SEARCH FOR THE HIGGS BOSON DECAYING TO TWO TAU LEPTONS IN PROTON-ANTIPROTON COLLISIONS AT A CENTER OF MASS ENERGY OF 1.96 TEV

A Dissertation

by

ANDREY LVOVICH ELAGIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2011

Major Subject: Physics

SEARCH FOR THE HIGGS BOSON DECAYING TO TWO TAU LEPTONS IN PROTON-ANTIPROTON COLLISIONS AT A CENTER OF MASS ENERGY OF 1.96 TEV

A Dissertation

by

ANDREY LVOVICH ELAGIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Alexei Safonov
Committee Members,	Bhaskar Dutta
	Stephen Fulling
	Teruki Kamon
	Peter McIntyre
Head of Department,	Edward Fry

December 2011

Major Subject: Physics

ABSTRACT

Search for the Higgs Boson Decaying to Two Tau Leptons
in Proton-Antiproton Collisions
at a Center of Mass Energy of 1.96 TeV. (December 2011)

Andrey Lvovich Elagin, B.S.; M.S., Moscow Institute of Physics and Technology
Chair of Advisory Committee: Alexei Safonov

A search for the Higgs boson decaying to $\tau\tau$ using 7.8 fb^{-1} of $p\bar{p}$ collisions at 1.96 TeV collected with CDF II detector is presented. The search is sensitive to four production mechanisms of the Higgs boson: ggH, WH, ZH and VBF. Modes where one tau decay leptonically, and another decay, hadronically, are considered. Two novel techniques are developed and used in the search. A Probabilistic Particle Flow Algorithm is used for energy measurements of the hadronic tau candidates. The signal is discriminated from backgrounds by the Missing Mass Calculator, which allows for full invariant mass reconstruction of $\tau\tau$ pair. The data are found to be consistent with the background only hypothesis. Therefore a 95% confidence level upper limit on the Standard Model Higgs boson cross section was set. At M_H=120 GeV/c^2 observed limit is $14.9\times\sigma_{SM} \times Br(H \rightarrow \tau\tau)$.

ACKNOWLEDGMENTS

This work would be impossible without people at Fermilab who kept the Tevatron accelerator running and the CDF detector taking high quality data. The hard work of hundreds of people is behind this data. I would like to thank all of them.

My journey to the Texas A&M graduate school began at Snowmass conference in Summer 2005 where I met many enthusiastic physicists. My thesis advisor, Alexei Safonov, was one of them. Over the course of the social part of the conference I learned a lot about possible career paths available in physics. The decision to move to US was not easy. I am greatfull to Anna, my wife to be at that time, for her understanding and being on board with me on that difficult decision. I would like to thank Georgy Chelkov, my advisor at Dubna, for keeping our warm relationships after I left. I would like to thank my HARP-CDP colleagues, Friedrich Dydak, Alexey Guskov, Alexey Zhemchugov and Igor Boyko, for keeping me involved in the analysis of the HARP experiment data. A big part of my scientific experience would be missing today without that.

I was fortunate enough to meet so many great people at College Station, who helped me to adjust to a new lifestyle in US. I'm thankful to Slava Krutelyov who helped me to settle in town. Special thanks goes to Russell and Ryan for being such awesome roommates. I wouldn't see so much of Texas if Russell and Libby didn't take me on so many trips out of town. The trip to Sonora was particularly amazing. Parties organized by Ryan and Paulette in our house will be my best memories of College Station. It was also great to have Stephen as a new roommate later.

I would like to thank all TAMU professors for the great work they do in classes. Some of the courses were really tough (nothing, of course, compares to Peter McIntyre's Electrodynamics) but I am happy that I was able to enrich my knowledge in the

core courses. Special thanks goes to Glen Agnolet and Valery Pokrovsky for helpful discussions on their classes. I am also very glad that we have Dave Toback and Teruki Kamon on the Department. I am thankful to them for warm welcome to TAMU High Energy Group during my first semester and for keeping track of my progress later on. Our discussions about research and career prospects has been always insightful.

I would like to thank Sandi Smith for the enormous help with all kind of paperwork and administrative issues that international students face from time to time.

When I moved to Fermilab it was another new start. It was great to have Max Goncharov as a knowledgeable colleague and a good friend. It is very hard to list every person who made my life in Chicago area interesting and exciting. This list will never be complete. Thank to Sasha Paramonov, Vadim Khotilovich, Andrew Ivanov, Peter Svoisky, Yuri Oksuzian, Oscar Gonzalez, JJ Schmidt, Dee Hahn, and their friends, Fermilab feels like a second home. Many thanks to Barbara Alvarez for inviting us to play in the Fermilab Volleyball League. That helped me finally find a hobby I really enjoy. Special thanks to Ana and Zarko for being our the best friends. I thank all former PARTI summer students at Fermilab for the great time we had.

Working together with Pasha Murat and Sasha Pronko at Fermilab was an extremely valuable experience. Not only they deserve the major credit for the MMC algorithm, one of the main tool used in my dissertation, but also they have been always very helpful with many other aspects of my analysis. If I were stuck I knew that all it takes is to discuss the problem with Pasha, who is the true expert in every aspect of the CDF experiment, or Sasha, who is always ready to help. I also got a lot of help on software issues from Ray Culbertson. I thank Tom Junk for the help he gave me on statistical methods. Thanks to every member of the CDF Higgs group for valuable suggestions and comments on the analysis. I am thankful to Jeff Roe for the energy and enthusiasm he brought to our group at the final stage of the analysis.

The strength of the final result is owing to the enthusiasm of Alexei Safonov, my advisor, who have always been setting ambitious goals to his students and guided them to achieve them. I have learned a lot working with Alexei. That was a very important experience. Many thanks to Alexei for the suggestions he made to improve my dissertation. I also thank Bhaskar Dutta, Stephen Fulling, Peter McIntyre, Terukin Kamon and Ricardo Eusebi for serving on the dissertation Committee and making themselves available sometimes at a very short notice.

I also thank my entire family for their support during these years. My parents Lev and Tatyana and my brother Igor deserve sincere gratitude for their trust in my success. I can't thank enough Anna, my wife. Only you know what it really took me to complete this dissertation. I am extremely lucky to have you in my life. I admire your patience and I am greatfull for your support, encouragement and love. I am deeply indebted to you for all the great things you have done for me while not seeing me enough for six years. It's just impossible to express how happy I am to be with you.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

The current knowledge about the fundamental building blocks of matter and their interactions is encompassed in the framework of the so-called Standard Model (SM) of Particle Physics. The SM contains two different types of elementary particles, fermions, which constitute the matter, and bosons, the carriers of the fundamental forces between fermions and themselves. There are four known forces or types of interactions in nature: strong, electromagnetic, weak and gravity. Owing to its relatively low strength, the gravitational force does not produce effects observable in the experiments with elementary particles at the energies achievable today and is not part of the SM. The heart of the Standard Model is the electroweak symmetry breaking (EWSB) responsible for providing masses to the gauge bosons without violating the local gauge invariance. The EWSB implies existence of at least one neutral scalar particle, the Higgs boson. Despite a tremendous amount of high precision experimental evidence supporting the validity of the Standard Model, the Higgs boson remains the only undiscovered particle predicted by the Standard Model. Discovery of the Higgs boson or an affirmative proof that it does not exist in nature is therefore a key priority for the experimental high energy physics. Incidentally, nearly every other theory predicting "‘new physics" beyond the SM also requires a Higgs boson or its equivalents, further motivating the search for the Higgs boson.

Experimental searches for the Higgs boson are a key part of the research programs of all major collider experiments in the High Energy Physics (HEP). Because of a small signal production cross section of the Higgs boson and presence of very

---

This dissertation follows the style of Physical Review D.

large backgrounds, no single measurement today is capable to discover or exclude the existense of the Higgs boson on its own. Therefore, searches for the Higgs focus on combining statistical strength of many individual analyses relying on specific channels of the Higgs production and decay. While each such analysis has moderate sensitivity to Higgs boson production, statistical combination of the results obtained for different channels, experiments and even different colliders significantly increases the overall sensitivity of such global search for the Higgs boson.

The focal point of the work described here is the search for the Higgs boson in the channel where Higgs decays to a pair of tau leptons. While this channel has been known to have the potential for making a strong contribution to the overall sensitivity for Standard Model Higgs boson, the actual achievable sensitivity of experimental analyses in this channel has been significantly reduced by several experimental challenges. Apart from very large background overhelming the potential signal and challenge of maintaining highly efficient data selection for the final states with hadronically decaying tau leptons, one key challenge arises from the difficulty in reconstructing the invariant mass of the Higgs boson to separate Higgs decay events from very similar decays of Z bosons to pairs of taus. This difficulty has two underlying causes, one is the poor momentum measurement of visible momenta of hadronically decaying taus in the environment with frequent energy overlaps. The other one stems from the lack of an efficient reconstruction technique for reconstructing the invariant mass in the presence of large cancelation of invisible momenta of neutrinos from the decays of the two tau leptons in the event. Addressing these serious challenges is therefore a pre-requisite for exploring the full potential of the $H \to \tau\tau$ channel.

This dissertation outlines two novel techniques developed to improve the sensitivity of the search for the Higgs boson in di-tau channel, the Probabilistic Particle Flow Algorithm (PPFA) and the Missing Mass Calculator (MMC). These techniques are then deployed in the analysis searching for Higgs at the CDF experiment at Fermilab using proton-antiproton collisions from the Tevatron collider described in the second part of the manuscript.

CHAPTER II

THE STANDARD MODEL AND THE HIGGS BOSON

The Standard Model describes strong, electromagnetic and weak interactions using the principles of relativistic quantum field theory. In the 1960s Glashow [1], Weinberg [2] and Salam [3] developed unified electroweak theory describing electromagnetic and weak interactions in a single framework. The Lagrangian of electroweak theory is invariant under the combined gauge transformation $SU(2)_L \times U(1)_Y$ and it describes massless gauge bosons and massless fermions.

The SM uses the Higgs mechanism [4] to allow the gauge bosons acquire experementally observed masses while preserving local gauge invariance via spontaneous electroweak symmetry breaking (EWSB). It is achieved by introducing a scalar field potential so that the ground states do not share the symmetry of the Lagrangian while the local gauge invariance is preserved.

A. The Standard Model

Here we give a brief overview of the Standard Model and introduce the Higgs mechanism.

1. Particles and Forces

There are 12 fermions, particles of spin 1/2 (in units of $\hbar$, the Planck constant), that are considered in the SM to be truly elementary particles with no internal structure. Fermions are grouped in three generation. Each generation consist of two quarks (electric charge +2/3 and -1/3) and two leptons (electric charge -1 and 0). Each fermion has a corresponding antiparticle. The defining property of the quarks is the color charge enabling strong interaction between quarks. Because of non-vanishing electric charge and weak isospin quarks can also interact with other fermions via electroweak interaction. Leptons do not cary color charge and can only participate in electroweak interactions. Electrically neutral leptons are called neutrinos and can only interact via weak force.

Interactions between fermions are mediated by gauge bosons, spin 1 particles. Eight massless gluons mediate strong interactions between colored quarks. Gluons carry color charge and can self interact. Photons mediate the electromagnetic force between particles with non-zero electric charge and also are massless. W^+, W^- and Z gauge bosons are responsible for weak interactions between fermions. $W^\pm$ and Z bosons are experimentally known to be massive, requiring a special mechanism to introduce their masses into the theory while preserving its gauge invariance.

2. Electroweak Interactions and Gauge Invariance

The gauge invariance principle is best illustrated by an example from the Quantum Electrodynamics (QED). If we introduce a gauge field corresponding to a photon, A_μ, and a covariant derivative $D_\mu = \partial_\mu + ieA_\mu$ with e being the coupling strength (electromagnetic charge), the Lagrangian of the QED, can be written as

$$L_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.1}$$

where ψ is a spinor field of spin 1/2, γ^μ are Dirac matrices and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. This Lagrangian is invariant under local gauge transformation

$$\psi(x) \to e^{-i\alpha(x)}\psi(x) \tag{2.2}$$

$$A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha(x) \tag{2.3}$$

If photons were massive, the mass term would have a form of $m_\gamma^2 A_\mu A^\mu$. This term is not invariant under local gauge transformation and therefore gauge invariance of the L_{QED} would be broken if photons were to be massive. The requirement of photons being massless inspired by the gauge invariance is experimentally verified with very high precision.

Electromagnetic and weak force can be described as a single force under the $SU(2)_L \times U(1)_Y$ gauge group formalism. The $SU(2)_L$ group is associated with weak isospin, T, three gauge fields $W_\mu^{1,2,3}$ with coupling g. The $U(1)_Y$ group is associated with hypercharge, Y, and one gauge field B_μ with coupling g'. The local gauge transformation now take the form

$$\Psi_L \rightarrow e^{ig\alpha(x)\cdot T + ig'\beta(x)Y}\Psi_L \tag{2.4}$$

$$\Psi_R \rightarrow e^{ig'\beta(x)Y}\Psi_R \tag{2.5}$$

where $\Psi_L = \frac{1}{2}(1-\gamma_5)\Psi$ and $\Psi_R = \frac{1}{2}(1+\gamma_5)\Psi$ are left- and right handed fields, respectively. The Electroweak Lagrangian, L_{ew}, invariant under these new transformation is similar to L_{QED}. The covariant derivative is substituted by $D_\mu = \partial_\mu + igW_\mu \cdot T + ig'\frac{1}{2}B_\mu Y$ and the kinetic term $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ is substituted by $-\frac{1}{4}W_{\mu\nu} \cdot W_{\mu\nu} - \frac{1}{4}B_{\mu\nu} \cdot B_{\mu\nu}$, where $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ and $W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - g \cdot W_\mu \times W_\nu$.

