the pre-selection region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). The stacked histograms are by construction normalized to the data.

C.5 Signal Region

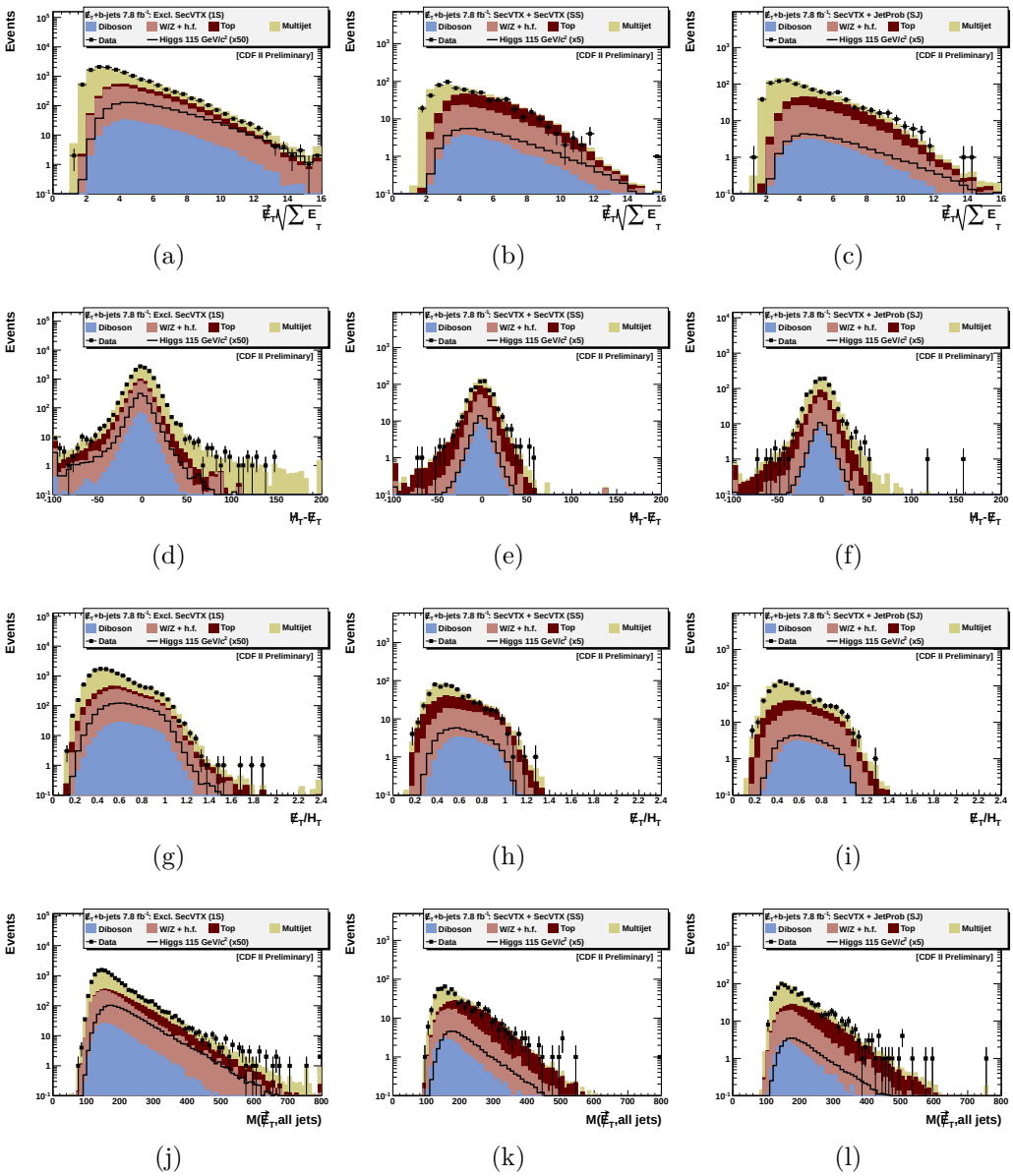


Figure A13.: Distribution of several input variables to NN_{QCD} in the signal region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). We obtain the normalization of the QCD background form events with $NN_{QCD} < 0.1$ (cf. section 8.7.5).

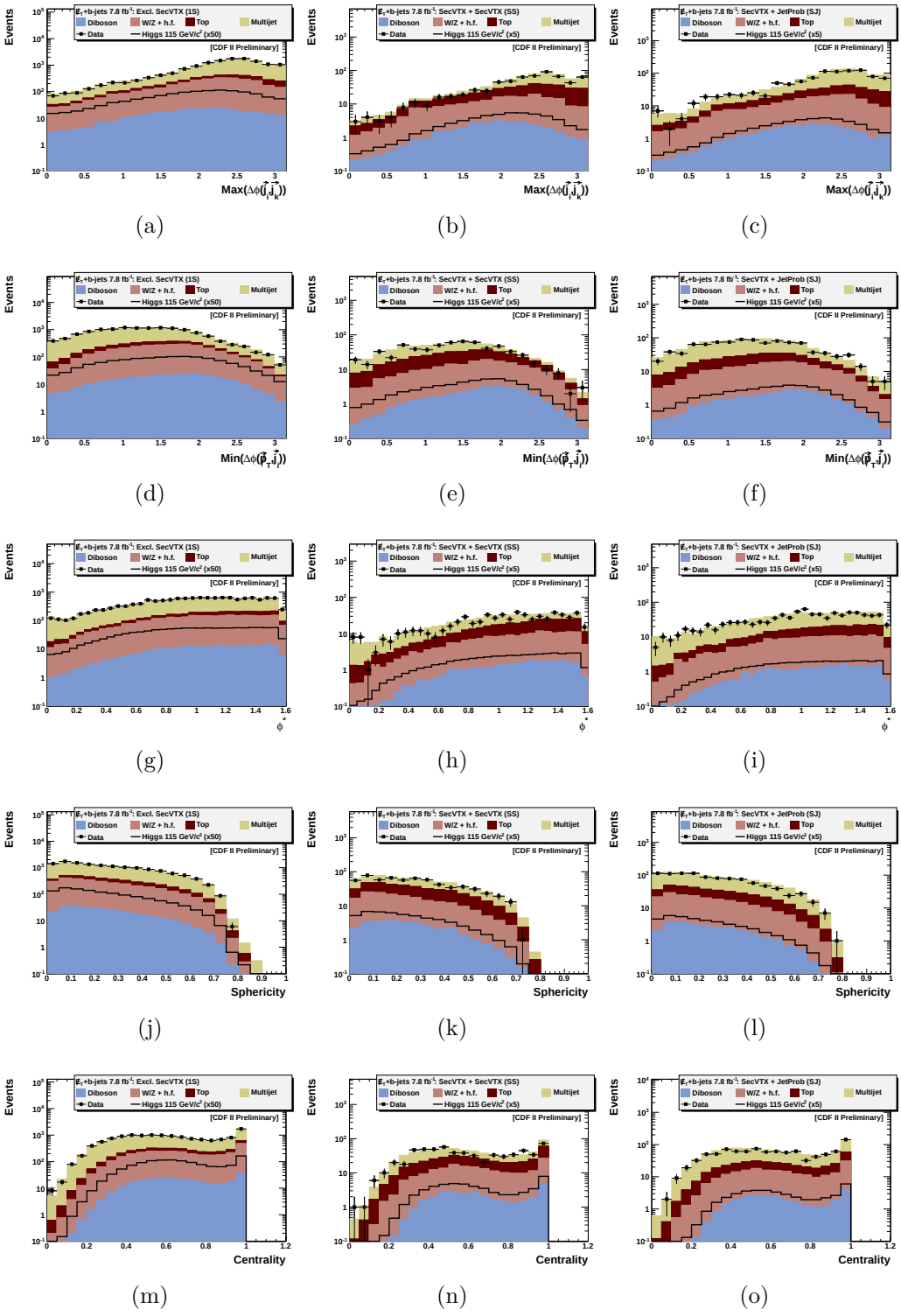


Figure A14.: Distribution of several input variables to NN_{QCD} in the signal region for exclusive SecVtx (left), SecVtx + SecVtx (center), and SecVtx + JetProb (right). We obtain the normalization of the QCD background form events with $NN_{QCD} < 0.1$ (cf. section 8.7.5).

D. MEASURING THE SINGLE TOP CROSS-SECTION AND THE $|V_{TB}|$ PARAMETER

D.1 Motivation

At the Tevatron, the dominant standard model (SM) mechanism, and the discovery mode, for top quark production in $p\bar{p}$ collisions is the production of $t\bar{t}$ pairs through strong interaction, with a cross section of approximately 7.0 pb [172]. Top quarks can also be produced singly through electroweak processes. The electroweak production of top quarks probes quark mixing. Its cross section is directly proportional to $|V_{tb}|^2$, thus allowing a measurement of this parameter in the Cabibbo-Kobayashi-Maskawa (CKM) matrix of quark mixing [33,34] without the assumption of the unitarity of the matrix. A value of V_{tb} significantly smaller than 1 could thus indicate the presence of a fourth family of quarks [213]. On the other hand, a V_{tb} value significantly greater than one could point to the existence of a heavy W-like boson enhancing the cross section. Moreover, the electroweak production of top quarks is a source of 100% polarized quarks. A review of new physics models affecting the single-top production cross section is given in [214].

At the Tevatron, single top can be produced together with a b-quark in the s-channel, or paired with a light quark in the t-channel. The t-channel process also has a large next-to-leading-order contribution which produces an additional b-quark. The Feynman diagrams for the above processes are shown in figure A1. The predicted single top production cross-sections are 0.88 ± 0.11 pb for the s-channel and 1.98 ± 0.25 pb for the t-channel [157, 160], for an assumed top quark mass of 175 GeV/c^2.

Electroweak production of top quarks is a difficult process to isolate due to the low cross section and very large backgrounds. The top quark has a lifetime roughly 10^{-25}s, and decays into a W boson and a b-quark almost 100% of the time. The W boson

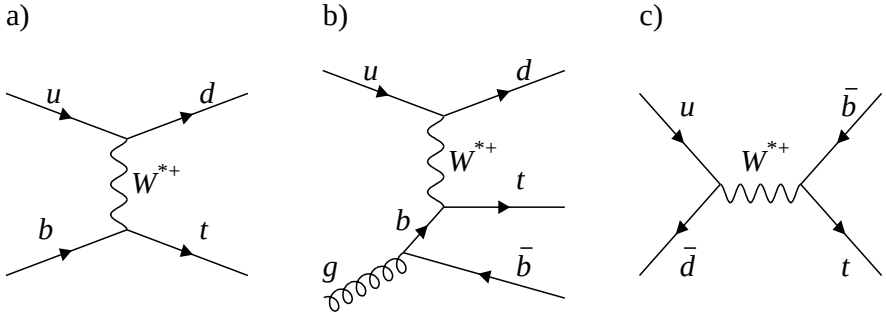


Figure A1.: Representative Feynman diagrams of single top quark production. Figures (a) and (b) are t-channel processes, and figure (c) is the s-channel process.

subsequently decays to either a quark-antiquark pair or a lepton pair. Events where decays as $W \to e\nu$ and $W \to \mu\nu$ [1] are the preferred identification mode at a hadron collider, due to the presence of the charged lepton and large missing transverse energy due the neutrino which suppress the large QCD background. The first evidence of electroweak top production has been achieved by looking at the above signature using $0.9\,\mathrm{fb}^{-1}$ by the D0 collaboration [215, 216]. A measurement in the same decay mode has also been released by the CDF collaboration using $2.2\,\mathrm{fb}^{-1}$ [217]. The D0 (CDF) published measurement reached a sensitivity of 3.4σ (3.7σ).

The above analyses all have in common the presence of identified leptons in their samples. We present here an analysis looking at events discarded by the other analyses, which uses for the first time events with a final state characterized by a many-jet topology, large missing transverse energy and a veto on reconstructed electrons or muons. The channel under study here has acceptance to $W \to \tau\nu$ decays, with subsequent hadronic τ decays; and to $W \to e\nu$ or $W \to \mu\nu$ decays where the e, μ are unidentified; this measurement is thus statistically independent from the other measurements.

[1] In addition to the aforementioned decays, there is extra acceptance to $W \to \tau\nu$ events where the tau decays to e or μ, and the latter appear isolated.

This analysis has been combined with the other CDF analyses performed in the sample with reconstructed leptons to achieve the $5.0\,\sigma$ excess over the backgrounds. In March 2009, both collaborations achieved for the first time the observation (5.0σ) level, using $2.6\,\mathrm{fb}^{-1}$ and up to $3.2\,\mathrm{fb}^{-1}$ respectively for the D0 collaboration [32] and the CDF collaboration [31].

The sample with no identified leptons is mainly dominated by QCD jet background, where a mismeasurement of one or more of the jet energies yields large energy imbalance in the detector. This analysis has the challenge of having much larger backgrounds with respect to the ones with identified leptons. Finally, the lack of reconstructed charged leptons, and the presence of the neutrino leave us with underconstrained kinematics, and the final state cannot be fully reconstructed. To improve the signal-to-background ratio, we select jets identified as originating from b-quarks using b-tagging algorithms. Even after these requirements, the signal-to-background ratio S/B is still too low to achieve sensitivity to this difficult process. Additional requirements based on the kinematic and topological characteristics of single top events are combined into a neural network and applied to the data.

The single top events analyzed here share the same signature as the search for the Higgs boson in association with a W or Z boson, where the Z decays to neutrinos, or the W decays leptonically, in either hadronic τs or unidentified e or μ. The methods used in the single top measurement are shared with the ones deployed in the Higgs search [114]. In that analysis, particular attention has been dedicated to increasing acceptance to events where the Higgs is produced toghether with a W boson, which has a very similar signature to the events which are here of interest. In order to reach such a goal, particular care has been devoted also to the background composition understanding and modeling.

Since the single top events analyzed here are a background to the Higgs search, measuring the single top production cross section in this sample means getting a step closer to reaching sensitivity to the Higgs signal, by giving us more confidence in the tools and techniques presented here.

The results reported here are based on the data taken with the CDF II detector between July 2002 and August 2007, corresponding to an integrated luminosity of 2.1 fb^{-1}.

The basic event pre-selection has been described in the previous chapters, together with the modeling of the physics processes we expect to be sensitive to. However, the results presented in this appendix were obtained before certain improvements in the technique. For instance, all MC modeling is based on PYTHIA except for the signal which is modeled using MADEVENT. We proceed here with the description of the steps specific to this analysis: the event selection, the discrimination of the single top signal from the backgrounds, the treatment of systematics, and the measurement of the single top production cross-section and of the V_{tb} element of the CKM matrix.

D.2 The Event Selection

D.2.1 Preselection and Topology Requirements

The event pre-selection, shared with the search for the standard model Higgs boson presented in ref. [114], is described in the previous chapter. It is briefly summarized below, commenting on the evolution of the sample properties.

To avoid trigger inefficiencies, we require a $\not\!\!E_T \geq 50$ GeV, leading jet energy $E_{T,1} \geq 35$ GeV, and second leading jet $E_{T,2} \geq 25$ GeV. We define the topology of the signal region by selecting events with a number of jets $2 \leq N_{\text{jets}} \leq 3$. In the case of three jet events, only jets with $E_T \geq 15$ GeV are considered. In order to be orthogonal to the sample used in single top cross-section measurements using events with identified electrons or muons, we veto events containing any well identified electrons or high-p_T isolated tracks.

About 523,000 events pass these preselection requirements. The signal-to-background ratio (S/B) for single top events after this selection is about 1/2800. At this stage of the analysis, most background is due to QCD events where mismeasurement and resolution effects give the very large $\not\!\!E_T$; these events are characterized by having

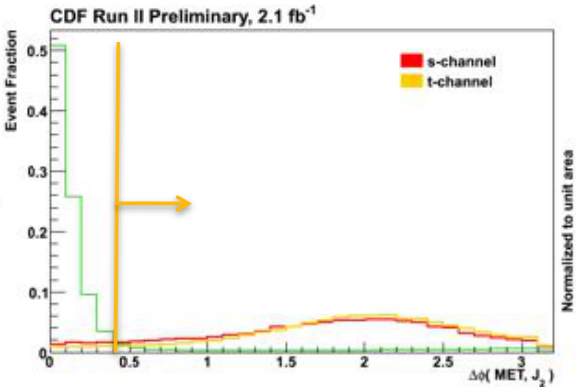


Figure A2.: Effect of cutting out events with $\not\!\!E_T$ aligned to the second jet.

the $\not\!\!E_T$ vector aligned in ϕ direction with one of the jets in the event. We require $\Delta\phi(\not\!\!E_T, j_1) \geq 1.5$ and $\Delta\phi(\not\!\!E_T, j_{2,3}) \geq 0.4$ to avoid such topologies. As can be seen in table D.1 and in figure A2, these cuts remove $\sim 10\%$ of the signal while removing around an order of magniture of QCD events, but still leave us with a S/B of $\sim 1/340$, where the majority of the background is still composed of QCD multijet production. Thus, we need to exploit additional properties of these events in order to further increase the purity of the sample.

Overview of selected single top events

After the event selection listed above, the number of signal events, subdivided in the three tagging categories, are presented in table D.1.

We checked the acceptance to hadronic decays of the W from single top events, which resulted very small (less than 2%). As can be seen in tables D.2 and D.3, we have acceptance roughly half of the time to events where the W decays to taus, and the rest split in electrons and muons. Since electrons and taus leave energy in the calorimeter, they can be reconstructed as jets by JetClu. As seen again in tables D.2 andD.3, this happens roughly 40% of the time (but mostly taus).

Monte Carlo variables are used to determine which single top decays we accept. We found that the largest category of single top decays is the $W \to \tau\nu$ ($\sim 50\%$ of

Selection cut	#Events s-channel	Acc.(%)	# Events t-channel	Acc.(%)	# Events DATA
Trigger driven selections	131.3	100%	200.4	100%	523207
Electron veto	102.9	78.3%	154.6	77.1%	512383
Isolated track veto	71.3	54.2%	110.0	54.8%	500813
$\not\!\!E_T$ not collinear to any jet	63.6	48.4%	101.0	50.3%	56301
SecVtx + SecVtx	8.4	13.1%	1.9	1.8%	190
SecVtx + JetProb	7.1	11.1%	1.9	1.8%	300
Exclusive SecVtx	18.5	29.1%	37.3	36.9%	3520

Table D.1: Acceptance of signal and backgrounds after sets of cuts in 2.1 fb^{-1}. The acceptance is relative to the events surviving the trigger requirements. The events were weighted by cross-section and efficiencies after applying the final selection

s-channel	$W \to e\nu$	$W \to \mu\nu$	$W \to \tau\nu$
all events	19%	30%	51%
2 jet events	20%	34%	46%
lepton-matched	1.2%	–	4.4%
3 jet events	17%	24%	59%
lepton-matched	5.4%	–	31%

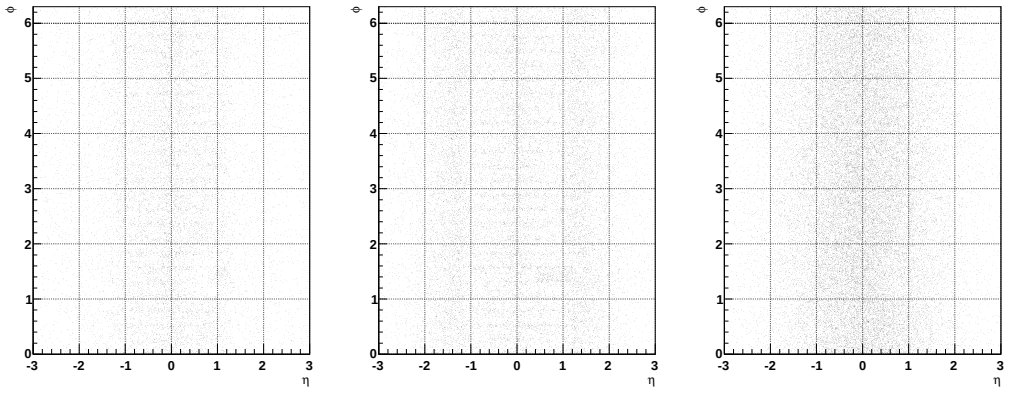
Table D.2: Contributions to 2/3 jet events from different leptonic decay modes of the W-boson in single top s-channel events

t-channel	$W \to e\nu$	$W \to \mu\nu$	$W \to \tau\nu$
all events	19%	29%	51%
2 jet events	21%	32%	47%
lepton-matched	3.3%	–	6.9%
3 jet events	17%	23%	61%
lepton-matched	6.5%	–	36%

Table D.3: Contributions to 2/3 jet events from different leptonic decay modes of the W-boson in single top t-channel events

the times), followed by $W \to \mu\nu$ ($\sim 30\%$)and $W \to e\nu$ ($\sim 20\%$). We also look at the direction of the lepton using the single top Monte Carlo interfaced to the detector simulation. The τ decays are mostly hadronic. They are emitted uniformly along ϕ, and preferentially in the region $|\eta| < 1$. The electron events show a clear pattern which

highlights the cracks in the ϕ and η directions in the EM calorimeter. Muon events are the less central of the three decay types and populate the uninstrumented regions of the muon system. The granularity along the ϕ direction follows the cracks in the EM calorimeter, hence not allowing identification of the muons. The uninstrumented region at $\eta \in [0.5 - 1]$ and $\phi \simeq \pi/2$, a service gap to accommodate for the cryogenic electronics for the solenoid, is clearly revealed. The majority of τ decays we accept are hadronic. Therefore, the angular distribution of $W \to \tau\nu$ events is not biased by vetoing the leptons and reflects the production kinematics.



(a) $l = e$ (b) $l = \mu$ (c) $l = \tau$

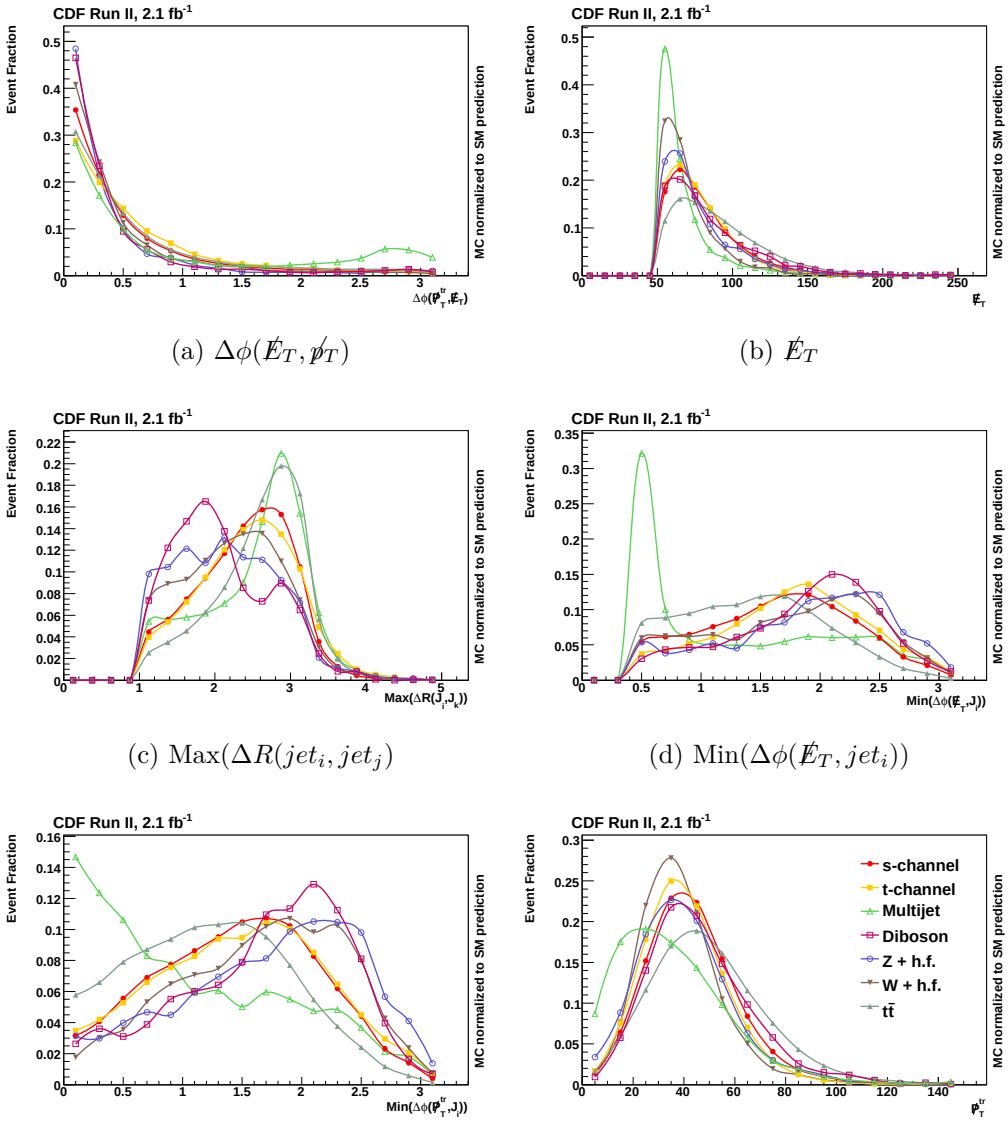
Figure A3.: ϕ vs. η distribution of the lepton from $W \to l\nu$ decay in s-channel.

D.2.2 Neural Network based event Selection

We introduce here a neural network approach whose primary goal is to recognize and separate QCD multijet events with mis-measured jets, in which the $\not\!\!E_T$ is instrumental, from events with real $\not\!\!E_T$ from neutrinos. In addition, the neural network is designed to also reject events with mis-tagged light flavor jets. The network is the *Multi Layer Perceptron* [176, 218] network architecture as implemented in the TMVA package [184], found in ROOT [97].

Table D.4: Input variables to the QCD-removing neural network.