Massless gauge fields of the electroweak theory are related to the experimentally observable massive $W^\pm$ and Z bosons and massless photon via a superposition of $W_\mu^{1,2,3}$ and B_μ fields:

$$A_\mu = cos\theta_W B_\mu + sin\theta_W W_\mu^3 \tag{2.6}$$

$$Z_\mu = -sin\theta_W B_\mu + cos\theta_W W_\mu^3 \tag{2.7}$$

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \pm W_\mu^2), \tag{2.8}$$

where θ_W is the Weinberg mixing angle defined as $cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}}$. However, so far W and B remain massless in the theory leaving the theory inconsistent with the experimental observations.

3. Higgs Mechanism

The three bosons mediating electroweak force W^+, W^- and Z are massive, while the Lagrangian describing electroweak interactions is not allowed to have a mass term. If the mass term were added explicitly to the Lagrangian, the gauge invariance would not be preserved. Higgs mechanism was formulated in 1964 to include massive particles into the model in a gauge invariant fashion. In this approach, an additional term $L_H = |D_\mu \Phi|^2 - V(\Phi^\dagger \Phi)$, where $V = \mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2$ is added to the gauge invariant electroweak Lagrangian. The vacuum expectation value is required to be non-zero by a specific choice of parameters μ and λ: $|\Phi_0| = \sqrt{-\frac{\mu^2}{\lambda}}$. When the Lagrangian L_H is expressed using $\Phi' = \Phi_0 + h(x)$, where Φ_0 is the ground state, the Lagrangian acquires a term -$2\mu^2 h^2$, corresponding to a massive scalar Higgs field h with mass $m_H = \sqrt{-2\mu^2}$. This mechanism where the ground state do not share the symmetry of the Lagrangian is called spontaneous symmetry breaking. When spontaneous symmetry breaking is introduced the Lagrangian of the theory remains invariant under the local gauge transformations while allowing the three gauge bosons to acquire mass.

B. Searches for the Higgs Boson

Because of the higher order corrections, indirect constraints on the Higgs mass can be obtained from the global fit of available data for the electroweak parameters sensitive to the mass of the Higgs boson, m_H. Examples of such parameters include the W boson mass, logarithmically dependent on m_H and the mass of the top quark, quadratically dependent on m_H (see Fig. 1). The result of the global fit of 18 electroweak parameters projected on m_H [5] is shown in Fig. 1. This fit results in $m_H = 89^{+35}_{-26}$ GeV/c^2 at the 68% confidence level (CL)[1].

LEP experiments have performed direct search for the Higgs boson in e^+e^- collisions at the center-of-mass energy, $\sqrt{s}$, ranging from 189 GeV to 209 GeV. The experiments observed no statistically significant evidence for the production of the Higgs boson and their combined lower limit on the Higgs mass is $m_H >$ 114.4 GeV/c^2 at the 95% CL. There are a number of ongoing searches for the Higgs boson in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV at the Tevatron by CDF and D0 collaborations and in pp collisions at $\sqrt{s} = 7$ TeV at the LHC by ATLAS and CMS collaborations. As of July 2011 the Tevatron has excluded the Higgs in the mass range $156 < m_H < 177$ GeV/c^2 at 95% CL [6]. ATLAS excluded mass ranges $155 < m_H < 210$ GeV/c^2 and $290 < m_H < 400$ GeV/c^2 at 95% CL [7]. CMS excluded $149 < m_H < 206$ GeV/c^2 and $300 < m_H < 440$ GeV/c^2 at 95% CL [8]. The sensitivity of the direct searches by Tevatron and LHC experiments is shown in Fig. 2.

While the LHC has taken the lead over the Tevatron in the search for the Higgs boson, the Tevatron data remain particularly valuable in the low mass region. As the low mass region is preferred by indirect measurements, a strong multi-experiment effort is aimed at improving the sensitivity to the Higgs boson in the mass region just above LEP exclusion.

---

[1] This fit does not include Summer 2011 Tevatron and LHC results.

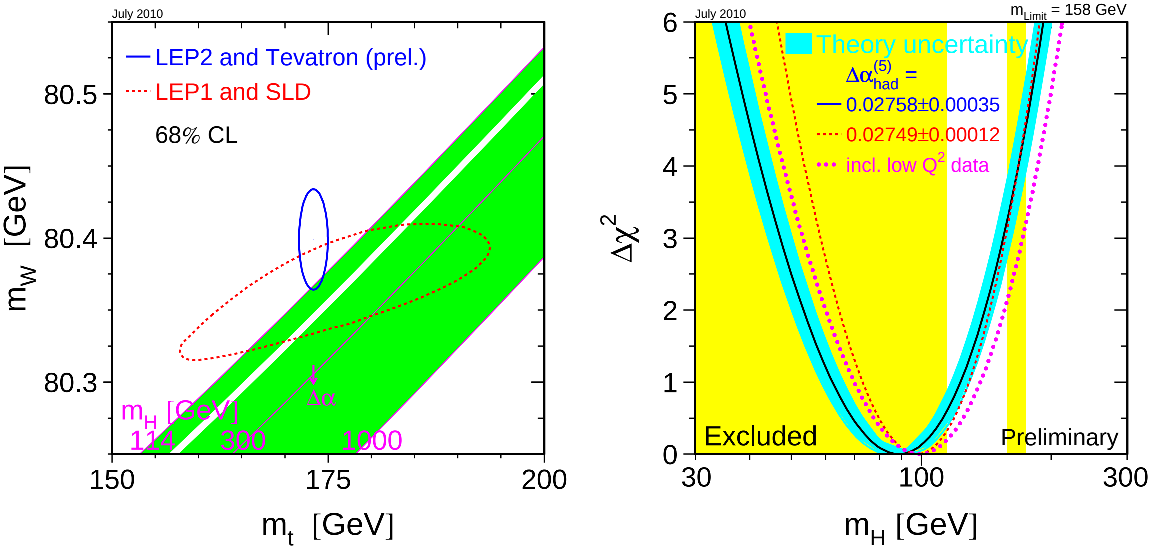

Fig. 1. Left: The comparison of the indirect constraints on m_W and m_t based on LEP-I/SLD data (dashed contour) and the direct measurements from the LEP--II/Tevatron experiments (solid contour). In both cases the 68% CL contours are plotted. Also shown is the SM relationship for the masses as a function of the Higgs mass. Right: The observed value of $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ derived from the fit of electroweak data as a function of the Higgs boson mass. The dark blue line the result of the fit using all high-Q^2 data; the light blue band is an estimate of the theoretical uncertainties. The dashed line is the result using alternative evaluation for the contribution of light quarks to the photon vacuum polarization, $\Delta\alpha_{had}^{(5)}$, as explained in Ref. [9] and references therein. The dotted curve corresponds to a fit including also the low-Q^2. The yellow shaded region shows mass region excluded at 95% CL by direct searches at LEP and the Tevatron (Summer 2011 results from the Tevatron and the LHC are not yet included).

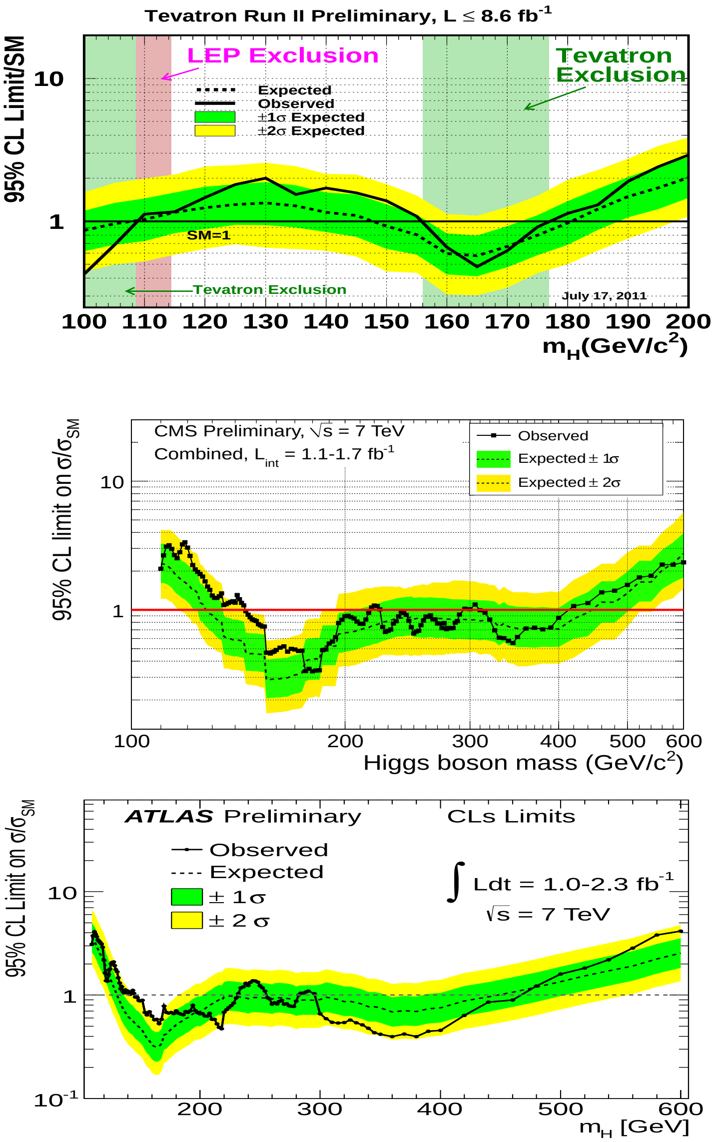

Fig. 2. Expected (dashed lines) and observed (solid lines) 95% CL upper limits on SM Higgs production cross-sections branching ratios as a function of Higgs mass. The green and yellow bands represent the 68% and 95% probability bands about the expected limits respectively. Shown are combined limits obtained at the Tevatron experiments (top) and CMS (middle) and ATLAS (bottom) experiments at the LHC.

C. Higgs Production at the Tevatron

As shown in Fig. 3, gluon fusion is the dominant production mechanism of the Higgs boson at the Tevatron. In the low mass region (115-150) the dominant decay modes for the Higgs boson are: $b\bar{b}$ ($\approx$90%), $\tau\tau$ ($\approx$7%) and $\gamma\gamma$ ($\approx$0.2%). Figure 3 shows the branching ratios for the Higgs decay as a function of its mass. It is important to note that in the low mass region no single channel has sensitivity to the Standard Model Higgs boson at the Tevatron. Only a combination of all production mechanisms and decay modes may provide enough experimental power to make conclusion about the existence of the Higgs boson. While the $b\bar{b}$ channel dominates in terms of the number of expected signal events, it has the highest background due to the high rate of the QCD jet production. Experimentally, the $b\bar{b}$ channel can only be efficiently utilized for the search for the Higgs boson, when Higgs is produced in association with W or Z bosons [10], [11], [12]. The channel where the Higgs boson decays to $\gamma\gamma$ has been recently investigated by the CDF and D0 collaborations [13], [14]. $H \rightarrow \tau\tau$ channel has many handles to suppress QCD background and has been already explored at CDF and D0 with smaller datasets [15], [16].