Variable
Absolute amount of the $\not\!\!E_T$
Absolute amount of the $\not\!\!p_T$
$\Delta\phi$ between $\not\!\!E_T$ and $\not\!\!p_T$
Maximum of ΔR between two jets
Minimum of $\Delta\phi$ between the $\not\!\!E_T$ and each jet
Minimum of $\Delta\phi$ between the $\not\!\!p_T$ and the jets
Maximum of $\Delta\phi$ between two jets
$\not\!\!H_T$ / $\not\!\!E_T$
$\Delta\phi(j_1, j_2)$ in the 2-jet rest frame
$\not\!\!E_T/H_T$
$\not\!\!E_T/\sqrt{\sum E_T}$
Invariant mass of $\not\!\!E_T$, j1 and j2
$\sum P_T^{chgd}/P_T^{jet}$ for the leading jet
$\sum P_T^{chgd}/P_T^{jet}$ for the 2nd leading jet
Sphericity

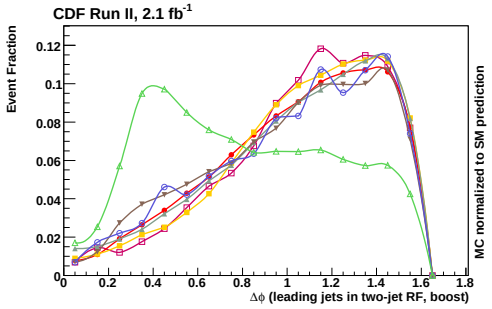


(a) $\Delta\phi(\not\!\!E_T, \not\!\!p_T)$ (b) $\not\!\!E_T$

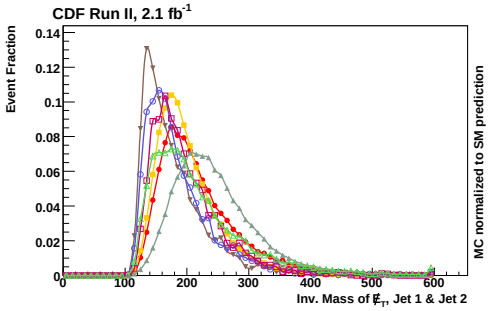
(c) $\mathrm{Max}(\Delta R(jet_i, jet_j)$ (d) $\mathrm{Min}(\Delta\phi(\not\!\!E_T, jet_i))$

(e) $\mathrm{Min}(\Delta\phi(\not\!\!p_T, jet_i))$ (f) $\not\!\!p_T$

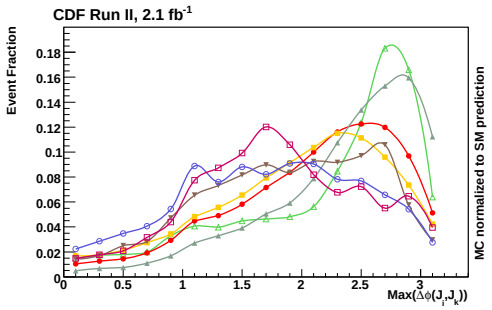
Figure A4.: Kinematic distributions for the single top s- and t-channel Monte Carlo (full circle and full square, respectively) events, and background events, for events passing the event preselection. All histograms are normalized to unity.



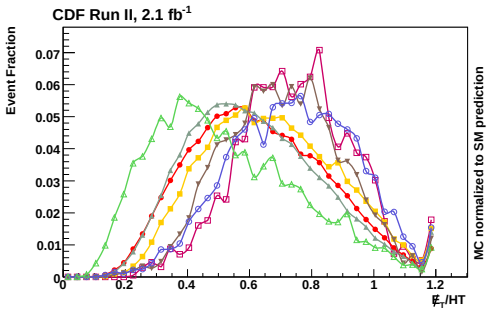
(a) $\Delta\phi(jet_1, jet_2)$ with respect to the dijet system direction



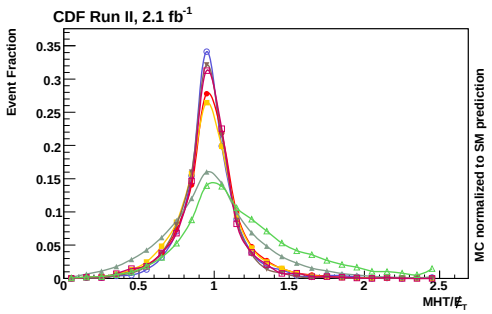
(b) Invariant mass of $\not\!\!E_T, jet_1$ and jet_2



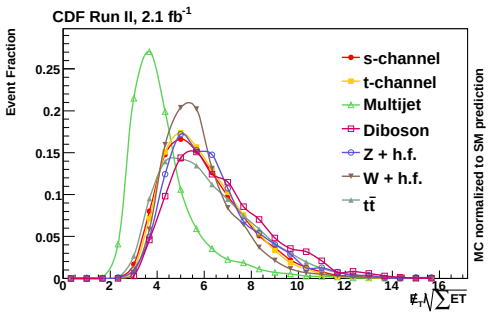
(c) Max($\Delta\phi(jet_i, jet_j)$)



(d) $\not\!\!E_T/H_T$

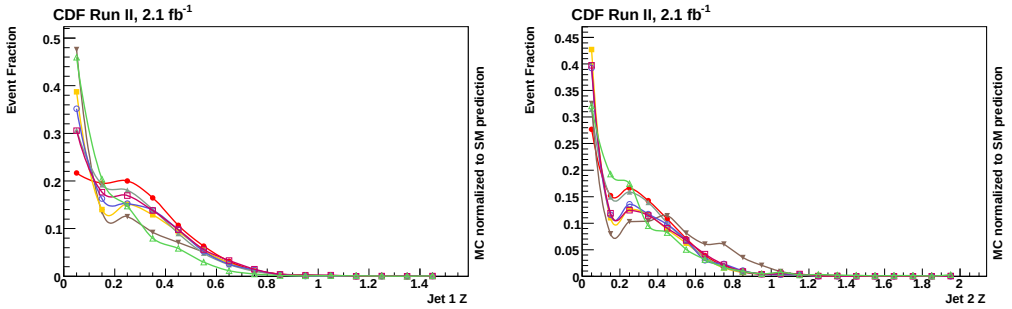


(e) $\not\!\!H_T/\not\!\!E_T$

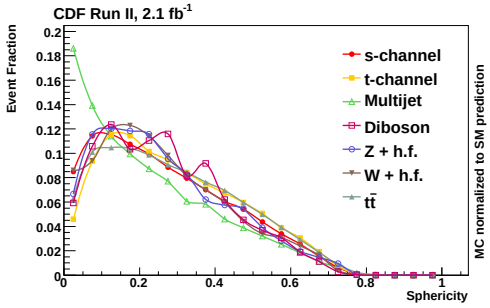


(f) $\not\!\!E_T/\sqrt{\sum E_T}$

Figure A5.: Kinematic distributions for the single top s- and t-channel Monte Carlo (full circle and full square, respectively) events, and background events, for events passing the event preselection. All histograms are normalized to unity.



(a) $\sum P_T^{chgd}/P_T^{jet_1}$ (b) $\sum P_T^{chgd}/P_T^{jet_2}$



(c) Sphericity

Figure A6.: Kinematic distributions for the single top s- and t-channel Monte Carlo (full circle and full square, respectively) events, and background events, for events passing the event preselection. All histograms are normalized to unity.

We use variables which describe the energy flow in the detector: the absolute amount of the missing transverse momentum $\not\!\!p_T = |-\sum_{\text{tracks}} \vec{P_T^i}|$, the absolute amount of the missing transverse energy $\not\!\!E_T$, the $\not\!\!E_T$ significance, $\not\!\!E_T/\sqrt{\sum E_T}$, the ratio of $\not\!\!E_T$ and H_T, the ratio of $\not\!\!H_T$ and $\not\!\!E_T$, and the invariant mass of the $\not\!\!E_T$ and the two leading jets. We also use variables that discriminate between heavy flavor quark jets, and jets originating from light flavor quarks or gluons: $\sum P_T^{chgd}/P_T^{jet}$ for the leading and second leading jets, where P_T^{chgd} is the transverse momentum of a charged track. We use the following angular informations: the difference in ϕ between $\not\!\!E_T$ and $\not\!\!p_T$, the maximum of the difference in the R space between two jets, taking two jets at the time; the minimum of the difference in ϕ between the $\not\!\!E_T$ and each jet j_i, considering all two or three pairings; the minimum of the difference in ϕ between the $\not\!\!p_T$ and the jets, considering all two or three pairings; the maximum of the difference in ϕ between two jets directions, taking two jets at the time; the $\Delta\phi$ between the direction of the leading jets in two-jet rest frame and the direction of the boost; the event sphericity S = 1.5 $\times(\lambda_2 + \lambda_3)$ where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ are the eigenvalues of the sphericity tensor $M_{ab} = \sum_j P_{ja} P_{jb}$. The 15 variables used as inputs to the neural network are summarized in table D.4. Comparisons of the kinematic distributions for background and signal events for the 15 variables are shown in figures A4 to A6, where for simplicity only events containing at least one SecVtx-tagged jet are shown. The investigated variables all have some discriminating power.

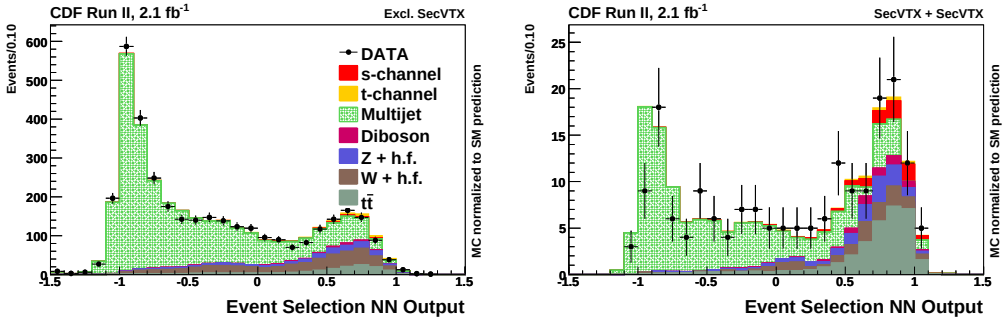
The use of tracking information provides excellent handles on QCD multijet events and mistags. The $\not\!\!p_T$ estimated from tracks is more correlated to the neutrino's energy and direction for central neutrinos, as it is less affected by resolution effects than $\not\!\!E_T$ from calorimeter information (but has less coverage in η).

Besides tracking related variables, the angular separation between $\not\!\!E_T$ and the jets, the $\not\!\!E_T$ significance and the $\Delta\phi$ between the direction of the leading jets in two-jet rest frame and the direction of the boost are very useful for discrimination. However, the latter can be improved by using the correlation between the input variables. In that respect, the use of a conventional cut-based approach is clearly not optimal, for it

cannot exploit that correlation. There is an incentive to use a multivariate technique, here a neural network, to obtain a better result than a cut-based approach, combine all the information contained in the variables, and give as output a single distribution that will be used to cut away the QCD and the mistags.

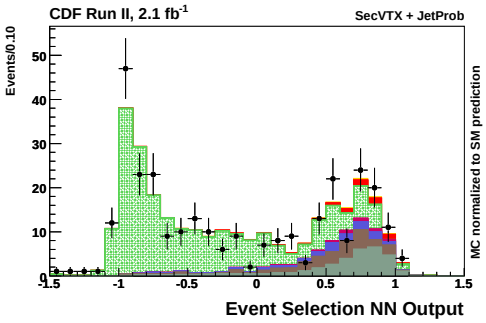
The QCD background kinematics do not vary significantly with the heavy flavor content and one neural network, trained with events with one SecVtx-tagged jet, is used in the b-tagged subsamples. The single top signal used for the training is a mixture of 50% s-channel events and 50% t-channel events, which is roughly what we expect to accept after the event selection described in the previous paragraph. The reference top mass chosen for the training is $M_{\rm top} = 175\,{\rm GeV}/c^2$. We use only events with at least one SecVtx-tagged jet for the training. For the background, we only use the QCD multijet background model previously described in section 8.7. This procedure is new, for in the Higgs analysis the neural network was trained using a Monte Carlo simulation for QCD. The neuron activation function is the hyperbolic tangent. The other parameters are set to the default values in TMVA. We train 500 epochs and check the convergence plots and found there is no overtraining. The TMVA package splits our samples in two: one for training the neural network and one for testing the network output for overtraining. Both samples contain 21 000 signal and 14 000 QCD multijet background events. Among the configurations investigated the one which behaves the best, providing the largest expected S/B, has all the 15 variables defined above as inputs, two hidden layers with 30 and 15 hidden nodes respectively, and 1 output node. Figure A7 shows the value of the output node, NN_{QCD}.

As mentioned above, we select the signal region to maximize signal significance keeping high signal efficiency. By cutting on the $NN_{QCD} > -0.1$, we were able to reduce the main background, due to QCD multijet events faking high $\not\!\!E_T$, by 77% (the overall background was reduced by 65%) while keeping 91% of the signal. We improve the signal significance $S/\sqrt{S+B}$ by 50% and the S/B ratio by 150%, from 1/50 to 1/20. This approach greatly outperforms a cut-based event selection. All



(a) Events with exactly one SecVtx-tagged jet

(b) Events with two SecVtx-tagged jets



(c) Events with one SecVtx-tagged and one JetProb-tagged jet

Figure A7.: NN_{QCD} output for events passing the event pre-selection, separated according to flavor content. As can be seen from the plot, the kinematics of the QCD background are very different from the other backgrounds. These events are largely removed cutting on $NN_{QCD} < -0.1$. In the signal region with $NN_{QCD} > -0.1$, the excess of single top events starts to be visible over the remaining backgrounds.

variables provide a description of events with $\not\!\!E_T$ coming from mismeasured jets or of events with mistags and help improving the NN performance. On the other hand, the use of additional variables negligibly affects performance.

Normalization of the QCD Multijet Processes

We use the region defined with $NN_{QCD} < -0.1$ to compute the normalization for QCD multijet production comparing the prediction given by the weighting the events for the tag rate parameterization, and the observed number of data events. We found the need to multiply the prediction by scale factors which depend on the heavy flavor content of the sample. The values extracted are 1.08 ± 0.05 for events with a jet tagged by the SecVtx algorithm, 0.79 ± 0.10 for events in the SecVtx + SecVtx category, and 0.76 ± 0.07 for events in the TJ category.

Those numbers are obtained by propagating the errors:

$$\text{SF} = \frac{\text{DATA - Exp. (MC) Signal - Exp. MC Background}}{\text{QCD (before SF)}} \tag{D.1}$$

$$\frac{\Delta\text{SF}}{\text{SF}} = \sqrt{\left(\frac{\text{MC}_{\text{ERR}}}{\text{DATA - MC}}\right)^2 + \left(\frac{\text{QCD}_{\text{ERR}}}{\text{QCD}}\right)^2}, \tag{D.2}$$

where MC is the number of expected events from Monte Carlo simulation (including signal), QCD is the QCD prediction before applying the scale factor, QCD_{ERR} is the statistical error on the QCD prediction (before SF) and MC_{ERR} is is computed adding the errors (stat.+syst.) on its components in quadrature. The values fo these for each tagging category are shown in table D.5.

D.2.3 Event Yields in Signal Region

In Table D.6 we show the contribution of signal and background events in the signal sample, subdivided according to the heavy flavor content of the sample.

Process	1b-tag	2 b-tags (SS)	2 b-tags (SJ)
DATA	2343	73	166
QCD before SF	1992 +/- 65	86.9 +/- 11.3	207 +/- 20
Exp. (MC) Signal	5.9 +/- 0.2	0.46 +/- 0.05	0.65 +/- 0.06
Exp. MC Background	188 +/- 66	4.4 +/- 1.2	8.0 +/- 2.1

Table D.5: Number of expected and observed events in the $NN_{QCD} < 0.1$ region in all tagging categories used to derive the QCD normalization scale factor.

Process	1b-tag	2 b-tags(SS)	2 b-tags(SJ)
s-channel	15.7±2.0	7.6±0.9	6.3±0.8
t-channel	31.2±4.9	1.7±0.2	1.6±0.2
$t\bar{t}$	125±23	30.3±5.8	29.2±5.7
WW/WZ/ZZ	33.0±6.5	4.9±0.6	4.2±0.6
W + h.f.	269±113	12.7±7.5	22.7±13.7
Z + h.f	105±53	11.8±5.8	11.8±6.0
QCD Multijet	592±27	28.9±3.8	58.5±5.8
Signal	46.9±5.2	9.3±1.0	7.9±0.8
Background	1125±169	89±15	126±21
Total	1172±169	98±15	134±21
Data ($2.1\,\mathrm{fb}^{-1}$)	1167	113	131

Table D.6: Number of expected and observed events in the signal region in all tagging categories.

After requiring $NN_{QCD} > -0.1$, the dominating backgrounds are QCD multijet production, W/Z + jets and $t\bar{t}$. We studied the dynamic of those events to develop a NN with the goal of discriminating the surviving backgrounds from the interesting signal. It should be noted at this stage that the "QCD multijet" background in the signal region consists more of events with mis-tagged W/Z + heavy flavor jets rather than QCD production of jets.

D.3 Control Regions

Before looking at any signal, we must make sure that our modeling of the various backgrounds is robust and that using it will not bias our results. For this purpose, we test our understanding of observed standard model processes in several *control* regions that are depleted of signal events. We define them *a priori* based on the event topologies we expect from each type of processes. The events in the kinematic regions were our signal is expected, the *signal box*, is excluded at first, until the method of analysis, selection cuts, background estimates and systematic uncertainties are fixed. This method is that of the *blind analysis*. The signal box is larger than our actual signal region, to ensure no bias. Only when a good agreement in the control regions is achieved, i.e. experimental data agrees with our model within uncertainties, are the final optimizations for the final signal region (which is still hidden at this stage). As a final step, we open the signal box and compare our predictions to observed data and proceed to the measurements.

We now proceed with the description of the control regions used to test our data-driven and Monte Carlo background modeling.

D.3.1 The QCD Control Region

After our event pre-selection, the $\not\!\!E_T$+jets data sample is mainly composed of QCD multijet production (including light flavor mis-tagged jets). The $\not\!\!E_T$ is these events comes mainly from mis-measurement of the jet energies, which cause the $\not\!\!E_T$ to point along one of the jets, but also from real sources of $\not\!\!E_T$ such as neutrinos or muons from semi-leptonic b-decays. The data-driven method used to describe these events is defined in 8.7. It is based on events with no leptons, two or three high energy jets, $50 \leq \not\!\!E_T \leq 70$ GeV, and $\Delta\phi(\not\!\!E_T, j_2) < 0.4$. These form the TRM region.

In addition to the checks performed in the TRM region while building the model, we use the QCD control region to cross-check our data-driven modeling. This region is made of events with no leptons, two or three high energy jets, $\not\!\!E_T > 70$ GeV, and

$\Delta\phi(\not{E}_T, j_2) > 0.4$ and is a high statistics region containing almost exclusively QCD events.

Figures A8 to A10 show the validation against data of the kinematic distribution used as input to the NN_{QCD} in the QCD control region and for the three tagging categories.

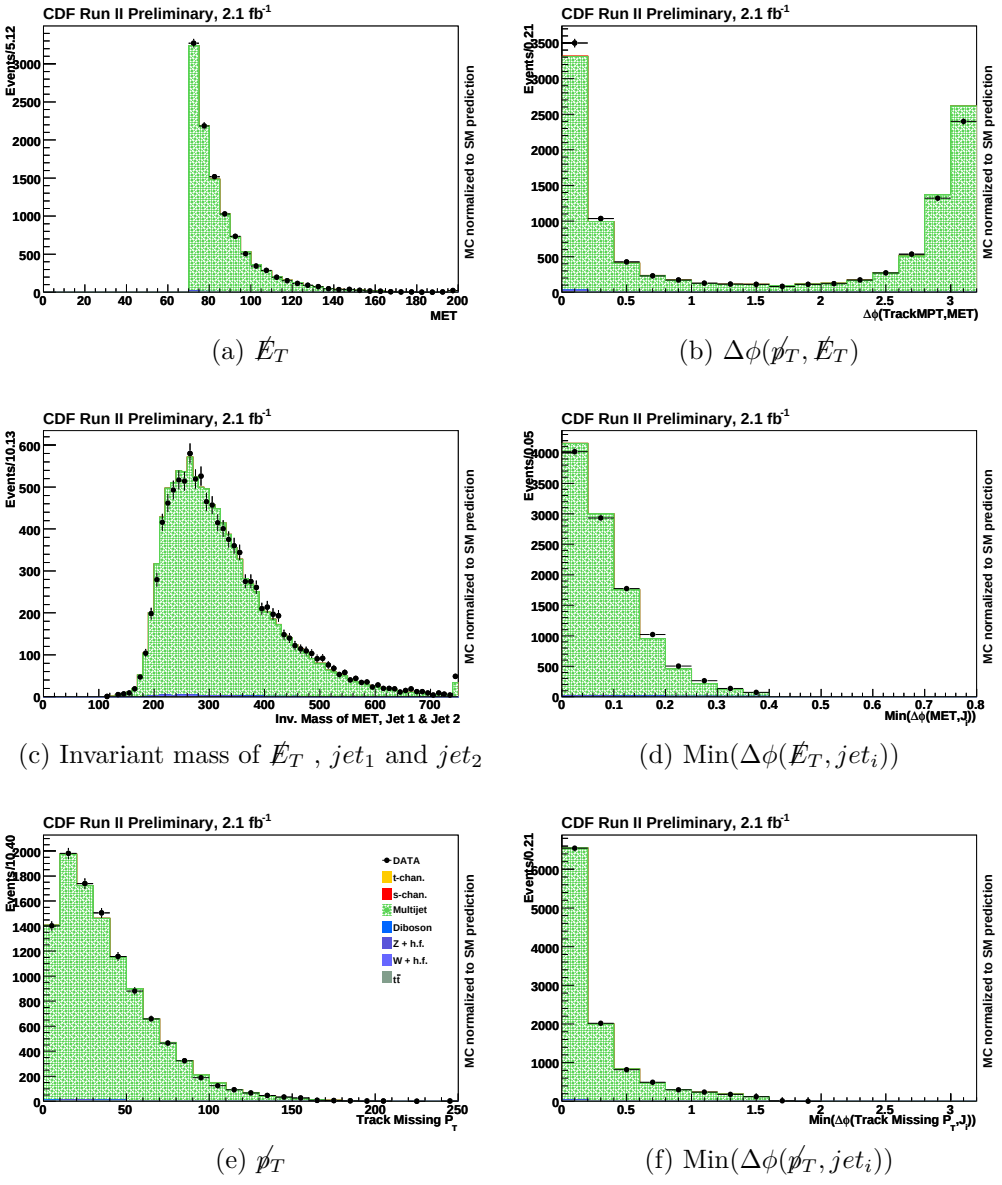
Table D.7 shows the event yields in the QCD control region for the three tagging categories. The expectations are normalized to data by scaling the QCD multijet accordingly.

Process	1b-tag	2 b-tags(SS)	2 b-tags(SJ)
Single Top S	0.6 ± 0.1	0.23 ± 0.04	0.27 ± 0.04
Single Top T	0.5 ± 0.1	0.02 ± 0.01	0.03 ± 0.01
Top Pair	7.0 ± 0.5	1.6 ± 0.2	1.9 ± 0.3
Di-Boson	0.9 ± 0.5	0 ± 0	0.1 ± 0.1
W + h.f.	20.5 ± 21.6	0.5 ± 0.3	1.8 ± 1.2
Z + h.f	26.8 ± 22.4	3.0 ± 1.4	3.9 ± 2.0
QCD Multijet	8751 ± 141	597 ± 32	1308 ± 55
Exp. Signal	1.1 ± 0.1	0.26 ± 0.04	0.31 ± 0.04
Exp. Background	8806 ± 148	602 ± 32	1316 ± 55
Total Expected	8807 ± 148	602 ± 32	1316 ± 55
DATA	8807	602	1316

Table D.7: Number of expected and observed events in the QCD control region for the three tagging categories. The expectations are normalized to data by scaling the QCD multijet accordingly.

D.3.2 The Electroweak Control Region

To test the modeling of processes with real $\not{E}_T$, such as W or Z plus heavy flavor jets, top pair production, diboson production and single top production. Since we want to remain unbiased to our signal region, we test those in events with at least one lepton (all leptons are vetoed in the signal region). This region is used to check the shapes of the Monte Carlo predictions and also serves as a low statistics check for the QCD multijet model. This control region, however, should only be considered able to

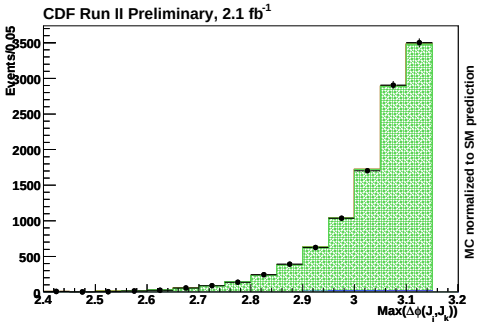


(a) $\not\!\!E_T$ (b) $\Delta\phi(\not\!\!p_T, \not\!\!E_T)$

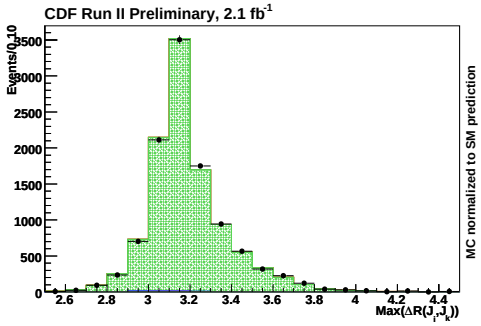
(c) Invariant mass of $\not\!\!E_T$, jet_1 and jet_2 (d) Min($\Delta\phi(\not\!\!E_T, jet_i)$)

(e) $\not\!\!p_T$ (f) Min($\Delta\phi(\not\!\!p_T, jet_i)$)

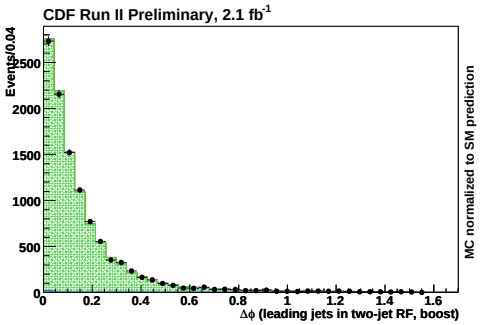
Figure A8.: Comparison of our background modeling to data in the QCD control region for the NN_{QCD} input variables in events with at least one b-tag. The variables show good agreement.



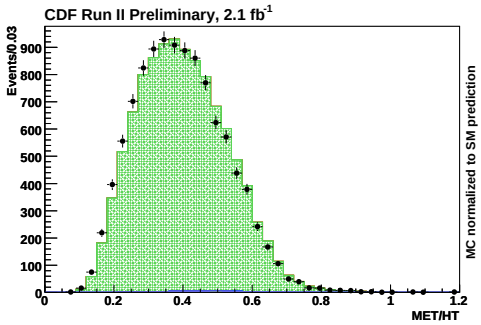
(a) Maximum of $\Delta\phi$ between two jets



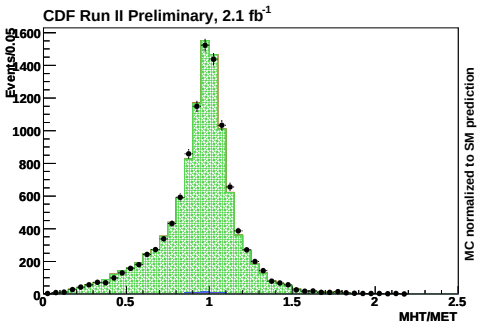
(b) Maximum of ΔR between two jets



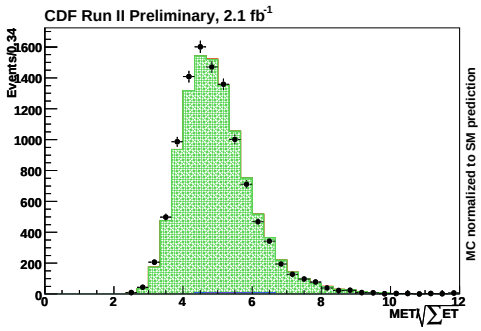
(c) $\Delta\phi(j_1, j_2)$ in the 2-jet rest frame



(d) $\not\!\!E_T/H_T$)

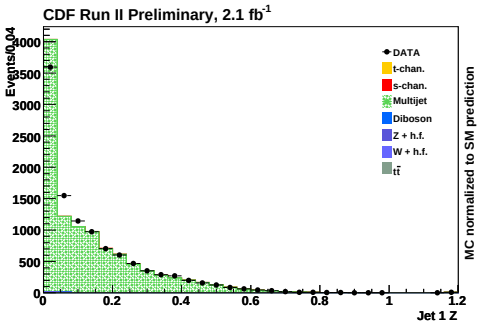


(e) $\not\!\!H_T$ / $\not\!\!E_T$

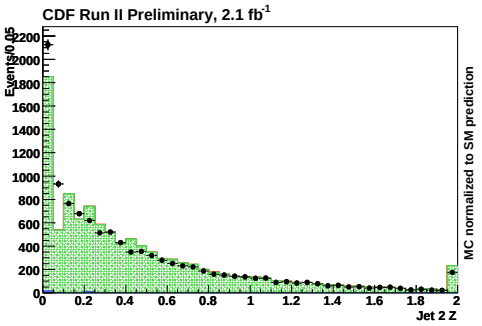


(f) $\not\!\!E_T/\sqrt{\sum E_T}$

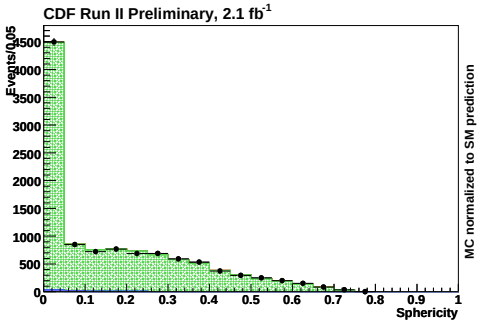
Figure A9.: Comparison of our background modeling to data in the QCD control region for the NN_{QCD} input variables in events with at least one b-tag. The variables show good agreement.