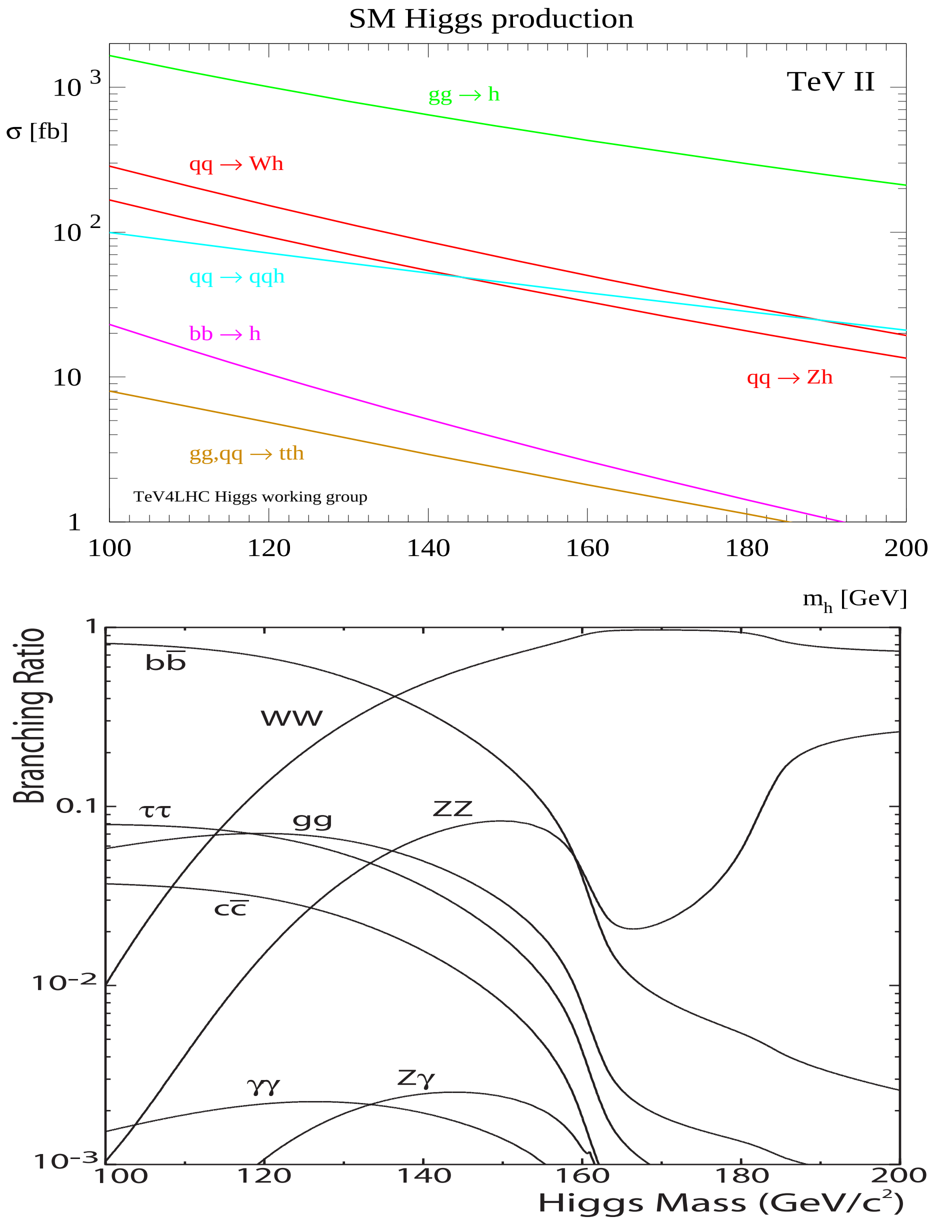

Fig. 3. The Higgs production cross section at the Tevatron (top) and decay branching ratio (bottom) as a function of the Higgs boson mass.

CHAPTER III

THE TEVATRON COLLIDER AND THE CDF DETECTOR

A. The Fermilab Accelerator Complex and The Tevatron

The Fermilab accelerator complex [17] is designed to produce protons and antiprotons beams and accelerate them for collisions at the center-of-mass energy $\sqrt{s}$=1.96 TeV. Figure 4 shows a schematic view of the complex. It consists of several major facilities each having its own dedicated task: production of the protons and antiprotons, their acceleration to the collision energy and maintaining achieved energy of the proton and antiproton beams.

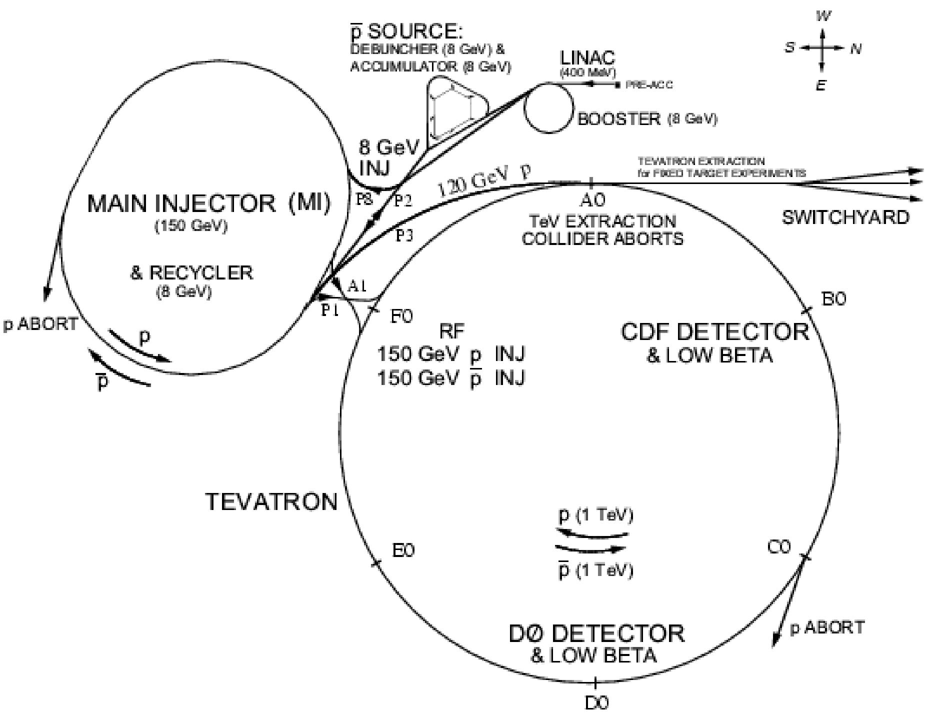

Fig. 4. The Tevatron accelerator chain.

1. The Proton Accelerator Chain

The Pre-accelerator, or Preacc, is the first accelerator in the Fermilab accelerator chain. It produce negatively charged hydrogen ions, H^-, and accelerate them to the energy of 750 keV. The Preacc repetition rate is 15 Hz. After beam exits the accelerating column, it travels through a transfer line to the Linac, the next level of acceleration for the negatively charged hydrogen ions. It takes the ions with an energy of 750 KeV and accelerates them to an energy of 400 MeV. The acceleration in the linac is achieved by the use of Radio Frequency (RF) cavities. RF cavities produce a standing wave which timing is such that particles traveling through a series of cavities receive additional energy in each cavity.

Booster is the next level of acceleration. It takes the 400 MeV negative hydrogen ions from the Linac. The electrons are removed from the ions leaving only protons which are accelerated to 8 GeV. The Booster is the first circular accelerator, or synchrotron, in the chain of accelerators. It consists of RF cavities for the acceleration and a series of magnets to keep particles in the ring of the 75 m radius. The repetition rate of the Booster is 15 Hz.

2. The Antiproton Source and Recycler

The antiprotons are created by 120 GeV proton beam from the main injector hitting the nickel alloy target. Magnets are used to extract antiprotons with momentum of 8 GeV out of other secondary particles produced on the target. The antiprotons are captured by into the Debuncher - triangular-shaped synchrotron with a mean radius of 90 meters. Efficient capture of the high momentum spread of the antiprotons is accomplished by RF manipulation called bunch rotation. Antiprotons are coolled in the Debuncher by stochastic cooling technique. After the Debuncher the antiproton beam is transfered into the Accumulator. It is also triangular-shaped synchrotron with mean radius of 75 meters and is located in the same tunnel as the Debuncher. The Accumulator serves as the storage ring of 8 GeV antiprotnos before they get transfered into the Recycler. The Accumulator has also a number of cooling systems.

The Recycler is an antiproton storage ring in the Main Injector tunnel. It accepts antiprotons from the Accumulator and cools them further than the Accumulator is capable. Apart from stochastic cooling used in the Recycler, electron cooling system is required for high intensity antiproton beam when stochastic cooling looses effectiveness. The particles in the Recycler are kept at a constant energy of 8 GeV until the Tevatron is ready to accept them, when they are passed to the Main Inhector.

3. The Main Injector and The Tevatron

The Main Injector accepts 8 GeV protons from the booster and 8 GeV antiprotons from the Recycler. The protons and antiprotons are accelerated to the energy of 150 GeV before they are injected to the Tevatron. The protons can be also accelerated to 120 GeV in the Main Injector and be used for antiproton production.

The Tevatron is a 4 miles circular synchrotron with eight accelerating cavities. The Tevatron can accept both protons and antiprotons from Main Injector and accelerate them from 150 GeV to 980 GeV in 85 s. In Collider mode, the Tevatron can store beam for hours at a time. For the high energy collisions the Tevatron is filled with 36 bunches of $\sim 3\times 10^{1}1$ protons and $\sim 7\times 10^{10}$ antiprotons. The bunches are groupped in three bunch trains of 12 bunches. The separation between bunch trains is 2.64 μs. Separation between bunches within a bunch train is 396 ns. Proton and anti-proton beams are orbiting in a helix trajectory in the same beam-pipe. Electrostatic separators produce a strong electric field to keep the two beams appart. One turn in the Tevatron takes 21 μs. The Tevatron employs superconducting magnets cooled with liquid hellium at 4.2 K. The magnets produce dipole field of 4.4 T.

B. The CDF II Detector

The CDF II detector [18] [19] is a general purpose detector for precision measurment of the momentum, energy and position of the particles produced in proton-antiproton collisions at the Tevatron. CDF II refers to a significantly upgraded version of the original CDF detector [20]. The CDF II detector[1] has been designed and constructed to perform a broad physics program [21] including detailed studies of the QCD processes, Heavy Flavour sector, measurement of the properties of top quark, Z and W bosons, searches for the Higgs boson and searches for the physics beyond the Standard Model.

Figure 5 shows the detector's cylindrical layout with accelerator beamline being the central axis of the cylinder. There are three major components of the detector. Tracking system is designed for precise reconstruction of the charged particles trajectories, and is the inermost part of the CDF detector embeded in 1.4 T uniform magnetic field produced by a superconducting solenoid. Calorimeter modules are located outside the solenoid arranged in a projective tower geometry and provide energy measurements for neutral and charged particles. The muon system, built as a series of drift chambers, constitute the outermost part of the detector. It is used to detect muons which are minimum-ionizing particles that usually pass throught the calorimeter.

---

[1] In the following referred to as simply "CDF".

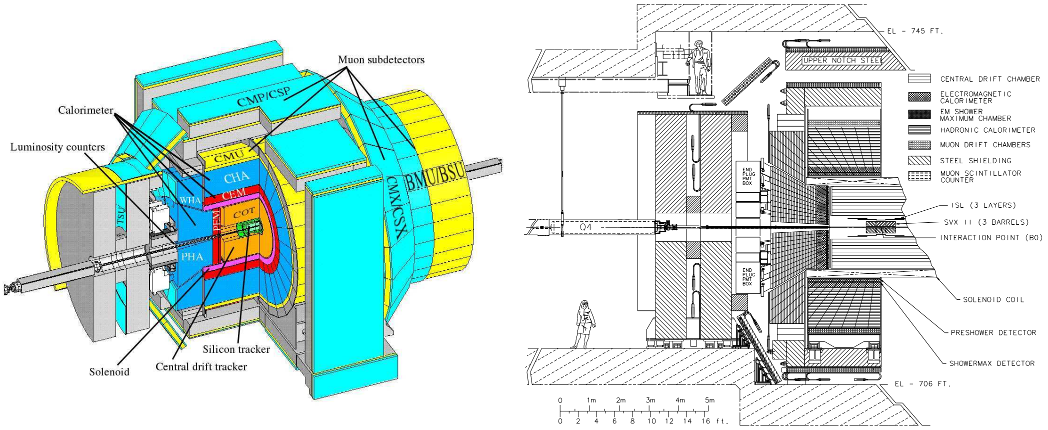

Fig. 5. The CDF detector: an isometric (left) and elevation (right) views.

1. Coordinate System

The z-axis of the CDF coordinate system is defined as the direction of the incoming protons and coinsides with the direction of the magnetic field. A particle direction at the origin is described by θ, the polar angle relative to the incoming proton beam; ϕ, the azimuthal angle about this beam axis; and z_0, the intersection point of the particle trajectory with the beam axis. Instead of θ, an alternative variable called pseudorapidity, $\eta = -ln(tan\frac{\theta}{2})$, is frequently used. This is because energies of the partons within proton and antiprotons are not equal to the beam energy of 980 GeV. Therefore the particle occupancy per unit polar angle, $dN/d\theta$ depends on the boost in z direction, since θ is not Lorentz invariant. Pseudorapidity is the ultrarelativistic approximation of the Lorentz invariant rapidity,

$$y = \frac{1}{2}ln\frac{1+\beta cos\theta}{1-\beta cos\theta} = \frac{1}{2}ln\frac{E+p_z}{E-p_z}. \tag{3.1}$$

For the high energy particles $p \approx E$ or $\beta \approx 1$ and therefore $y \approx \eta$. Now the number of particles per unit of pseudorapidity, $dN/d\eta$, is invariant under Lorentz transformation.

It is convenient to describe distance between the objects (particles or group of particles) in the detector in $\eta - \phi$ space by using ΔR variable defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The transverse momentum, p_T, is a component of the momentum projected on a plane which is normal to the beam axis, $p_T = p \cdot sin\theta$. Similarly, the transverse energy, E_T, is given by $E \cdot sin\theta$.

2. Cherenkov Luminosity Counters

The total integrated luminosity, $\int Ldt$, is calculated from the rate of the inelastic $p\bar{p}$ collisions measured with Cherenkov Luminosity Counters (CLC) using the expression

$$\int Ldt = \frac{\mu \times f}{\sigma_{in}} \tag{3.2}$$

where μ is the average number of inelastic $p\bar{p}$ collisions per bunch crossing, f is the Tevatron bunch crossing frequency (1.515×10^7 MHz), and σ_{in} is total inelastic $p\bar{p}$ cross section at 1.96 TeV (σ_{in}=60.7±2.4 mb).