(a) $\sum P_T^{chgd}/P_T^{jet}$ for the leading jet



(b) $\sum P_T^{chgd}/P_T^{jet}$ for the 2nd leading jet



(c) Sphericity

Figure A10.: Comparison of our background modeling to data in the QCD control region for the NN_{QCD} input variables in events with at least one b-tag. The variables show good agreement.

reveal a significant mis-modeling in the larger backgrounds, for it not only contains single top events but these are actually looked at in the analyses in the leptons plus jets sample. We briefly mention it here since it was used to derive the parts common to $\not\!\!E_T$ plus jets Higgs analysis and the one being presented here.

Figures A11 to A13 show the validation against data of the kinematic distribution used as input to the NN_{QCD} in the electroweak control region and for the three tagging categories.

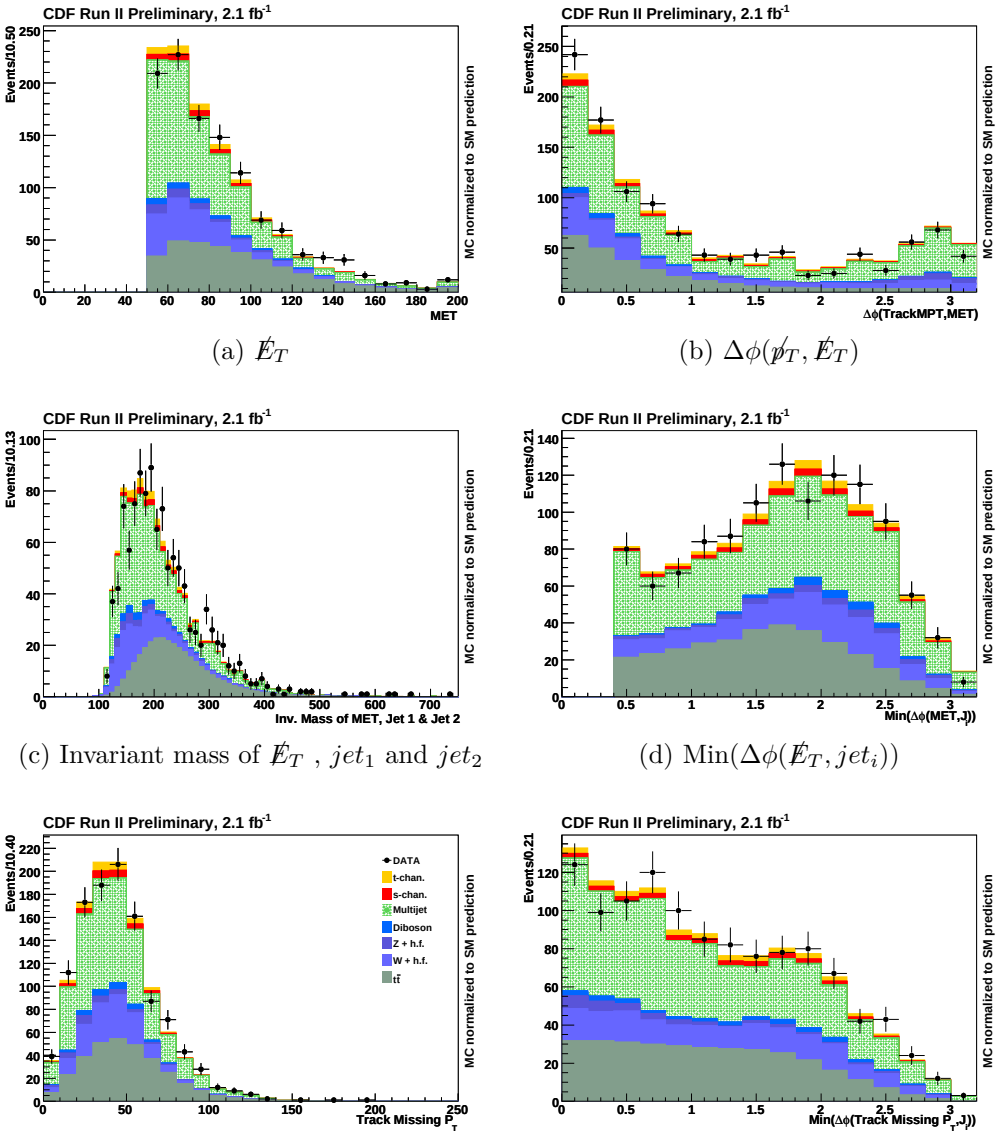
Table D.8 shows the event yields in the electroweak control region for the three tagging categories. The expectations are normalized to data by scaling the QCD multijet accordingly.

Process	1b-tag	2 b-tags(SS)	2 b-tags(SJ)
Single Top S	15.8 ± 0.3	7.6 ± 0.2	6.4 ± 0.2
Single Top T	31.9 ± 4.6	1.7 ± 0.1	1.4 ± 0.1
Top Pair	204.9 ± 25.7	61.8 ± 12.2	56 ± 12.1
Di-Boson	24.7 ± 6.4	2.4 ± 0.4	2.4 ± 0.5
W + h.f.	148.0 ± 57.9	10.1 ± 2.7	11.6 ± 3.1
Z + h.f	27.7 ± 13.8	1.7 ± 0.9	2.8 ± 1.3
QCD Multijet	455 ± 32	21.6 ± 9.5	44.5 ± 13.6
Exp. Signal	47.7 ± 4.6	9.3 ± 0.2	7.9 ± 0.2
Exp. Background	860 ± 83	98 ± 16	117 ± 19
Total Expected	908 ± 83	107 ± 16	125 ± 19
DATA	908	107	125

Table D.8: Number of expected and observed events in the electroweak control region for the three tagging categories. The expectations are normalized to data by scaling the QCD multijet accordingly.

D.3.3 The $NN_{QCD} < -0.1$ Control Region

This region is peculiar in that it is not *stricto sensu* only a control region. While it serves the purpose of checking the shapes of our data-driven model, it is also the region where the normalization of the data-driven model is performed. This region

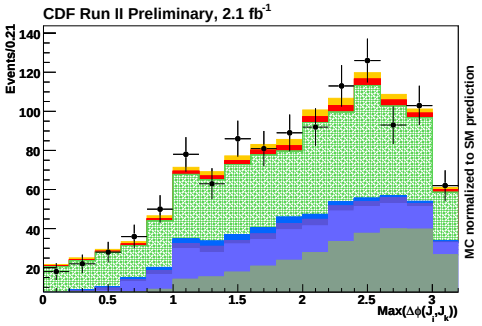


(a) $\not\!\!E_T$ (b) $\Delta\phi(\not\!\!p_T, \not\!\!E_T)$

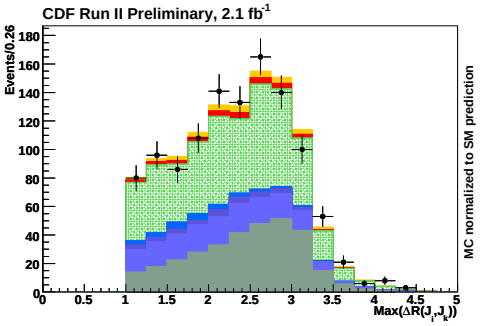
(c) Invariant mass of $\not\!\!E_T$, jet_1 and jet_2 (d) Min($\Delta\phi(\not\!\!E_T, jet_i)$)

(e) $\not\!\!p_T$ (f) Min($\Delta\phi(\not\!\!p_T, jet_i)$)

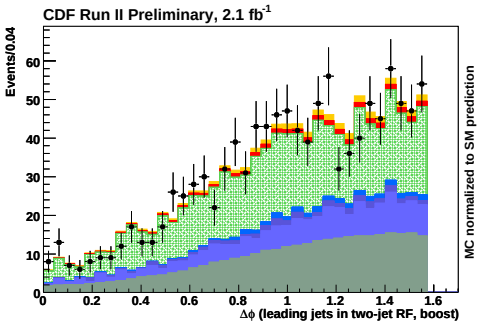
Figure A11.: Comparison of our background modeling to data in the electroweak control region for the NN_{QCD} input variables in events with at least one b-tag. The variables show good agreement.



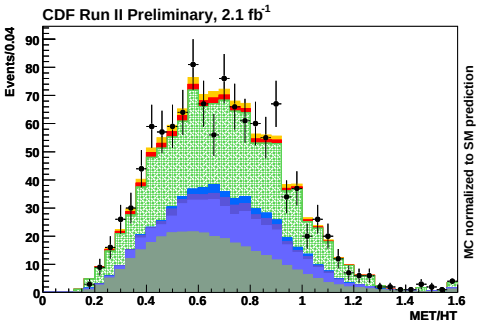
(a) Maximum of $\Delta\phi$ between two jets



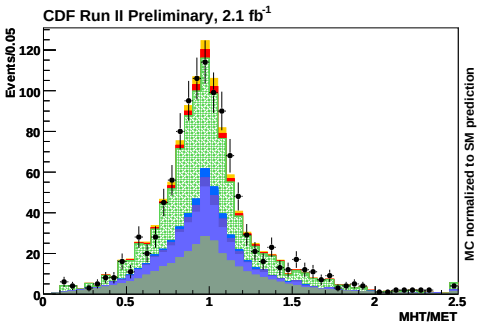
(b) Maximum of ΔR between two jets



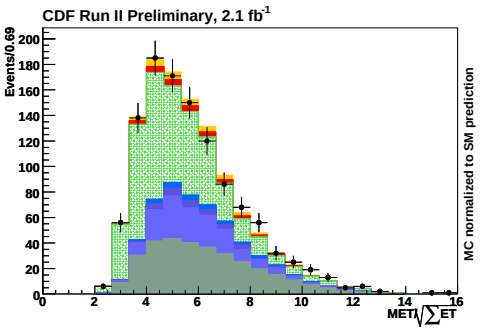
(c) $\Delta\phi(j_1, j_2)$ in the 2-jet rest frame



(d) $\not\!\!E_T/H_T$)

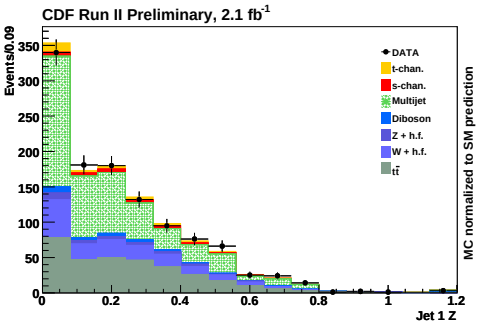


(e) $\not\!\!H_T$ / $\not\!\!E_T$

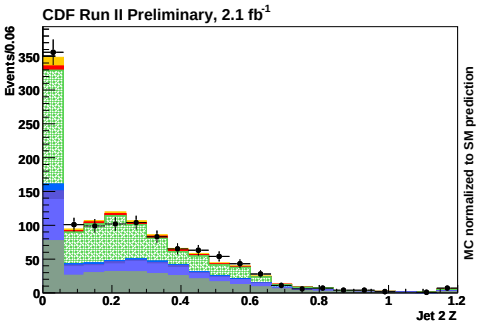


(f) $\not\!\!E_T/\sqrt{\sum E_T}$

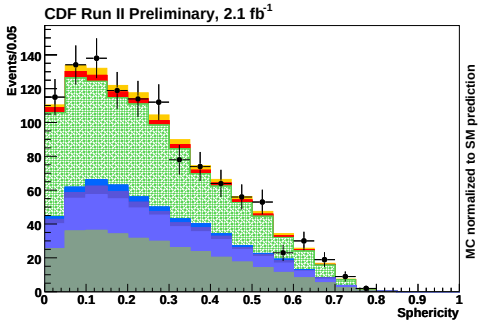
Figure A12.: Comparison of our background modeling to data in the electroweak control region for the NN_{QCD} input variables in events with at least one b-tag. The variables show good agreement.



(a) $\sum P_T^{chgd}/P_T^{jet}$ for the leading jet



(b) $\sum P_T^{chgd}/P_T^{jet}$ for the 2nd leading jet



(c) Sphericity

Figure A13.: Comparison of our background modeling to data in the electroweak control region for the NN_{QCD} input variables in events with at least one b-tag. The variables show good agreement.

is built from the pre-selection cuts and the requirement of $NN_{QCD} < -0.1$, i.e. it is the region of the pre-selection not part of the signal region.

Figures A14 to A16 show the validation against data of the kinematic distribution used as input to the NN_{QCD} in the $NN_{QCD} < -0.1$ control region and for the three tagging categories.

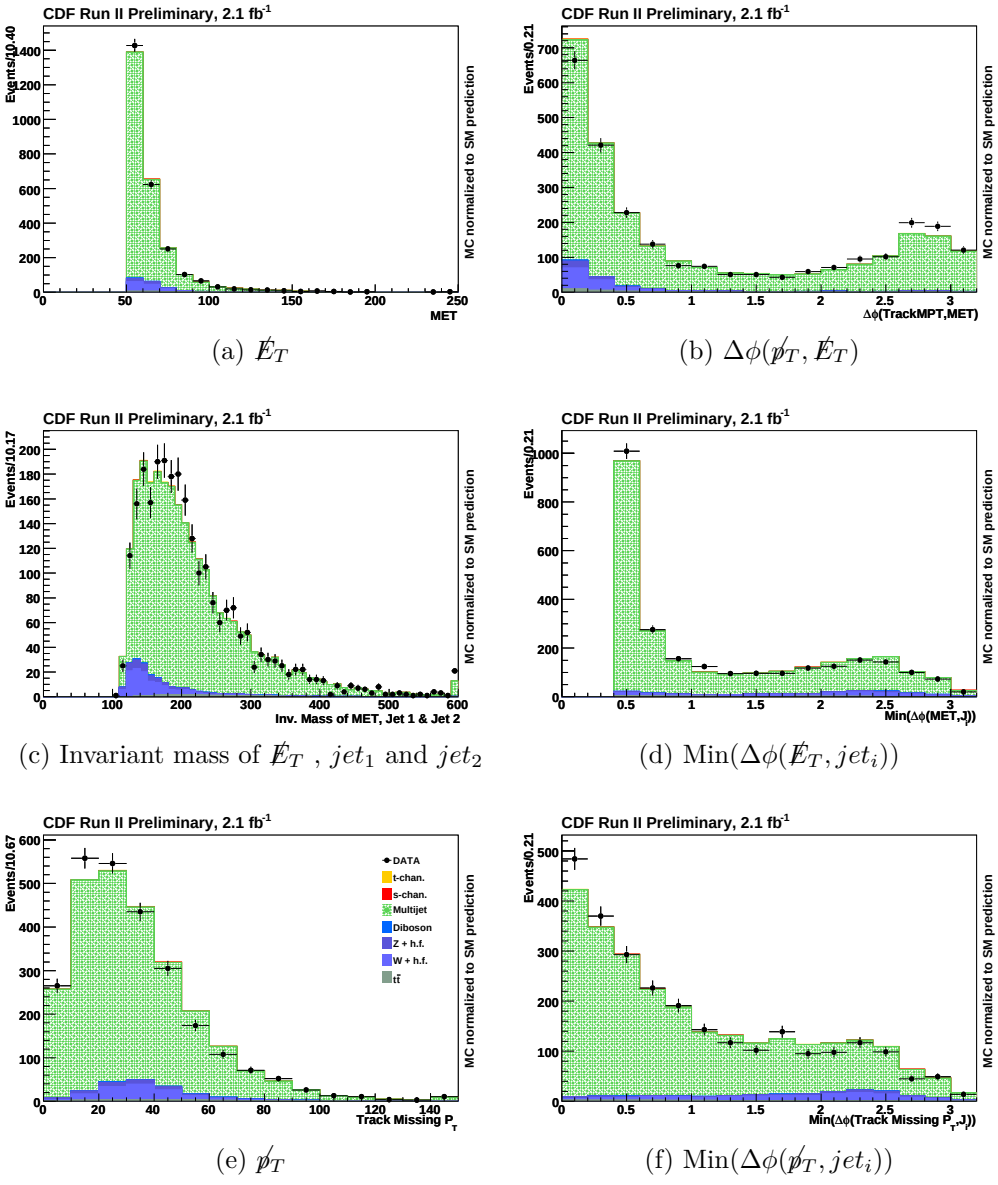
Table D.9 shows the event yields in the $NN_{QCD} < -0.1$ control region for the three tagging categories. The expectations are normalized to data by scaling the QCD multijet accordingly. This scaling is used in the signal region to derive the expected event yields for QCD multijet production.

Process	1b-tag	2 b-tags(SS)	2 b-tags(SJ)
Single Top S	2.1 ± 0.1	0.36 ± 0.04	0.48 ± 0.05
Single Top T	3.8 ± 0.2	0.09 ± 0.02	0.17 ± 0.03
Top Pair	21.7 ± 0.8	2.0 ± 0.2	2.7 ± 0.3
Di-Boson	6.7 ± 1.1	0.3 ± 0.2	0.4 ± 0.2
W + h.f.	119.0 ± 52.2	0.7 ± 0.6	2.8 ± 1.3
Z + h.f	40.6 ± 13.8	1.3 ± 0.6	2.1 ± 0.8
QCD Multijet	2149 ± 97	68.2 ± 8.9	157.4 ± 15.7
Exp. Signal	5.9 ± 0.2	0.46 ± 0.05	0.65 ± 0.06
Exp. Background	2337 ± 117	72.5 ± 9.0	165.4 ± 15.8
Total Expected	2343 ± 117	73 ± 9	166 ± 16
DATA	2343	73	166

Table D.9: Number of expected and observed events in the $NN_{QCD} < -0.1$ control region for the three tagging categories. The expectations are normalized to data by scaling the QCD multijet accordingly. This scaling is used in the signal region to derive the expected event yields for QCD multijet production.

D.4 Multivariate discrimination

After having developed an event selection which maximizes the signal purity of the sample, we need to determine the systematic uncertainty on the background prediction for events passing our selection requirements. The background systematic uncertainty is approximately 2 to 4 times the size of the signal we seek, depending

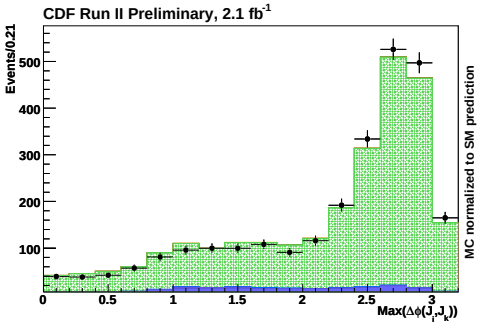


(a) $\not\!\!E_T$ (b) $\Delta\phi(\not\!\!p_T, \not\!\!E_T)$

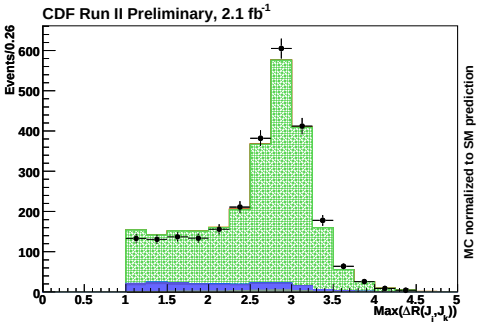
(c) Invariant mass of $\not\!\!E_T$, jet_1 and jet_2 (d) Min$(\Delta\phi(\not\!\!E_T, jet_i))$

(e) $\not\!\!p_T$ (f) Min$(\Delta\phi(\not\!\!p_T, jet_i))$

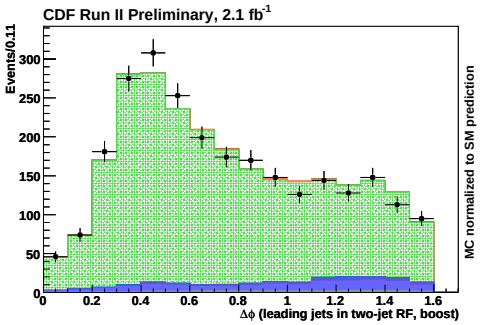
Figure A14.: Comparison of our background modeling to data in the $NN_{QCD} < -0.1$ control region for the NN_{QCD} input variables in events with at least one b-tag. The variables show good agreement.



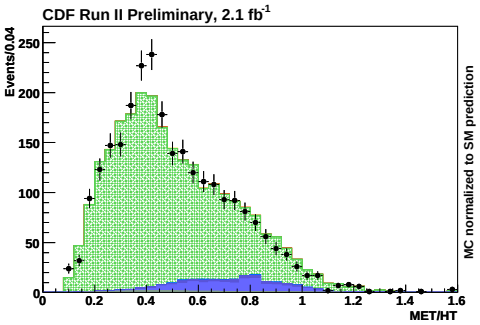
(a) Maximum of $\Delta\phi$ between two jets



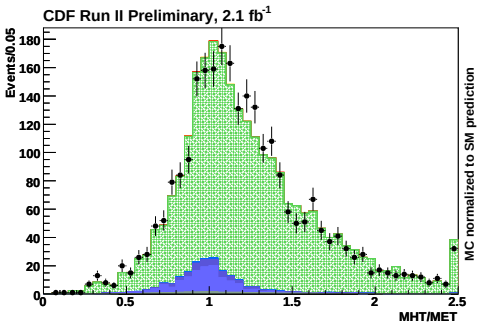
(b) Maximum of ΔR between two jets



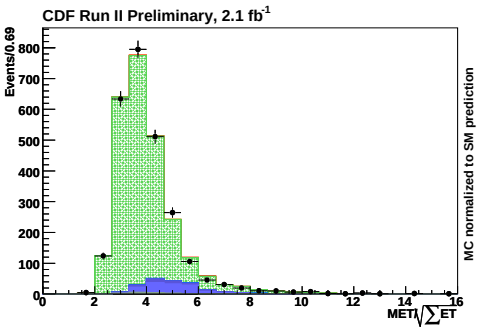
(c) $\Delta\phi(j_1, j_2)$ in the 2-jet rest frame



(d) $\not\!\!E_T/H_T$)

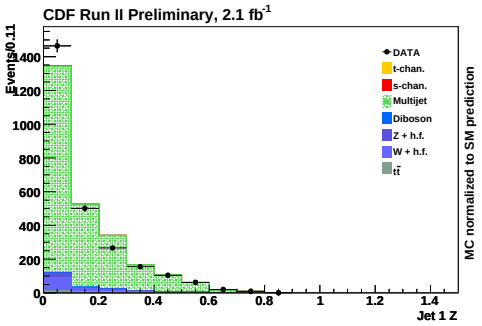


(e) $\not\!\!H_T$ / $\not\!\!E_T$

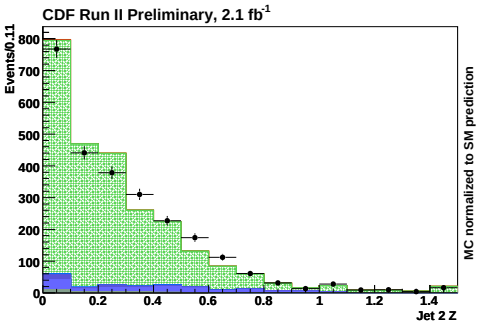


(f) $\not\!\!E_T/\sqrt{\sum E_T}$

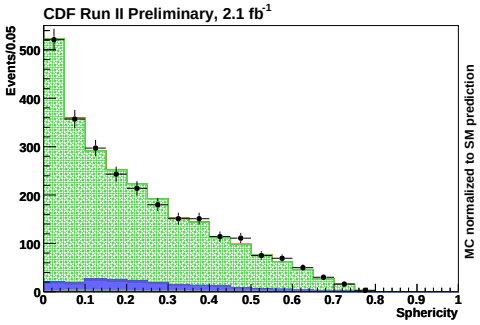
Figure A15.: Comparison of our background modeling to data in the $NN_{QCD} < -0.1$ control region for the NN_{QCD} input variables in events with at least one b-tag. The variables show good agreement.



(a) $\sum P_T^{chgd}/P_T^{jet}$ for the leading jet



(b) $\sum P_T^{chgd}/P_T^{jet}$ for the 2nd leading jet



(c) Sphericity

Figure A16.: Comparison of our background modeling to data in the $NN_{QCD} < -0.1$ control region for the NN_{QCD} input variables in events with at least one b-tag. The variables show good agreement.

on the channel. Simply counting events which pass our event selection will not test for the presence or absence of single top production if the systematic uncertainty on the background remains large. Further separation of the signal from the background is required, based on the differences in kinematics expected between signal and background events. Events which appear to be more signal-like are used to test for the presence of single top production and to measure the cross-section, and events which appear to be more background-like are used to constrain the uncertain background rates.

D.4.1 A NN to discriminate the signal from the backgrounds

At this stage of the analysis, the S/B ratio is about 1/20. As show in previous section, the background which is the most different from the signal we look after is the QCD instrumental background, which has already been greatly reduced. Unfortunately, all surviving backgrounds have topology and kinematics very similar to single top events since charged and neutral leptons are undetected and the final state kinematics are largely unconstrained. In order to increase statistical power of the analysis, and to minimize the effect of the background systematic uncertainties, we study the signal sample to take advantage of the small residual differences between the signal and backgrounds.