The CLC consist of two modules installed around the beampipe at each end of the detector and provides coverage in the region $3.6<|\eta|<4.6$. Each module consists of 48 counters filled with isobutane gas at 22 psi. The counters arranged in three concentric layers of 16 counters around beam-pipe. The counters in the inner layer are 1.1 m long and the counters in the two outer layers are 1.8 m long. Each counter is made of reflective aluminized Mylar with light reflector to collect the Cherenkov light into photomultiplier tubes with high quatnum efficiency in the ultraviolet wave lenght range. The half-angle of the Cherenkov light cone is 3.1 which correspond to a momentum threshold for light emmision of 9.3 MeV/c for electrons and 2.6 GeV/c for pions. The overall uncertainty on the luminosity measurements is 5.9% which is dominated by the uncertainty in the absolute normalization of th CLC acceptance for a single $p\bar{p}$ inelastic colision.

3. Tracking System

The silicon strip vertex detector is the inermost tracking detector located immediately around the beam pipe and serves for high precision tracking and secondary vertex detection. It consists of three modules: Layer 00 (L00), Silicon Vertex Detector (SVX II) and Intermediate Silicon Layer (ISL). L00 is a single-sided silicon microstrip detector arranged in wedges at radius of 1.35 cm and 1.62 cm from the center of the beam pipe and covering η<4.0. The SVX II is doubled-sided silicon microstrip detector extending in radial direction from 2.1 cm to 17.3 cm covering η<2.0. The ISL is also double-sided silicon microstrip. The central layers of the ISL are located at a raduis of 22 cm and cover η<1.0. Forward and backward layers are located at r=20 cm and r=28 cm and cover 1.0<η<2.0.

The Central Outer Tracker (COT) is located directly outside of the silicon tracking detectors in the radial directions. The COT chamber has over 300,000 redout channels and cover the region $|\eta| < 1$. The chamber consist of 8 superlayers of 310 cm length cells at radii between 40 and 132 cm from the beam axis. Each superlayer consist of 12 layers of sense wires located between alternating layers of potential wires. Four out of eight superlayers are parallel to the beam axis (axial layers) and the other four superlayers are are strung at an angle of ±2 degree with respect to the beam axis. This allows to perform particle tracking in $z-$direction.

The COT is filled with argon and ethane mixes equally. This mixture ensures a fast drift velocity of ≈50 μm/ns, which is necessitated by the short intervals between beam bunch crossings. The maximum drift distance in the chamber is 0.88 cm which correspond to a drift time of the order of 200 ns.

The charge and momentum of a particle associated with a track are determined by the track curvature in the magnetic field of 1.4 T. The uniformity of the field in the tracking region $|z| <$150 cm and $r <$150 cm is 0.1%. The transverse momentum of a reconstructed track is determined as $p_T = 0.3 \cdot q \cdot B \cdot \rho$, where B is the magnetic field measured in T, q is the electric charge of the particle measured in the magnitude of the electron charge and ρ is the radius of the curvature of the track. The resolution decreases with p_T as track bend less and the precision of the curvature measurement degrades. The resolution of the COT is $\delta p_T/p_T^2 \approx$0.0015 (GeV/c)$^{-1}$.

4. Time of Flight System

Although data analysis presented in this dissertation does not use Time of Flight System (TOF) of the CDF detector, we summarize its design and performance for completeness. TOF was added in 2001 to improve particle identification. It was designed with the goal to distinguish low momentum kaons, pions and protons. It is achieved by measuring particle travel time from the interaction point to the TOF. The system consist of 216 scintillator bars 3 m length with cross section area of 4×4 cm^2. The bars are arranged around the COT at radius of 1.5 m. TOF system timing resolution is about 110 ps, which provides two standard deviation separation between kaons and pions for momentum below 1.6 GeV/c.

5. Calorimeter System

Calorimeter modules provide the energy measurement for both charged and neutral particles. These modules are sampling scintillator calorimeters arranged in towers around the outer edges of the central tracking volume. There are two calorimeter sections in the detector: a central barrel ($|\eta|$<1) section and plug section (1.1<η<3.64). Scintilator layers are parallel to the beam in the central calorimeter and perpendicular to the beam in the plug. The size of the projective readout towers of the central barrel is 0.1 in η and 15 degrees in ϕ. The size of the plug tower varies from 0.1 in η and 7.5 in ϕ at η=1.1 to 0.5 in η and 15 in ϕ at η=3.64. Each calorimeter module has electromagnetic and hadronic part. Central part of the calorimeter is extensively used in this dissertation. Central electromagnetic (CEM) and hadronic (CHA) calorimeters cover the pseudorapidity region of $|\eta| < 1$. CEM is a lead-scintillator calorimeter with resolution $\delta E_T/E_T = 0.135/\sqrt{E_T} \oplus 0.02$. CHA is an iron-scintillator calorimeter with the single pion energy resolution of $0.5/\sqrt{E_T} \oplus 0.03$. The Shower Maximum (CES) detector, consisting of a set of strip-wire chambers embedded inside the CEM at the expected maximum of the electromagnetic shower profile, enables measurement of the position of electromagnetic showers with a few mm accuracy by reconstructing clusters formed by strip and wires. While rarely used to measure energy of the electromagnetic showers, CES cluster's pulse height provides a measurement of electromagnetic shower energy with the resolution of $\delta E/E = 0.23$.

6. Muon System

Muons typically pass through the calorimeter modules without singificant losses in energy. There are two main parts of the muon system. The central part ($|\eta|$<0.6) consist of the Central Muon Detector (CMU) and the Central Muon Upgrade (CMP). Central Muon Extension Detector (CMX) provides coverage to a region with η<1.

The CMU consist of single-wire drift chambers located directly around the outer edge of the central calorimeter module and arranged in four concentric radial layers. The wires are strung parallel to the beam line. Wire pairs in layers 1 and 3 and layers 2 and 4 project radially back to the beam interaction point which allows for a coarse p_T measurement using the difference in signal arrival times on the two wires within a pair. Precision position measurement in ϕ direction is made by converting signal arrival times into drif distances in the transverse plane. The wires of cells in neighbouring stacks are connected via resistive wires at the non-redout end cells to provide for a coarse measurement of the hit positions along $z-$direction. In Run II these chambers operate in proportional mode [2]. The maximum drift time within a CMU cell is 800 ns which is longer than the 396 ns spacing between bunch crossings in the accelerator. The ambiguity as to which beam crossing a particular CMU hit originates from is resolved in both the trigger and the offline reconstruction using timing information associated with a matched COT track and/or matching energy in the calorimeter.

CMP and CMX are wire drift chambers run in proportional mode. They are identical except for their lengths along the direction of the wire which is larger for CMP chambers. These drift cells are roughly a factor of two wider than those in the CMU detector resulting in a longer maximum drift time of 1.8 μs. Occupancies in these chambers are small enough to uniquely determine the appropriate beam-crossing from COT track matching. The CMP chambers are arranged in a box-like structure around the outside of the CMU detector and an additional 3λ of steel absorber which is sandwiched between the two detectors. The additional steel greatly reduces hadronic punchthrough into the CMP chambers and allows for cleaner muon identification. Drift cells in the CMX detector are arranged in conical arrays of eight staggered layers. The partial overlap between drift tubes in the CMX conical arrangement allows for a rough hit position measurement in the z coordinate utilizing the different stereo angles of each cell with respect to the beam axis.

---

[2] CMU chambers were operated in streamer mode during Run I.

CHAPTER IV

HADRONIC TAU RECONSTRUCTION AND THE PROBABILISTIC PARTICLE FLOW ALGORITHM

A. Introduction to the Particle Flow Algorithm

Accurate measurement of energy of hadronic jets is critical both for precision verification of the Standard Model (SM) as well as searches for new physics at current and future collider experiments. A standard technique for measuring jet energy relies on clustering spatially close energy depositions in the calorimeter, the detector designed to measure the energy of particles that produce electromagnetic or hadronic showers in the absorber material. Given that on average about 70% of a typical jet energy is carried by particles interacting hadronically[1] (mostly $\pi^{\pm}$, but also $K^{\pm}$, K_L^0, protons, neutrons), accuracy of the jet energy measurement is driven by the calorimeter precision in measuring energy of hadronic showers. While the energy of electromagnetic showers can be measured very well, large fluctuations in the development of hadronic showers lead to a significantly lower precision[2]. The non-equal response of the non-compensating calorimeters to electromagnetic and hadronic showers[3], further biases the overall jet energy scale and reduces resolution requiring special corrections, which can only partially recover this reduction. While presence of many particles in a jet averages out fluctuations in the measurement of energy of individual hadronic show-

---

[1] the remaining 30% is mainly due to neutral pions decaying to pairs of photons, which produce electromagnetic showers.

[2] A typical example is the CDF calorimeter, which has good electromagnetic calorimeter resolution $\delta E/E \sim 0.135/\sqrt{E}$ while the response to stable hadrons, e.g. charged pions, is substantially less precise $\delta E/E \sim 0.5/\sqrt{E}$.

[3] E.g., main calorimeter systems at ATLAS, CDF, and CMS are all non-compensating.

ers, jet energy resolution remains poor for jets of low (∼10–30 GeV) and moderate (∼30–60 GeV) energies. Incidentally, resolution of low-to-moderate energy jets has a strong impact on sensitivity of many physics analyses, from electroweak precision measurements to searches for Supersymmetry or Higgs in $b\bar{b}$ and $\tau\tau$ channels, motivating development of improved jet energy measurement techniques. Furthermore, as mismeasurements in the energy of jets in an event bias the missing transverse energy, better jet energy resolution improves missing transverse energy measurement, a key discriminant in many searches for new physics.

A significant improvement in the jet energy resolution at hadron collider experiments has been achieved with the deployment of a technique known as the Particle Flow Algorithm (PFA). PFA achieves better jet energy resolution by reconstructing and measuring energy of individual particles in a jet using information from several detector sub-systems. For example, momentum of a charged hadron can be precisely measured using tracking (except for the case of very high transverse momenta, which is not relevant for this discussion) allowing one to replace with it the less accurate calorimeter measurement of the energy carried by charged hadrons in the PFA jet energy calculation:

$$E_{jet} = \sum_{tracks} E_{trk} + \sum_{\gamma' s} E_{\gamma} + \sum_{n} E_{n}, \tag{4.1}$$

where the first term is the energy of the charged particles in the jet, the second term accounts for energy of photons accurately measured in the electromagnetic calorimeter, and E_n is the energy of stable neutral hadrons, e.g. neutrons or K_L^0, which still relies on the hadron calorimeter. The corresponding relative jet energy resolution can be written in terms of single particle relative resolutions as:

$$\frac{\sigma^2(E_{jet})}{E_{jet}^2} = \frac{1}{E_{jet}^2} \times \left(\sum_{tracks} E_{trk}^2 \frac{\sigma^2(E_{trk})}{E_{trk}^2} + \sum_{\gamma' s} E_{\gamma}^2 \frac{\sigma^2(E_{\gamma})}{E_{\gamma}^2} + \sum_{n' s} E_{n}^2 \frac{\sigma^2(E_{n})}{E_{n}^2} \right) \tag{4.2}$$

Note that only the last term depends on the potentially poor calorimeter resolution for energy of hadronic showers. However, because the average fraction of the jet energy carried by stable neutral hadrons is on average only around 10%, it's contribution to the overall jet energy uncertainty is strongly suppressed by $\sum E_n/E_{jet}$. With the remaining 90% of energy accurately measured either in the tracker or in the electromagnetic calorimeter, achievable jet energy accuracy can substantially outperform the traditional calorimeter-only based measurements. Furthermore, the bias in the energy scale related to calorimeter non-compensation is significantly reduced as it is only present in the suppressed third term and can be easily corrected.

Apart from an obvious pre-requisite of highly efficient tracking, the performance of a PFA-based reconstruction in a realistic setting depends critically on one's ability to correctly identify and separate calorimeter energy depositions from spatially close particles. One example illustrating the issue is an overlap of energy deposits in the calorimeter due to a charged pion and a neutron. In this case one has to "guess" the fraction of the measured calorimeter energy deposited by the charged pion, so that the excess can be attributed to a neutral hadron. The dependence of jet energy resolution on the overlap effects is sometimes parameterized by amending Eq. (4.2) with the so called "confusion term" σ^2_{conf}. The size of the confusion term depends on the power of the algorithm and the detector design features, but it generally increases with the coarser calorimeter segmentation and higher particle densities. In extreme cases, the large size of the confusion term can completely eliminate the advantages of the PFA schema over traditional calorimeter-based measurement.

PFA-based algorithms have been successfully implemented at LEP in the 1990's [22] and have been pursued in developing physics program at the International Linear Collider (ILC) [23]. At hadron collider experiments, a simplified version of a PFA-based algorithm was implemented for reconstructing hadronically decaying tau leptons at

CDF at the end of Run I [24] providing strong improvement in hadronic tau jet energy resolution. It was further improved and used at CDF for Run II analyses [15]. A more comprehensive implementation of the same technique has been shown to improve generic jet resolution at CDF compared to calorimeter-only reconstruction. However, the confusion term associated with frequent energy overlaps owing to the coarse segmentation of the CDF calorimeter towers has allowed only a limited improvement. A complete PFA algorithm has been developed by the CMS experiment [25] allowing a strong improvement in jet energy and missing transverse energy scale and resolution. The CMS detector is nearly ideally suited for PFA-based reconstruction due to its fine granularity of the electromagnetic calorimeter and the longitudinal profiling of hadronic showers, which improves their spatial resolution. However, the expected "High Luminosity LHC" upgrades causing large increases in particle occupancies per event will require developing techniques for resolving energy overlaps in order to maintain the performance of the PFA reconstruction.

In this chapter, we discuss the challenges and implications of deploying a PFA-based reconstruction in the environment with frequent energy overlaps (Section B). In Section C we present a technique designed to resolve the overlapping energy depositions of spatially close particles using a statistically consistent probabilistic procedure. In addition to improving energy resolution, the technique allows combining measurements from multiple detectors as opposed to "substituting" one measurement with another in existing algorithms and is nearly free of ad-hoc corrections thus minimizing distortions due to the discontinuities of the correction functions. It provides additional handles, such as the measurement of jet energy uncertainty on a jet-by-jet basis and the measure of the overall consistency of the measurement, improving sensitivity of physics analyses. In Section D, we describe implementation of this technique for reconstructing hadronic tau jets at CDF and illustrate its performance with real data in Section E.

B. Challenges of the Large Energy Overlap Environment

Reconstruction of hadronically decaying tau jets with the CDF detector is a good example of a problem with frequent overlaps of energy deposits by nearby particles. The CDF calorimeter has projective tower geometry with azimuthal segmentation $\phi = 15^\circ$ and pseudorapidity segmentation $\eta \approx 0.1$ and does not provide measurements of either lateral or longitudinal profile (except having two separate energy measurements for deposits in the electromagnetic and hadron compartments of a tower). With a typical angular size of a hadronic tau jet of the order of 0.05-0.1 rad, there is a substantial probability for several or even all particles within the tau jet to cross the face of the calorimeter within the boundaries of a single calorimeter tower. Treatment of frequent energy overlaps is therefore a key consideration in designing a PFA-based reconstruction at CDF.

To set the stage, we need to briefly remind charachteristics of the sub-detector systems used in tau reconstruction and identification, a full description of the CDF detector is available in Chapter 3. The CDF tracking system provides nearly 100% efficient tracking within the pseudorapidity range of $|\eta| < 1$ relevant to tau reconstruction. It's main element is the Central Outer Tracker (COT), a drift chamber covering radii from 0.4 m to 1.37 m, providing momentum resolution of $\delta p_T/p_T^2 \approx 0.0015(\mathrm{GeV}/c)^{-1}$. If available, hits from the silicon vertex detector (SVX) are added to the COT information further improving the resolution. Central electromagnetic (CEM) and hadronic (CHA) calorimeters cover the pseudorapidity region of $|\eta| < 1.1$. CEM is a lead-scintillator calorimeter with resolution $\delta E_T/E_T = 0.135/\sqrt{E_T} \oplus 0.02$. CHA is an iron-scintillator calorimeter with the single pion energy resolution of

$0.5/\sqrt{E_T} \oplus 0.03$. Both calorimeters have a projective tower geometry with tower size $\Delta\phi \times \Delta\eta \approx 15^\circ \times 0.1$ and neither of the calorimeters measures either the longitudinal or lateral shower profile. The Shower Maximum (CES) detector, consisting of a set of strip-wire chambers embedded inside the CEM at the expected maximum of the electromagnetic shower profile, enables measurement of the position of electromagnetic showers with a few mm accuracy by reconstructing clusters formed by strip and wires. While rarely used to measure energy of the electromagnetic showers, CES cluster's pulse height provides a measurement of electromagnetic shower energy with the resolution of $\delta E/E = 0.23$. Because pulse heights of the one-dimensional strip and wire clusters reconstructed for the same shower are typically within $\approx 7\%$ of each other[4], multiple showers within a single CES chamber can typically be correctly reconstructed by matching the 1D strip and wire clusters using their pulse heights. The much broader hadronic showers frequently extend over multiple CHA towers and their spatial position can only be inferred from the energy measured for each tower. Early hadronic showers can deposit part of their energy in CEM and may produce clusters in CES, which sometimes complicates reconstruction of CES clusters, e.g. if overlapping with showers produced by photons in the same jet.

Let us consider a relatively simple example of reconstructing a jet containing a charged pion π^+, a neutral pion π^0 decaying to two unresolved photons $\gamma_1\gamma_2$ (appearing as a single electromagnetic cluster), and possibly a neutral hadron n. While the π^+ momentum is known from the tracker, energy estimation for neutral particles relies on the calorimeter measurement. However, the energy registered in the electromagnetic and hadronic parts of the calorimeter, E^{EM}_{meas} and E^{HAD}_{meas}, is a sum of the

---

[4] The CES energy resolution is driven by the fluctuations in the amount of ionization produced inside the CES chambers and not by the measurement of the charge collected on strips and wires.

unknown deposits by each of the particles in the jet, including that by the charged pion:

$$E^{EM}_{meas} = E^{EM}_{\pi^+} + E^{EM}_{\gamma_1\gamma_2} + E^{EM}_{n} \tag{4.3}$$

$$E^{HAD}_{meas} = E^{HAD}_{\pi^+} + (E^{HAD}_{\gamma_1\gamma_2}) + E^{HAD}_{n}, \tag{4.4}$$

resulting in an under-constrained system with two equations and six unknowns. Because the leakage of the electromagnetic showers from photons into the hadron calorimeter is typically small, as illustrated in Fig. 6(a) showing E^{EM} vs. E^{HAD} for simulated electrons, the corresponding term, shown in parentheses in Eq.(4.4), could be neglected. While it helps in reducing the number of unknowns, solving the problem requires disentangling contributions from hadronically interacting particles. While $E^{EM}_{\pi^+}$ and $E^{HAD}_{\pi^+}$ terms are correlated with the accurately measured momentum of π^+, the correlation is not trivial as illustrated in Fig. 6(b) showing the 2D distribution of E^{EM} vs. E^{HAD} for a simulated sample of charged pions with $p_{\pi^+} = 25$ GeV/c. The complex shape of the dependence owes to the large fluctuations in the development of hadronic showers and the non-compensating nature of the CDF calorimeter. Because $E^{EM}_{\pi^+}$ cannot be reliably estimated, and E^{EM}_{n} is completely unconstrained, momentum of π^0 cannot be calculated directly. Estimating the jet energy directly in the PFA approach is therefore hampered by two issues: (i) difficulty in estimating E^{EM} for hadronically interacting particles required to evaluate π^0 momentum, and (ii) difficulty in estimating $E^{HAD}_{\pi^+}$ required to estimate momentum of n. Measuring momentum of a combined $\pi^0 + n$ system, e.g. by "guessing" charged pion depositions and assigning the rest to the $\pi^0 + n$ system is nearly exactly equivalent to measuring jet energy using the calorimeter only thus negating all advantages of the PFA technique.

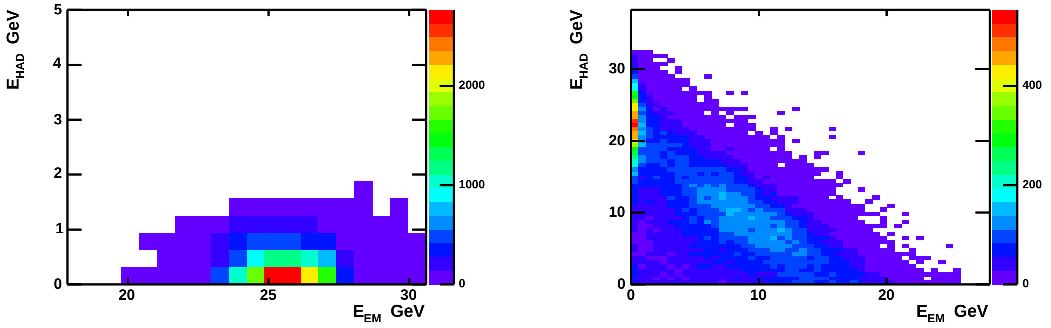

Fig. 6. Examples of the calorimeter response for (a) simulated isolated electrons with $p = 25$ GeV/c and (b) simulated isolated charged pions with $p = 25$ GeV/c in the plane E^{EM} versus E^{HAD}.

An algorithm based on solving Eqs.(4.3,4.4) directly with a specific purpose of reconstructing hadronic tau jets was implemented in the "tracks+π^0's" algorithm at CDF and used in the early Run-II analyses. The idea was to simplify the problem by assuming absence of neutral hadrons and estimate $E^{EM}_{\pi^+}$ as an average electromagnetic deposition of a charged pion with the momentum measured in the tracker. Then the remaining portion of the measured electromagnetic energy can be taken as the energy of the π^0 (Eq.(4.3)). An alternative method tried was to assume minimal ionization by the charged pion in the electromagnetic calorimeter. While delivering a significant improvement over the calorimeter-only measurement for a large fraction of events, the algorithm featured long tails in the energy resolution due to jets with substantial overlaps of particles in the same calorimeter tower. In physics analyses, undermeasuring energy of quark or gluon jets containing neutral hadronically interacting particles also lead to an increase in background contamination. To address these issues, additional corrections based on detecting incompatibilities of the reconstructed energy with the initially unused Eq.(4.4) or gross disagreements with the low resolution measurement of π^0 energy in the Shower Maximum detector have been developed. These adjustments improved algorithm performance by correcting jets with the most obvious inconsistencies. However, the ad-hoc nature and complexity of the corrections as well as the algorithm's inability to consistently treat correlations and incorporate other available measurements, e.g. energy measurement in the hadronic calorimeter and the measurement of electromagnetic shower energy in the Shower Maximum detector, motivated developing a more comprehensive method.

C. PPFA: The Probabilistic Particle Flow Algorithm

The challenge of solving an underconstrained system with high correlations and additional redundant measurements outlined in previous section can be addressed with a probabilistic approach. For every particle content hypothesis (the number of particles of each type) in a jet, one can build a probability estimator (likelihood) for any given set of momenta p_i of each particle in a jet ($i = 1, ..., i_p$ is the particle index) to result in a particular set of detector measurements (which could represent energy counts in calorimeter towers, cluster energies, track momenta or any other measurement) E^j_{meas} ($j = 1, ..., j_m$ runs over all available measurements):

$$\mathcal{L}(\vec{p}\,|\vec{E}_{meas}) = \int \prod_{j=det} \mathcal{M}(E^1_1, ..., E^{j_m}_{i_p}, E^1_{meas}, ..., E^{j_m}_{meas}) \times \mathcal{P}_{ij}(E^j_i|p_i) dE^j_i, \qquad (4.5)$$

where $\mathcal{P}_{ij}(E^j_i|p_i)$ is the "response function" for particle i with true momentum p_i to produce a contribution E^j_i to a measurement j, and the matrix $\mathcal{M}$ contains information about correlations between contributions of each particle to each measurement. One example of the latter is the correlation between the the deposits of energy E^i_j in electromagnetic calorimeter tower j by all particles crossing it giving a term $\delta(\sum\limits_i E^j_i - E^j_{meas})$ in $\mathcal{M}$. Another example is the correlation between the energy deposits by particle i in the electromagnetic and hadron calorimeter clusters (or towers) j_1 and j_2 it crosses, in which case the corresponding term may have a fairly complex form $f(E^i_{j_1}, E^i_{j_2})$. If such global likelihood function were constructed, $\vec{p}^0$ corresponding to its maximum will determine the most probable set of particle momenta, thus achieving the goal of fully reconstructing the event using all available detector information. The type of each particle and their number can be taken as parameters of the global likelihood allowing one to also determine the most probable particle content of a jet.

While building a global and fully inclusive likelihood is certainly possible, it is hardly practical. However, this approach can be used to solve specific problems like measuring jet energy in the environment with frequent energy overlaps in the calorimeter. Here, we will describe an example of such possible PPFA implementation. For simplicity, this example will use energy of pre-reconstructed electromagnetic and hadronic calorimeter clusters as the basic measurements E^j_{meas}, but an implementation using tower energy measurements would be very similar. The PPFA probability for a set of particles with momenta p_i to produce a set of calorimeter measurements E^j_{meas} for each cluster j in electromagnetic and hadron calorimeter can be written as follows:

$$\mathcal{L}_p(\vec{p}\,|\vec{E}_{meas}) = \int \prod_{j=clusters} \delta(\sum_i E^j_i - E^j_{meas})\mathcal{M}\mathcal{P}_{ij}(E^j_i|p_i)dE^j_i, \tag{4.6}$$

where $\vec{p}$ is the vector of particle momenta p_i, index i runs over the list of particles in a jet, index j runs over the available measurements (in our example, electromagnetic and hadronic calorimeters cluster energy measurements), E^j_i is the energy deposited by i^{th} particle in cluster j, E^j_{meas} is the measured energy for cluster j, and the response function $\mathcal{P}_{ij}(E^j_i|p_i)$ is the probability for particle i with true momentum p_i to deposit energy E^j_i in cluster j ($\mathcal{P}_{ij}$ depends on the type of particle), $\mathcal{M}$ is not yet defined correlation matrix. The likelihood $\mathcal{L}_p$ is essentially a sum of probabilities of all possible outcomes (specific values of energy deposited by each particle in the electromagnetic and hadronic calorimeter clusters) consistent with the actual cluster energy measurements. The probability of each outcome is a product of probabilities $\mathcal{P}_{ij}$ for each particle to deposit given amount of energy E_i in the hadronic and electromagnetic calorimeters given their assumed true momenta p_i. In this example, $\mathcal{M}$ in Eq.(4.6) is needed to account for the correlation of deposits by the same particle in the electromagnetic and hadronic calorimeters, e.g. early showering of a charged hadron leads to larger deposition in the electromagnetic calorimeter and reduced energy deposited in the hadron calorimeter. The easiest way to take that into account is to switch to two-dimensional response functions $\mathcal{P}^{CAL}(E_i^{EM_{j_1}}, E_i^{HAD_{j_2}}|p_i)$, where j_1 and j_2 are the indeces of the electromagnetic and hadronic calorimeter clusters the particle traverses. In this schema, the distributions previously shown in Fig. 6(a) and (b) can be normalized and used as $\mathcal{P}^{CAL}(E^{EM}, E^{HAD}|p)$ for electrons and charged pions, respectively.