We use the following variables to discriminate between signal and background processes: the invariant mass of the second jet and missing transverse energy; the second jet in the $W \to \tau\nu$ events is often a τ, thus this variable is the reconstructed W transverse mass for this background events; the scalar sum of transverse energy of all the jets in the event, H_T^3; the minimum of the difference in ϕ between the $\not\!\!E_T$ and each jet j_i, considering all two or three $(\not\!\!E_T, j_i)$ pairings; $\mathrm{Z}(j_{1(2)})$, the ratio of the sum the jet track p_Ts to the $P_T(j_{1(2)})$; the absolute amount of the missing transverse energy, $\not\!\!E_T$; the absolute amount of the missing transverse momentum, $\not\!\!p_T$; the $\Delta\phi$ between the direction of the leading jets in the two-jet rest frame and the direction

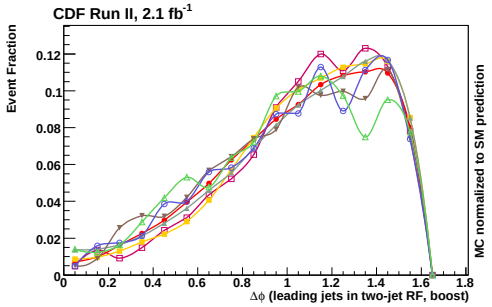
of the boost; the $\not\!\!E_T/H_T$; the invariant mass of $\not\!\!E_T$, j_1 and j_2; the invariant mass of all tight jets in the event, m_{3j}.

Using the above inputs, we train a multi-layer-perceptron neural network. We use the single top s- and t-channels in their respective proportions in the signal sample (approximately 50%-50%). For the training purpose, we select the background processes which account for more that 5% of the total background, i.e. the multijet, top pair production, $W \to \tau\nu$ and $Z \to \nu\nu$ processes. Both training and test samples contain 39,000 signal and 42,000 background events. The network architecture consists of an input layer with 11 nodes corresponding to the input variables shown in table D.10, plus 1 bias node; 2 hidden layers with 22 and 11 hidden nodes, and an output layer with the output node. We looked at the convergence plots and found no overtraining.

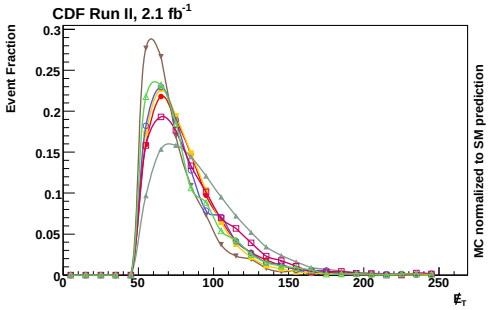
Table D.10: Input variables to the final discriminant neural network.

Variable
Invariant mass of $\not\!\!E_T$ and 2nd jet
Scalar sum of jet p_T's, H_T
Minimum of $\Delta\phi$ between the $\not\!\!E_T$ and each jet ji
$\sum P_T^{chgd}/P_T^{jet}$ for the leading jet
$\sum P_T^{chgd}/P_T^{jet}$ for the 2nd leading jet
Absolute amount of the $\not\!\!E_T$
Absolute amount of the $\not\!\!p_T$
$\Delta\phi(j_1, j_2)$ in the 2-jet rest frame
$\not\!\!H_T$ / $\not\!\!E_T$
Invariant mass of $\not\!\!E_T$, j1 and j2
Invariant mass of all jets in the event

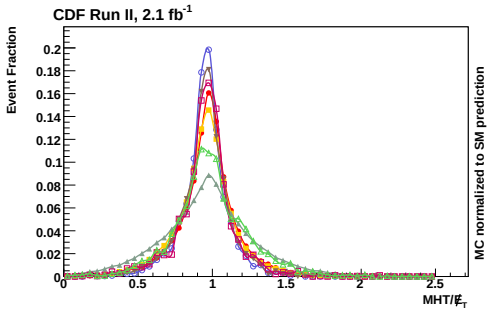
The distributions of the input variables are shown in the pre-selection region for all tagging categories in figure A17 and A18, where the shapes of the distributions for each group of physics processes are compared. In figure A19, the same variables are shown but validated against data. One can see that the predictions agree well with data. Validation against data was also performed in the control regions, even before



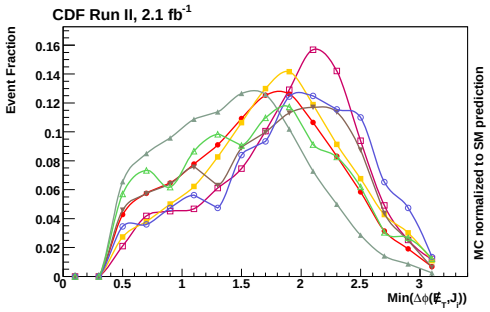
(a) $\Delta\phi(jet_1, jet_2)$ with respect to the dijet system direction



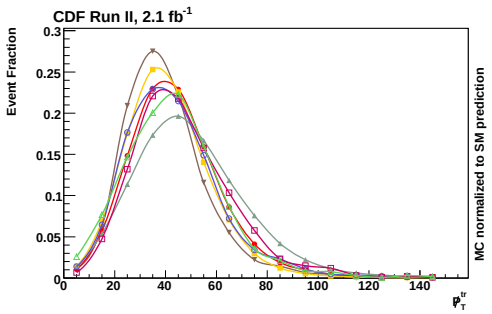
(b) $\not\!\!E_T$



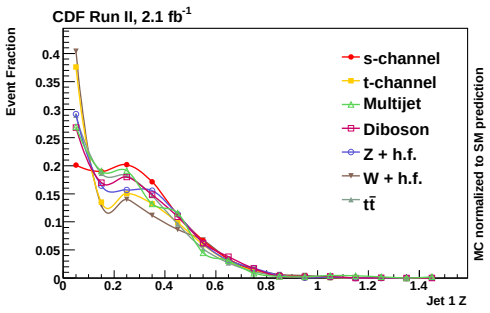
(c) $\not\!\!H_T/\not\!\!E_T$



(d) Min($\Delta\phi(\not\!\!E_T, jet_i)$)

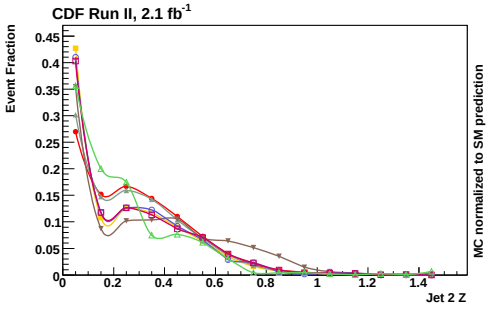


(e) $\not\!\!p_T$

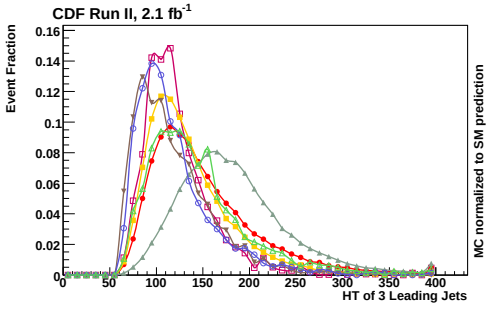


(f) $\sum P_T^{chgd}/P_T^{jet}$ for the leading jet

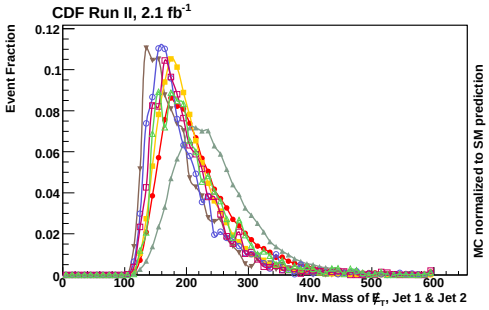
Figure A17.: Kinematic distributions for the single top s- and t-channel Monte Carlo (full circle and full square, respectively) events, and background events, for events in the signal region ($NN_{QCD} > -0.1$). All histograms are normalized to unity.



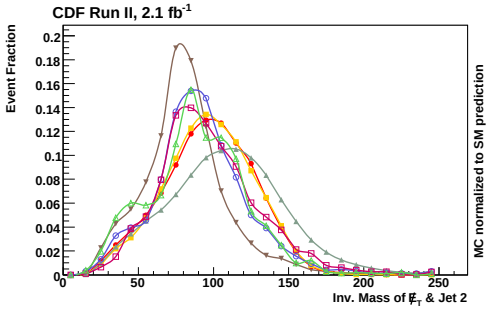
(a) $\sum P_T^{chgd}/P_T^{jet}$ for the 2nd leading jet



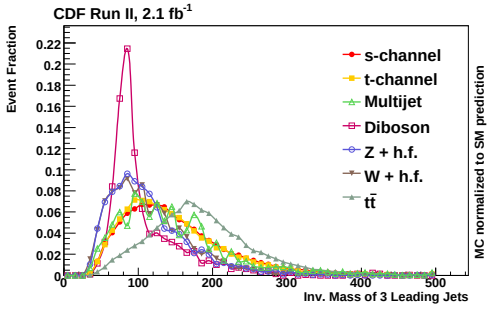
(b) Scalar sum of jet p_T's, H_T



(c) Invariant mass of $\not\!\!E_T$, jet_1 and jet_2



(d) Invariant mass of $\not\!\!E_T$ and 2nd jet

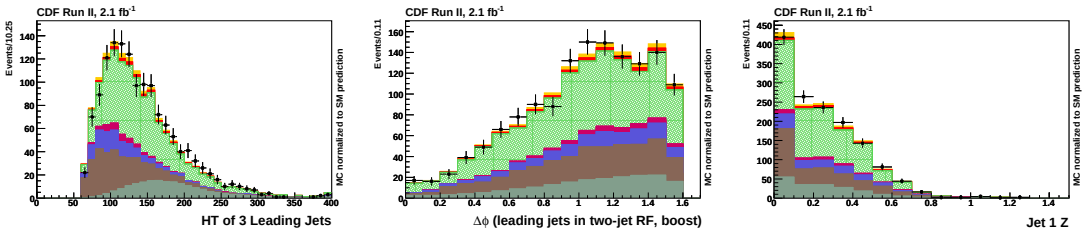


(e) Invariant mass of all jets in the event

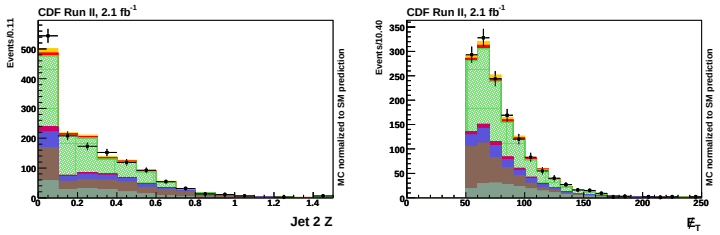
Figure A18.: Kinematic distributions for the single top s- and t-channel Monte Carlo (full circle and full square, respectively) events, and background events, for events in the signal region ($NN_{QCD} > -0.1$). All histograms are normalized to unity.

looking at the signal region plots. The output of this NN, which we call the final NN discriminant, is shown in Fig. A20. This output as well as the NN_{QCD} output was also validated against data in the control regions.

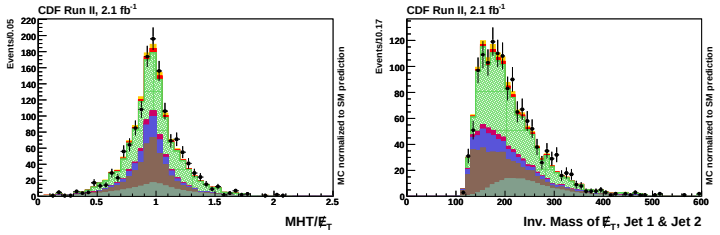
The output of the final NN discriminant peaks around 0.3 for the signal and -0.3 for the backgrounds. Although we use many variables, separation is hard to achieve because of the lack of reconstructed properties of the top quark. The separation is helpful in the measurement of the cross-section and the determination of the significance because backgrounds with large systematics are moved away from the bins where the signal peaks.



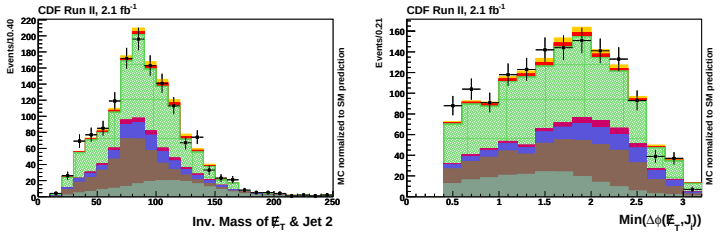
(a) Scalar sum of jet p_T's, H_T (b) $\Delta\phi(jet_1, jet_2)$ with respect to the dijet system direction (c) $\sum P_T^{chgd}/P_T^{jet}$ for the leading jet



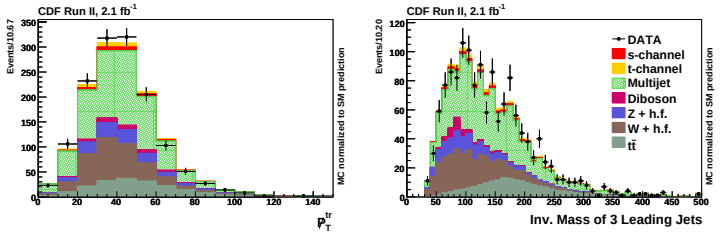
(d) $\sum P_T^{chgd}/P_T^{jet}$ for the 2nd leading jet (e) $\not\!\!E_T$



(f) $\not\!\!H_T/\not\!\!E_T$ (g) Invariant mass of $\not\!\!E_T$, jet_1 and jet_2

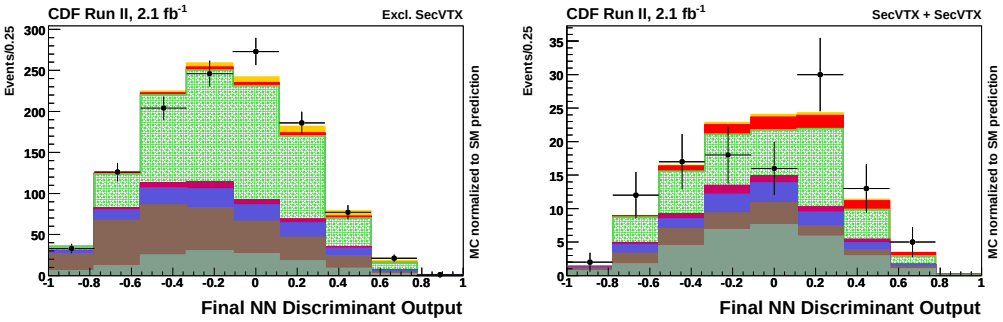


(h) Invariant mass of $\not\!\!E_T$ and 2nd jet (i) Min$(\Delta\phi(\not\!\!E_T, jet_i))$



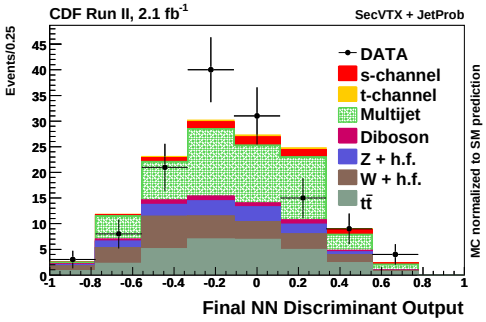
(j) $\not\!\!p_T$ (k) Invariant mass of all jets in the event

Figure A19.: NN discriminant input variables distribution in signal region ($NN_{QCD} > -0.1$). For clarity, all the three flavor categories have been merged.