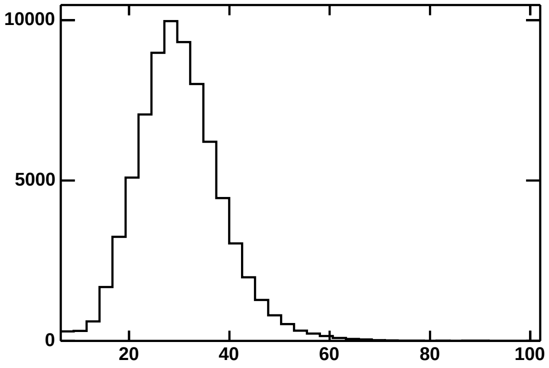

Fig. 7. Examples of the Shower maximum detector response functions for simulated isolated photons with momenta $p = 25$ GeV/c. Arbitrary units are chosen.

Additional measurements can be easily incorporated by modifying the likelihood function with Bayesian-like "priors". For example, tracking, preshower, or Shower Maximum information can be added by multiplying the initial likelihood function by a probability to measure a certain track momentum or pulse height given the assumed true momentum of a charged pions, electrons or photons. For example, distribution shown in Fig. 7 upon normalization can be used as the Shower Maximum response functions $\mathcal{P}_{\gamma}^{CES}(E^{CES}|p_{\gamma})$ for photons with momenta $p = 25$ GeV/c.

The most probable set of particle momenta $\vec{p}^0$ is obtained by maximizing the likelihood $\mathcal{L}_p(\vec{p}\ |\vec{E}_{meas})$, its shape in the $\vec{p}$ space can be used to evaluate of the uncertainty in the energy of each particle. If one primarily seeks to measure the energy of the entire jet, one can use the likelihood to obtain a "posterior" distribution for the jet energy (defined as a sum of the energies of the constituent particles).

$$\mathcal{L}_E(E_{jet}|\vec{E}_{meas}) = \int \mathcal{L}_p(\vec{p}\ |\vec{E}_{meas}) \times \delta(\sum_{i=1}^{N} p_i - E_{jet}) d\vec{p}, \tag{4.7}$$

as in the presence of correlations the latter provides a more convenient estimate of the jet energy and its uncertainty. The shape of the jet energy "posterior" allows estimating the uncertainty in the measured jet energy. Once the most likely set of particle momenta $\vec{p}^0$ is found, one can further test the "goodness" of the particle hypothesis. We define a p-value as the probability to observe a combination of detector measurements $\vec{E}'_{meas}$ that is equally or less likely than the actual set $\vec{E}_{meas}$ observed in the event, given that the true combination of particles and momenta is the one that maximizes the likelihood in Eq. (4.6):

$$p = \frac{\displaystyle\int_{\mathcal{L}_p(\vec{p}^0\ |\vec{E}'_{meas}) < \mathcal{L}_p(\vec{p}^0\ |\vec{E}_{meas})} \mathcal{L}_p(\vec{p}^0\ |\vec{E}'_{meas}) d\vec{E}'_{meas}}{\displaystyle\int \mathcal{L}_p(\vec{p}^0\ |\vec{E}'_{meas}) d\vec{E}'_{meas}} \tag{4.8}$$

In practice, the p-value can be easily calculated by generating "pseudo-experiments," in which one generates “pseudo-deposits” of energy by each particle with momenta p_i^0 towards each cluster energy measurement using the same response functions. The sum of the deposits of all particles crossing particular clusters yields a set of pseudo-measurements $\vec{E}'_{meas}$. The probability of the generated outcome is given by $\mathcal{L}_p$, and the integrated probability of observing equally or less probable set of measurements gives the p-value. Too low p-value will indicate that the initial hypothesis should be modified. Note that interpreting measured p-values has to be done carefully as arbitrary addition of new particles to make the observed calorimeter response “perfect” may degrade the resolution by biasing the measurement towards the calorimeter based jet energy measurement .

D. PPFA Implementation for Hadronic Tau Reconstruction at CDF

In this section we describe a practical implementation of the method developed for hadronic tau jet reconstruction at CDF. In the following, we discuss the CDF baseline hadronic tau jet reconstruction, which is used as a starting point for the algorithm. We then discuss the PPFA strategy, measurement of response functions, mathematical definition of the PPFA likelihood function and the “p-value,” and corrections. We conclude with evaluating algorithm’s energy resolution using simulation.

1. Baseline Hadronic Tau Jet Reconstruction at CDF

A hadronic tau jet candidate is defined as a narrow cluster of calorimeter energy with a seed tower of $E_T > 5$ GeV/c and at least one track with $p_T > 5$ GeV/c pointing to the calorimeter cluster. The narrow cluster is defined as a cluster with no more than six contiguous calorimeter towers with $E_T > 1$ GeV/c and is required to be in the central part of the detector ($|\eta| \leq 1$) to ensure high tracking efficiency. Given the size of the CDF calorimeter towers of $\Delta\eta \times \Delta\phi \sim 0.1 \times 0.2$, efficiency of the calorimeter requirements is very high for hadronically decaying taus with visible $p_T > 10$ GeV/c. Seed track p_T requirement brings a non-negligible inefficiency for tau jets of low-to-moderate visible momentum, but its strong power in rejecting quark and gluon jet backgrounds made it a standard in all CDF analyses involving hadronic tau jets. Next, all nearby (within a signal cone of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.17$ around the seed track) are associated with the tau candidate to be used in further analysis.

2. Implementation Strategy

The likelihood-based PFA algorithm starts with the initial hypothesis that every reconstructed track is a charged pion, every reconstructed cluster in the Shower Maximum detector with no track pointing to it is a photon, and no other particles are present in the jet. While this initial hypothesis can be corrected at a later point in the algorithm, in most cases it turns out to be correct owing to the low rate of track and Shower Maximum reconstruction failures and the low branching fraction of hadronic tau lepton decays for modes with neutral hadrons except π^0's. Next, we define the probability function using pre-calculated response functions (details for both are discussed in the following two sub-sections) and perform a scan in the multi-dimensional parameter space of momenta of the particles assumed to comprise the hadronic tau jet searching for the maximum of the likelihood function.

After the most likely combination of momenta of particles is determined, we construct the "p-value" which measures the probability that given particle content and momenta hypothesis could result in detector measurements less likely or equal than the observed response. If the p-value is too low, the hypothesized particle content is modified by adding a photon, which is assumed to be not reconstructed either due to a rare failure or an overlap with the extrapolated position of the charged track

(Shower Maximum cluster will be vetoed if it is reconstructed too close to the track), and the full calculation is repeated. If the p-value remains too low, the particle content is modified by adding a neutral hadron and the likelihood calculation is repeated. The procedure iterates until an acceptable outcome is achieved or after running out of the pre-set options as discussed in what follows.

3. Response Functions of the CDF Detector Sub-systems

As discussed earlier, the relevant detector measurements include tracking, measurements of energy deposited in the electromagnetic and hadronic calorimeter towers, and the measured CES cluster energy. Because the precision of the CDF tracking is much higher than the accuracy of other measurements, the tracker response function for charged pions as a function of pion momenta can be safely approximated by a delta function to simplify further calculations. To determine the calorimeter response functions for charged pions, we use CDF GEANT-3 [26] based simulation package, which was tuned using the test beam data. Isolated pions are selected using hadronic tau decays $\tau^{\pm} \rightarrow \pi^{\pm}\nu_{\tau}$ from an inclusive $Z/\gamma^{*} \rightarrow \tau\tau$ simulated sample of events generated with Pythia [27]. We calculate response functions for charged pions with momenta ranging from 1 to 100 GeV/c in steps of 1 GeV/c. Large fluctuations in the development of hadronic showers and their large lateral size, frequently spanning across several CHA towers, make it impractical to calculate responses separately for each tower in a multi-tower cluster. Instead, we measure hadronic calorimeter response for charged pions by summing tower energies in a square of 3×3 towers centered on the extrapolated position of the $\pi^{\pm}$ track. In the CEM, hadronic showers rarely deposit energy in more than a single tower, therefore charged pion electromagnetic deposition is calculated using the energy in the tower pointed at by the track associated with $\pi^{\pm}$. To take into account the strong correlation of the energy depositions in CEM and CHA, we define a 2-dimensional response function in the E^{EM} versus E^{HAD} plane. Fig. 6(b) shows an example of the calorimeter responses in CEM and CHA for simulated isolated charged pions with momenta $25 < p_\pi < 26$ GeV/c. We verify the accuracy of response functions obtained using simulation by comparing them with those obtained in a pure sample of isolated charged pions in data. When normalized to unity, these response functions represent the probability density functions (PDF) for a charged pion of particular momentum to produce given response in the calorimeter, which we will refer to as $\mathcal{P}_\pi^{CAL}(E^{EM}, E^{HAD}|p_\pi)$.

Vast majority of photons in tau jets originate from $\pi^0 \to \gamma\gamma$ and typically have energy of the order of a few GeV, making the accurate understanding of the calorimeter response for low energy photons particularly important. While the response functions for photons can be measured directly from the simulation, validating them with the data can be difficult owing to the challenges in obtaining a high purity sample of low energy photons in data. Fortunately, the calorimeter response to photons and electrons is nearly identical allowing use of a relatively high purity sample of electrons in data obtained by tagging photon conversions. Similar to the case of charged pions, we calculate 2-dimensional response functions for photons with momenta ranging from 1 to 100 GeV/c in steps of 1 GeV/c in the E^{EM} versus E^{HAD} plane. Fig. 6(a) shows an example of the calorimeter response function for photons with the true momenta $25 < p_\gamma < 26$ GeV/c. We denote the response functions of this type as $\mathcal{P}_\gamma^{CAL}(E^{EM}, E^{HAD}|p_\gamma)$.

As mentioned earlier, the CES energy measurement is used in the likelihood function as, despite its modest resolution, it can help correctly assign energies in difficult cases. Because photon candidates reconstructed in CES have highly correlated strip and wire pulse heights, we only use the strip based measurements to determine the energy associated with a given CES cluster. Examples of the CES response functions

$\mathcal{P}_{\gamma}^{CES}(E^{CES}|p_{\gamma})$ for isolated photons with energies $5 < p_{\gamma} < 6$ GeV and $25 < p_{\gamma} < 26$ GeV are shown in Figs. 7(a) and (b), respectively.

4. Computation of the PPFA Likelihood

As mentioned earlier, the initial particle hypothesis assumes that each reconstructed track is a charged pion and each reconstructed CES cluster not associated with a track is a photon (or perhaps two merged photons, which makes little difference). Tracking momentum measurement is taken to be exact due to the superior resolution of the CDF tracker. To include calorimeter measurements, the highest p_T track in a tau candidate is extrapolated to the CES radius and the corresponding calorimeter tower becomes a seed tower. A grid of 3x3 towers is formed around the seed tower. Each track and CES cluster is associated to one tower on the grid. Each electromagnetic tower provides its own measurement $\vec{E}_{meas}^{EM}$ (components of this vector will be denoted as $E_{meas}^{EM_j}$ $j = 1, ..., 9$) used in the likelihood. For the hadronic calorimeter, we sum energies of all nine towers into a single measurement, $E_{meas}^{HAD} = \sum E_{meas}^{HAD_j}$, for the entire "super-cluster". In assumption that decay products of tau are charged tracks and photons, the likelihood function has the following form:

$$
\begin{aligned}
\mathcal{L}_p(\vec{p}_{\pi}, \vec{p}_{\gamma}, \vec{p}_n | \vec{E}_{meas}^{EM}, E_{meas}^{HAD}, \vec{E}_{meas}^{CES}) = \int \prod_{i=1}^{N_{\gamma}} dE_{\gamma_i}^{HAD} \prod_{k=1}^{N_{\pi}} dE_{\pi_k}^{HAD} \prod_{l=1}^{N_n} dE_{n_l}^{HAD} \times \\
\prod_{j=1}^{9} dE_{\gamma_i}^{EM_j} dE_{\pi_k}^{EM_j} dE_{n_l}^{EM_j} \times \delta(\sum_{i=1}^{N_{\gamma}} E_{\gamma_i}^{HAD} + \sum_{k=1}^{N_{\pi}} E_{\pi_k}^{HAD} + \sum_{k=1}^{N_n} E_{\pi_n}^{HAD} - E_{meas}^{HAD}) \\
\times \delta(\sum_{i=1}^{N_{\gamma}} E_{\gamma_i}^{EM_j} + \sum_{k=1}^{N_{\pi}} E_{\pi_k}^{EM_j} + \sum_{k=1}^{N_n} E_{\pi_n}^{EM} - E_{meas}^{EM_j}) \times \mathcal{P}_{\gamma}^{CAL}(E_{\gamma_i}^{EM_j}, E_{\gamma_i}^{HAD} | p_{\gamma_i}) \\
\times \mathcal{P}_{\pi}^{CAL}(E_{\pi_k}^{EM_j}, E_{\pi_k}^{HAD} | p_{\pi_k}) \times \mathcal{P}_{\gamma}^{CES}(E_{\gamma_i\ meas}^{CES} | p_{\gamma_i}), \quad (4.9)
\end{aligned}
$$

where the integration runs over all possible depositions of energy by each individual particle in each available calorimeter measurement, the delta functions in the second line ensure that the sum of the deposits for each measurement is equal to the observed value, and the third line includes response functions for photons and charged pions in the calorimeter and in the CES detector. One can choose to convert Eq.(4.9) into a posterior probability distribution to estimate the hadronic tau jet energy as:

$$\begin{aligned}\mathcal{L}_E(E_{jet}|\vec{E}^{EM}_{meas}, E^{HAD}_{meas}, \vec{E}^{CES}_{meas}) = \int \mathcal{L}_p(\vec{p}_\pi, \vec{p}_\gamma|\vec{E}^{EM}_{meas}, E^{HAD}_{meas}, \vec{E}^{CES}_{meas}) \\ \times\delta(\sum_{i=1}^{N_\gamma} p_{\gamma_i} + \sum_{k=1}^{N_\pi} p_{\pi_k} + \sum_{l=1}^{N_n} p_{n_l} - E_{jet}) d\vec{p}_\pi d\vec{p}_\gamma d\vec{p}_n\end{aligned} \tag{4.10}$$

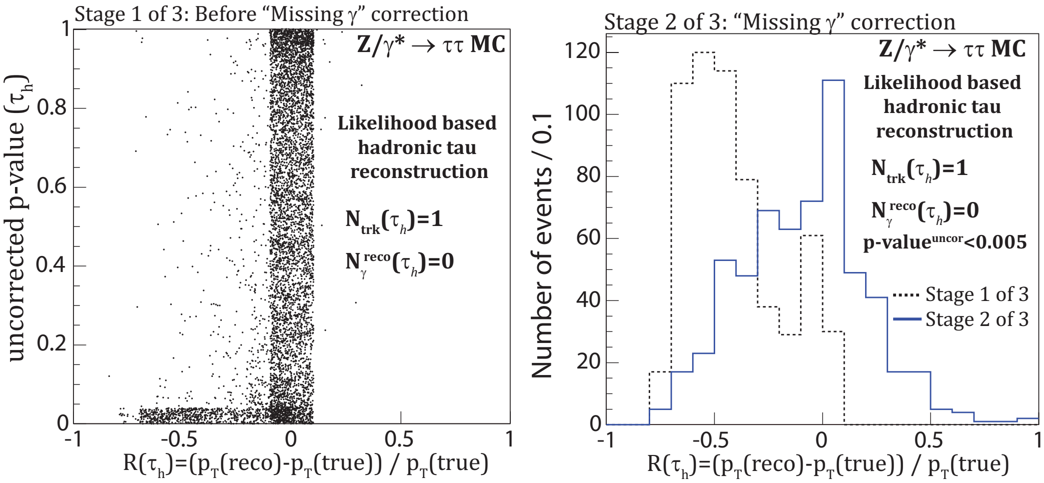

Fig. 8. Examples of $\mathcal{L}_E(E_{jet})$ for two representative simulated $Z \to \tau\tau$ events.

While the integral form presented in Eqs.(4.9,4.10) appears fairly complicated, it is straightforward to implement in the code and compute numerically using the Monte Carlo integration technique. Values of $\vec{p}_{\pi_k}$ and $\vec{p}_{\gamma_i}$ which maximize $\mathcal{L}(\vec{p}_\pi, \vec{p}_\gamma)$ in Eq.(4.9) represent the best estimate on energies of particles in tau decay in the assumption that the initial hypothesis on the particle content was correct. Figure 8 shows examples of the $\mathcal{L}_E(E_{jet})$ distributions for two representative events from a sample of simulated $Z \rightarrow \tau\tau$ events.

5. The Reduced p-Value Definition

Photon reconstruction failures or presence of a stable neutral hadron, e.g. K_L^0, may lead to an incorrect initial particle hypothesis making it important to detect and correct such cases. We define a p-value using Eq.(4.8), but we do two simplifications to the definition of the likelihood $\mathcal{L}_p$ in Eq. 4.9 to speed up the calculations. First, because in practice most of the cases affected by the incorrect initial hypothesis can be identified through inconsistencies between the available calorimeter and tracker measurements, we drop the terms associated with the CES in Eq. (4.9). Second, we combine the nine electromagnetic towers in the hadronic tau cluster into a single "super-tower" with energy $E^{EM} = \sum E^{EM_j}$, and define the "reduced" version of Eq. (4.9):

$$\mathcal{L}'_p(\vec{p}\,|(E^{EM}_{meas}, E^{HAD}_{meas}) = \int \prod_{j=1}^{9} dE^{EM_i} \times \mathcal{L}_p(\vec{p}\,|\vec{E}^{EM}_{meas}, E^{HAD}_{meas}) \times \delta(\sum_{i=1}^{9} E^{EM_i}_{meas} - E^{EM}_{meas}) \qquad (4.11)$$

We then define the "reduced" p-value according to Eq. (4.8) using the reduced $\mathcal{L}'_p$. The p-value evaluates how frequently a set of particles with true momenta $\vec{p}^0$ can produce a set of measurement equally or less probable than the one observed in data. The p-value is sensitive to inconsistencies in the available calorimeter measurements and can be used to detect mistakes in the initial particle content hypothesis. Figure 9(a) shows the distribution of the reduced p-value for all reconstructed hadronic tau jets in the sample of simulated $Z \rightarrow \tau\tau$ events as a function of the relative difference between the reconstructed visible tau jet energy at the maximum of the likelihood function and the true visible jet energy obtained at the particle generator level. It is evident that a vast majority of mismeasured jets have very low reduced p-value. As it will be shown next, most of these mismeasurements owe to the incorrect initial particle hypothesis.

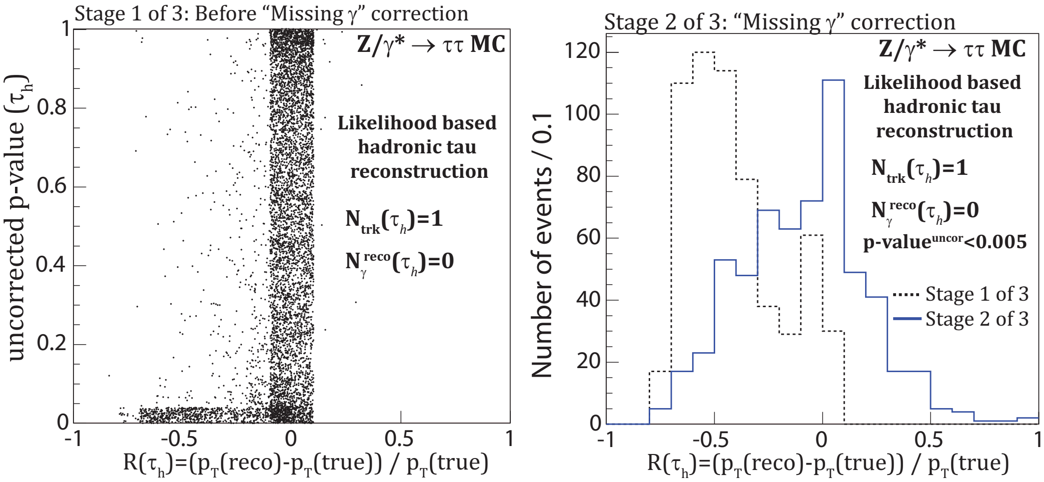

Fig. 9. $Z \rightarrow \tau\tau$ events in CDF II detector simulation: 1-prong taus with no photon candidate reconstructed by CES. Left: $p - value$ versus $R(\tau_h)$. Right: $R(\tau_h)$ for events with small p-value before correction (dashed black line) and after correction for missing photon (full blue line).

6. Corrections to the Particle Content Hypothesis

Based on our studies with the simulation, the majority of mismeasurements owing to the incorrect initial particle hypothesis fall into only two categories. The first category includes tau jets with one charged pion and typically one π^0 where none of the photons were reconstructed in the CES. This can happen for one of the following three reasons: (i) a simple CES reconstruction failure (either dead channels or a photon landing outside the fiducial volume of CES, (ii) the CES cluster is vetoed due to being too close to the extrapolated track position, or (iii) photon(s) falling into the cracks between the calorimeter ϕ-wedges. The last case is likely impossible to correct as the deposited electromagnetic energy is highly sensitive to small differences in the electromagnetic shower development. In addition photons hitting the cracks may deposit a substantial portion of their energy in the hadron calorimeter. All three cases lead to a substantial underestimation of the tau jet energy as only momentum of the track would count towards the measurement. To correct for this effect, we apply the following procedure: if a tau candidate with a single reconstructed track and no reconstructed photons has too small reduced p-value ($p < 0.005$), we first attempt to correct it by introducing an additional photon. Because no CES measurement is available for this photon, the term with $\mathcal{P}^{CES}$ in Eq.(4.9) is removed and the likelihood function with modified particle hypothesis $\mathcal{L}(p_\pi, p_\gamma)$ (or the corresponding $\mathcal{L}_E$) is recalculated. The new energy is taken as the updated energy of the tau jet. Figure 9(b) shows the relative difference of reconstructed and true jet energy for these jets before and after the correction. While the improvement is evident, the catastrophic cases where photons hit the cracks between the calorimeter wedges cannot be fully recovered contributing to reduced resolution. Another contribution to broadening of the distribution after correction comes from events in the second category which is discussed next.

Tau jets with one charged hadron and a stable neutral hadron (kaon), which is always omitted in the initial particle content hypothesis, typically have an excess of energy measured in the hadron calorimeter compared to what one would expect from a single charged pion. Because the excessive energy in the hadron calorimeter detected

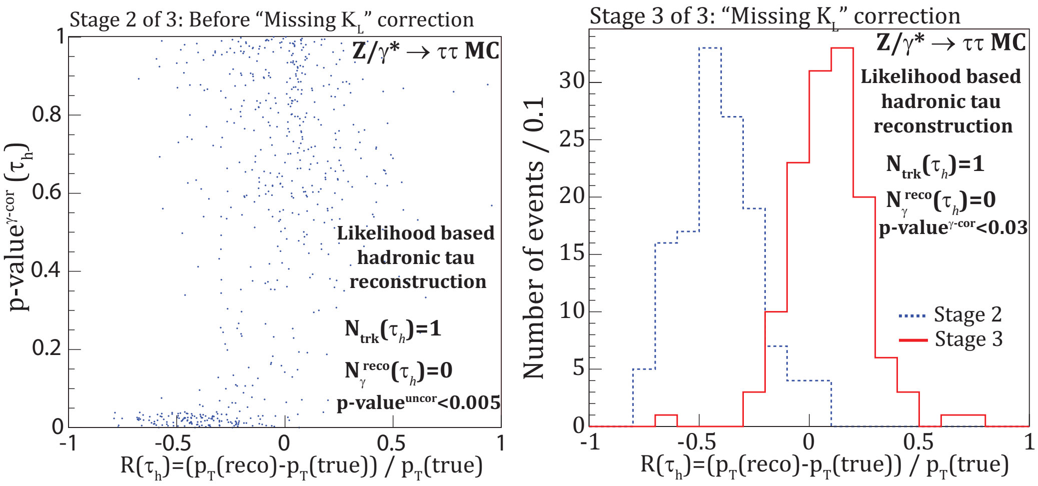

Fig. 10. $Z \to \tau\tau$ events in CDF II detector simulation: 1-prong taus with no photon candidate reconstructed by CES and $p - value^{uncor} < 0.005$. Left: $p - value$ after photon correction versus $R(\tau_h)$. Right: $R(\tau_h)$ for events with small $p - value^{\gamma - cor}$ before kaon correction (dashed blue line) and after correction for kaons (full red line).

using the p-value cannot be accounted for by adding a photon at the previous step, the p-value for these jets remains small after correcting the initial particle content hypothesis for a photon, as shown in Fig. 10(a). Therefore, for jets with exactly one reconstructed track and no reconstructed photons that had a low initial p-value ($p < 0.05$) and continued to have a low p-value after the photon correction (the threshold is $p < 0.03$), the particle content hypothesis is modified to contain one charged pion and one neutral kaon. Technically, it is accomplished by adding a term $\mathcal{P}_n^{CAL}(E^{EM}, E^{HAD}|p_n) = \mathcal{P}_\pi^{CAL}(E^{EM}, E^{HAD}|p_n)$ (as the calorimeter response for charged pions and neutral hadrons is very similar) in Eq.(4.9), and adjusting the argument of the delta-functions to include a new particle. The energy of the tau jet candidate is updated with the energy obtained from maximizing $\mathcal{L}_p(p_\pi, p_n)$ (or the corresponding $\mathcal{L}_E$). The relative difference of the reconstructed and true tau jet energy before and after the correction for this class of jets is shown in Fig. 10(b).