(a) Events with exactly one SecVtx-tagged jet

(b) Events with two SecVtx-tagged jets



(c) Events with one SecVtx-tagged and one JetProb-tagged jet

Figure A20.: Final NN discriminant output distributions in signal region.

D.5 Systematic Uncertainties

To address the fact that our predictions for various observables differ from their "true" values, we assign systematic uncertainties and propagate their effects on the measured cross-sections and on the significance of the signal.

Systematic uncertainties are grouped by their source. A given source of uncertainty may affect several templates (background and/or signal). The various systematic uncertainties related to a same source are considered 100% correlated. There are three categories of systematic uncertainties: (a) rate uncertainties, which are related to the predicted production rates of the various signal and background processes; (b) shape uncertainties, which express differences in the distributions due to a given systematic source, and (c) bin-by-bin uncertainties, which are statistical fluctuations arising from the limited size of the Monte Carlo or data samples (where applicable).

The sources of systematic uncertainties are presented in chapter 13. One source of systematic uncertainty is added to those described there:

- **Top mass dependence** The uncertainty associated with the variation of the top mass. We consider top masses of 170 and 180 GeV. This uncertainty is applied to all top processes only when extracting the value of V_{tb} and computing the p-value (defined in the next section).

For the analysis in this chapter, these are summarized in table D.11.

D.6 Cross-section Measurement

When quoting the single top cross-section, we assume a top quark mass $m_t = 175\text{GeV}/c^2$. We do not include the m_t uncertainty in the $t\bar{t}$ background or the signal when measuring the cross-section or its uncertainty. When extracting the cross-sections, the theoretical uncertainty on σ_{s+t} is also not included.

Systematic	Rate	Shape	Comment
Top pair cross-section	± 12%	-	
W/Z + h.f. cross-section	± 40%	-	
Diboson cross-section	± 11%	-	
Luminosity	6%	-	Not for QCD multijet
Trigger Efficiency	$< 2.6\%$	X	
B tagging scale factors	4.5% to 12%	-	
Lepton Veto	2%	-	
ISR/FSR	-4.5% …+16%	X	Only for top processes
JES	-14% …+23%	X	
PDF	$\pm 1\% \ldots \pm 2\%$	X	Shape only for single top
QCD multijet model	4.5% to 13%	X	Only QCD multijet
Background scaling	2%	-	
Signal cross-section	± 12%	-	Only for p-value and V_{tb}
Top mass dependence	-16% …+7.5%	X	

Table D.11: Summary of systematic uncertainties

Measurement of σ_{s+t}

We assume the standard model ratio between s- and t-channel production. We make use of a Bayesian marginalization technique [38] to incorporate the effects of systematic uncertainties.

$$L'(\beta) = \int L(\mathbf{D}|\beta, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{s}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\kappa}, \boldsymbol{\delta})\pi(\boldsymbol{\theta})\pi(\boldsymbol{\eta})d\boldsymbol{\theta}d\boldsymbol{\eta}, \tag{D.3}$$

where the π functions are the Bayesian priors assigned to each nuisance parameter. These priors are explained in Section D.5. The L' notation is condensed and not all parameters are present. The measured cross-section corresponds to the maximum of L', which occurs at $\beta^{\max}$:

$$\sigma_{s+t}^{\text{meas}} = \sigma_{s+t}^{\text{SM}}\beta^{\max}. \tag{D.4}$$

The asymmetric uncertainties quoted correspond to the shortest interval containing 68% of the integral of the posterior, assuming a uniform prior in β, $\pi(\beta) = 1$:

$$0.68 = \frac{\int_{\beta_{\text{low}}}^{\beta_{\text{high}}} L'(\beta)\pi(\beta)d\beta}{\int_0^{\infty} L'(\beta)\pi(\beta)d\beta}.$$

Extraction of Bounds on $|V_{tb}|$

Under the standard model,

$$|V_{tb}|^2 = \frac{\sigma_{s+t}^{\text{meas}}}{\sigma_{s+t}^{\text{SM}}} \tag{D.5}$$

with the assumption that $|V_{td}|^2 + |V_{ts}|^2 \ll |V_{tb}|^2$, and that only $|V_{tb}|$ incorporates new physics contributions. Hence, the theoretical uncertainty on σ_{s+t}^{SM} needs to be taken into account when setting the V_{tb} constraints. This is not the case when measuring the cross-section.

D.7 Results

By using the signal and background modeling described in section 8.5, scanning the multivariate discriminant described in section D.4.1 using the likelihood described in chapter 12, we expect to measure the single top cross-section as

$$\sigma_{s+t}^{exp} = 2.7^{+2.3}_{-2.1} \text{ pb}.$$

Applying our analysis to the first 2.1 fb^{-1} of data recorded by the CDF II experiment, we measure a single top production cross-section of

$$\sigma_{s+t}^{obs} = 4.9^{+2.5}_{-2.2} \text{ pb}.$$

i.e. roughly with a 50% precision. The measured value is thus consistent with a $+1\sigma$ statistical upward fluctuation with respect to the standard model cross-section. The probability to measure a cross-section this high or higher has been estimated to be

18% by running 150,000 pseudoexperiments. The distribution of the outcomes of the pseudoexperiments is shown in figure A21.

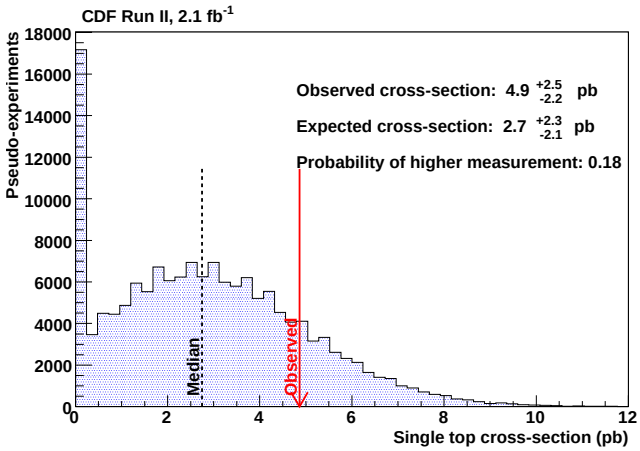


Figure A21.: Distribution of cross-section measurements outcome using pseudoexperiments. The arrow shows the observed cross-section. The probability to measure a cross-section this high or higher amounts to 18%.

We also perform the measurement by splitting the samples according to the heavy flavor content and the tagging algorithm used to estimate it. The results are shown in figure A22. While the precision is low, the three orthogonal measurements are in agreement with each other.

Using the statistical test described above, we compute the probablility that the background model (B) fluctuated equal or up to the observed value in the data (observed p-value) or to the median of signal plus background (S+B) pseudo-experiments (expected p-value). We obtain an expected p-value of 7.9×10^{-2} (1.4σ) and an observed p-value of 1.6×10^{-2} (2.1σ). The results are shown in figure A23.

Finally, we measure the V_{tb} element of the CKM matrix. Using an unconstrained flat prior, we find $|V_{tb}| = 1.24^{+0.34}_{-0.29} \pm 0.07$(theory) as shown in figure A24; the theoretical uncertainty is entirely due to the uncertainty on the standard model theoretical cross-section for single top production [157, 160]. Assuming a flat prior between 0 and 1, figure A24b shows that $V_{tb} > 0.36$ at 95% confidence level.

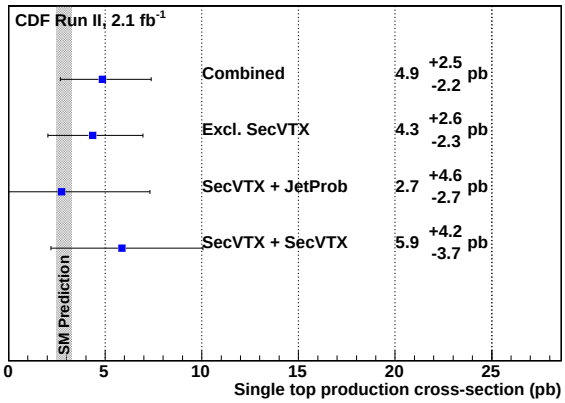


Figure A22.: Measurement of the single top cross-section production. We show the result in each tagging category, as well as the combination of all the available channels. All measurements are consistent with the standard model theoretical cross-section within uncertainties.

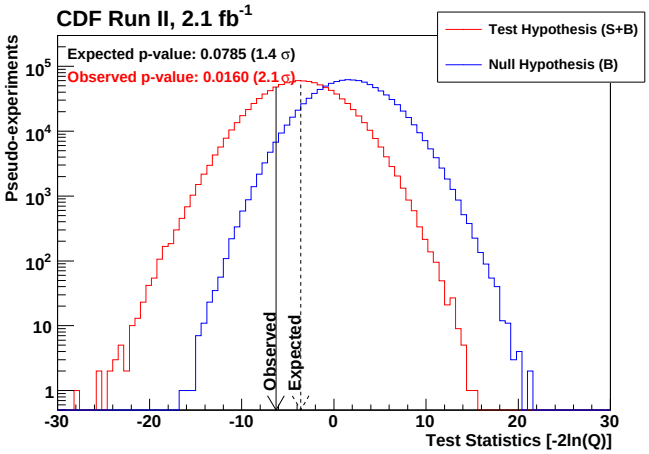
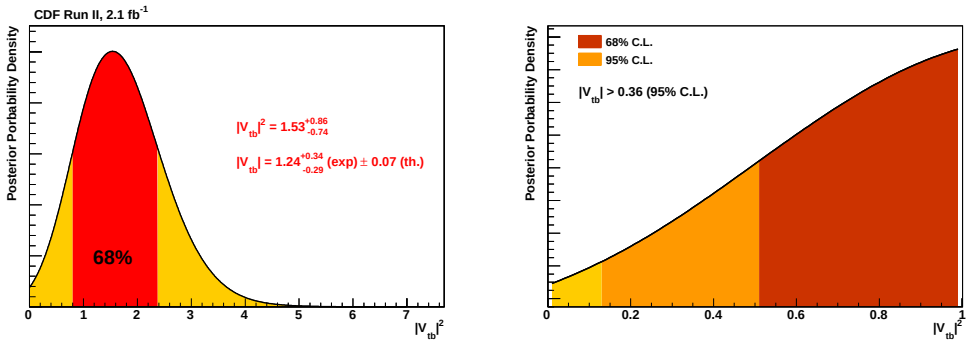


Figure A23.: Test statistics for expected cross-section measurement.

D.8 Summary

We have presented a search for s- and t-channel electroweak single top production in the $\not\!\!E_T$+jets channel. We have analyzed 2.1 fb^{-1} of CDF Run II data and measured the single top production cross-section for the first time in events where the lepton



(a) With unconstrained flat prior (b) With flat prior between 0 and 1

Figure A24.: Measurement of the V_{tb} element of the CKM matrix.

from the W decay is either not identified or reconstructed as a jet. We find, assuming a top quark mass of 175 GeV/c^2,

$$\sigma_{s+t}^{obs} = 4.9^{+2.5}_{-2.2} \text{ pb}.$$

The observed p-value is 1.6×10^{-2} (2.1 σ). While the sensitivity obtained with this analysis is much lower than that of the single top searches requiring the presence of identified leptons, this analysis played an important role in the combination. It allowed the extension of the acceptance by 33% using an orthogonal sample and improved the combined sensitivity to reach the 5σ observation of the single top production.

Finally, we measure the V_{tb} element of the CKM matrix :

$$|V_{tb}| = 1.24^{+0.34}_{-0.29} \pm 0.07(\text{theory}).$$

Assuming a flat prior between 0 and 1, we proceed to set the lower limit of $V_{tb} > 0.36$ at 95% confidence level.

[redacted]

Karolos Potamianos [redacted] He completed his high school education in 2002, and subsequently attended the Polytechnic School of the University of Brussels (ULB), from which he obtained a M.Sc. in Applied Physics Engineering in 2007. His master thesis topic was on quantum transport in semiconductors. During his time at ULB he also obtained a B.Sc. in Business Administration from the Solvay Business School (2003-2006). In the Fall of 2007, he entered the graduate program at Purdue University. He joined the CDF Collaboration at Fermilab in January 2008 under the supervision of Daniela Bortoletto. He was a URA Visiting Scholar at Fermilab during the Fall of 2009. He completed his Ph.D. in November 2011.