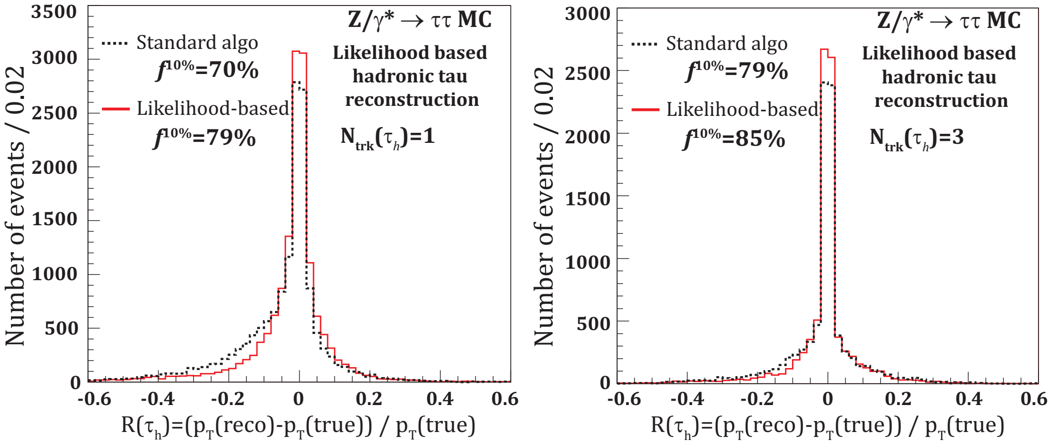

Fig. 11. Comparison between reconstructed transverse momentum and true transverse momentum of the hadronic tau for $Z \to \tau\tau$ events in CDF II detector simulation. The red solid line corresponds to the likelihood method, the black dashed line corresponds to standard CDF tau reconstruction.

7. PPFA Energy Resolution

Figures 11(a) and (b) show the relative difference between the PPFA reconstructed tau jet transverse momentum and the true visible transverse momentum obtained at generator level for one and three-prong hadronic tau jets. For comparison, the same plots show the performance of the standard CDF tau reconstruction (see [15] for details) shown as dashed line using the same simulated $Z \to \tau\tau$ events. It is evident that the PPFA algorithm has been able to converge to the correct energy without resorting to any kind of ad-hoc corrections used in the standard CDF reconstruction.

To quantify the level of improvement, we use the fraction of jets with the reconstructed energy falling within 10% of the true jet energy (denoted as $f^{10\%}$ in Fig. 11), on average the PPFA increases $f^{10\%}$ by about 10%. PPFA jet energy resolution distribution also has a more symmetric shape around the true energy and a reduced tail due to jets with underestimated reconstructed energy. More obvious improvements in the tail behavior for one-prong events owes to a larger fraction of jets containing neutral pions with significant contribution towards the total visible jet energy.

E. PPFA Performance in Data

To validate the performance of the algorithm in a real detector setting, we use a clean and well understood sample of CDF $Z \to \tau\tau$ events with one tau decaying to a lepton and another one decaying via one of the hadronic modes. We compare results in data with the simulation based expectations paying particular attention to events with large energy overlap and validate the p-value.

Note that unlike the case of $Z \to ee$ or $Z \to \mu\mu$ events where lepton momentum resolution can be inferred from the broadness of the dilepton mass spectrum, in the case of $Z \to \tau$ it is not possible due to the broad shape of di-tau mass owing to cancellation of the missing energy contributions associated with the neutrinos from tau decays. Therefore, to compare the performance of the PPFA based algorithm with the standard CDF reconstruction, we perform two side-by-side proto-analyses using similar selection that rely on discriminators provided by each of the two reconstruction methods. We start with a sample with loosened lepton isolations that has a

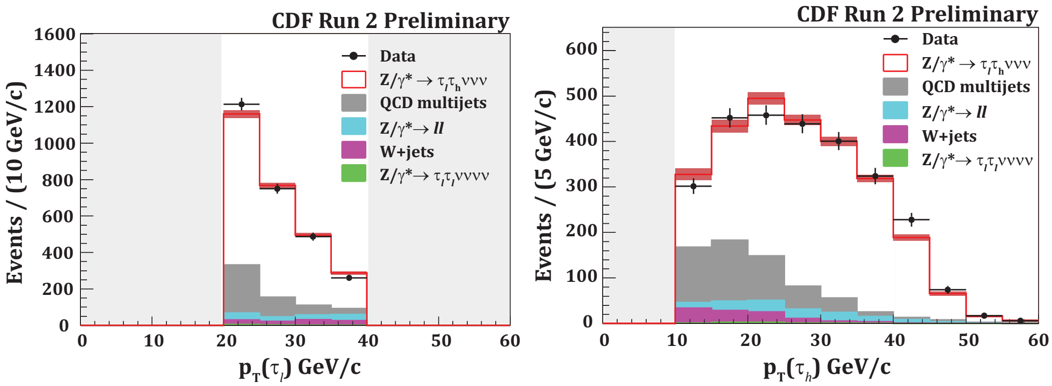

Fig. 12. Kinematic distributions demonstrating purity of the clean tau sample after $Z \to \tau\tau \to l\tau_h\nu\nu\bar{\nu}$ ($l = e$ or μ) events are extracted from CDF data with tight selection requirements: (a) transverse momentum of the light lepton, (b) transverse momentum of visible decay products of the hadronically decaying tau lepton, τ_h.

substantially higher level of QCD multi-jet background contamination and compare the signal to background ratio after applying selections exploring jet shape differences for hadronic taus and generic jets. The level of improvement, if present, could be used as a figure of merit in comparing PPFA performance with the standard method.

1. Validation of the PPFA Reconstruction Using $Z \to \tau\tau$

We use a fairly clean and very well understood sample of $Z \to \tau\tau$ events collected by CDF in Run-II in the channel where one tau decays hadronically ($\tau \to \tau_h\nu_\tau$) and the other decays to a light lepton (electron or muon). Selections require a tightly isolated reconstructed muon or an electron with $20 < p_T < 40GeV$ and a loose hadronic tau jet candidate. Tau jet is required to have a seed track with $p_T > 10$ GeV/c and no explicit requirement on the full momentum of the jet to exclude biases owing to the choice of a tau energy reconstruction algorithm. A set of event topology cuts are applied to reduce contamination due to cosmic rays, $Z/\gamma^* \to ee$, $Z/\gamma^* \to \mu\mu$ and W+jets events. A full list of selections is available in [28]. Remaining QCD multi-jet background is estimated from data using events with lepton and tau candidates having electric charge of the same sign. We rely on simulation to estimate $Z/\gamma^* \to \tau\tau$, $Z \to ee$, $Z \to \mu\mu$ and $W + jets$ contributions. These processes are generated using Pythia Tune A with CTEQ5L parton distribution functions [29] and the detector response is simulated using the GEANT-3 package [26].

Once the sample is selected, PPFA reconstruction is performed on data and simulation. Figures 12(a) and (b) shows lepton momentum and PPFA-based hadronic tau jet momentum distributions for the selected $Z/\gamma^* \to \tau\tau$ candidate events showing a good agreement between data and simulation as well as demonstrating the high purity of the sample. A thorough analysis of many other kinematic distributions has shown similar level of agreement.

Because a large fraction of one-prong hadronic taus decay via $\tau_h^\pm \to \nu_\tau \rho^\pm(770) \to \nu_\tau \pi^\pm$, the invariant mass of one prong tau jets allows evaluating the PPFA performance by comparing it to something known. Figures 13(a) and (b) show the invariant mass of the candidates for one prong tau candidates in data with the simulation predictions overlaid using the PPFA approach and the standard CDF tau reconstruction. Figures 13(c) and (d) show the distribution of the PPFA p-value for selected hadronic tau candidates with one or three charged tracks, also showing a good agreement. In addition, it is evident that the p-value itself can provide additional discriminating power against the jets from multi-jet QCD events and could be used in physics analyses to improve purity of the selected data.

2. New PPFA Tools and Physics Analyses

Table I. Selections used in the two proto-analyses using either PPFA or standard CDF selection for hadronically decaying tau jets. The first group of selections corresponds to standard CDF selections applied first in both analyses. The second group shows additional non-standard selections using the invariant mass and the narrowness of the tau candidate's jet cluster that can be applied to both analyses. The last selection uses the PPFA p-value and is only applied to the PPFA proto-analysis.

Selections	1-prong	3-prong
Kinematical selections:		
p_T GeV/c	> 10	> 15
Standard CDF ID selections:		
$m(\tau)$ GeV/c^2	< 2.5	< 2.5
Additional ID selections applied to both algorithms:		
$m(\tau)$	(0:0.25) or (0.375:1.4)	(0.8: 1.4)
$\Delta\theta$ rad	< 0.04	< 0.015
Additional PPFA-specific selections (only applicable to PPFA analysis):		
p-value	$pv > 0.008$ if $p_T < 20$ GeV/c	$pv > 0.06$ if $p_T < 30$ GeV/c

The new handles available in PPFA, such as the p-value or the jet-by-jet energy measurement uncertainty, can be used to improve sensitivity of physics analyses. They allow categorizing events with the different signal to background ratio or used as discriminating variables, either directly or as inputs to a multi-variate discriminators based on neural nets or decision tree techniques. Additional advantages of the

PPFA technique come from a better measurement of the momentum spectra and an improved measurement of jet shapes, which both allow better discrimination against jet backgrounds. To illustrate the PPFA potential in a real data analysis setting, we model two simple proto-analyses aiming to maximize the signal to background ratio for the $Z \to \tau\tau$ events by only exploiting the properties of tau jet candidates. One of the analyses only relies on variables calculable using standard CDF reconstruction and the other one relies on PPFA calculations (including the new p value variable). Both analyses start with a sample of candidate $Z \to \tau\tau$ events with a more typical for physics analyses level of background contamination due to the QCD multi-jet events[5]. Compared to the high purity sample, the "realistic" sample is obtained by loosening isolation and some other tight quality requirements on the lepton leg and removing the requirement on the absence of additional energetic jets in the event. The composition of this sample can be inferred from Fig. 14 showing several kinematic and jet shape variables. Note that the pedestal near $m = 0.14$ GeV/c^2 in the invariant mass plot is due to the one prong tau candidates that contain no reconstructed particles other than a single charged pion.

Table I shows the specific selections applied. Similar to the previously discussed invariant mass of the jet $m(\tau)$, $\Delta\theta(\tau)$ is calculated using the momenta of individual particles reconstructed in a jet and is the weighted angular width of the jet. It is defined as $\Delta\theta(\tau) = \frac{\sum E_i \times \theta_i}{\sum E_i}$, where the summation goes over particles in the jet, E_i being the particle energy and θ_i is the angle between the particle and the visible 4-momentum of the tau jet. To achieve a better signal to background ratio, we apply more elaborate selections for $m(\tau)$ and $\Delta\theta(\tau)$ than those customarily used at CDF.

---

[5] The clean sample used so far was obtained with the sole purpose of achieving a high purity sample of hadronic taus and features extremely tight selections applied to the lepton leg, which are not typical of physics analyses aiming to maximize sensitivity to a particular process of interest.

The specific cut values are picked by looking at distributions and are not unreasonable, but not necessarily optimal. These selections can be applied to both the standard and the PPFA-based analyses. Finally, we apply an additional p_T-dependent cut on the p-value. Because there is no equivalent of the p-value in the standard reconstruction, this selection can only be applied to the PPFA-based analysis. The distributions for these variables using PPFA definitions are illustrated in Figs. 14(b), (c) and (d). For the PPFA-based case, the effect of these selections is illustrated in Fig. 15(a) and (b) showing the "before" and "after" distributions for $m(l, \tau, \not\!\!E_{\mathrm{T}})$, the visible mass of lepton, tau and missing transverse energy, a quantity frequently used as the final discriminant in physics analyses [15, 16].

To obtain a quantitative figure of merit for the comparison of the two techniques, we calculate the signal to background ratio for $Z \rightarrow \tau\tau$ candidate events above a certain threshold in the visible mass, $m(l, \tau \not\!\!E_{\mathrm{T}}) > m_0$, as a function of m_0. Note that backgrounds are heavily dominated by the QCD multi-jet events. The resultant S/B ratio is shown in Fig. 15 for three cases: (i) the standard tau reconstruction with standard tau selections, (ii) the standard tau reconstruction with the new $m(\tau)$ and $\Delta\theta(\tau)$ selections, and (iii) PPFA based reconstruction with the new $m(\tau)$ and $\Delta\theta(\tau)$ selections and also the p-value-based selections. It is evident that the use of the PPFA tools provides a factor of 1.7 improvement in the signal to background ratio. It is expected that a more comprehensive utilization of the PPFA capabilities based on categorization of events using the jet energy uncertainty and the p-value can provide even larger improvement as compared to this simple example with box-like cuts.

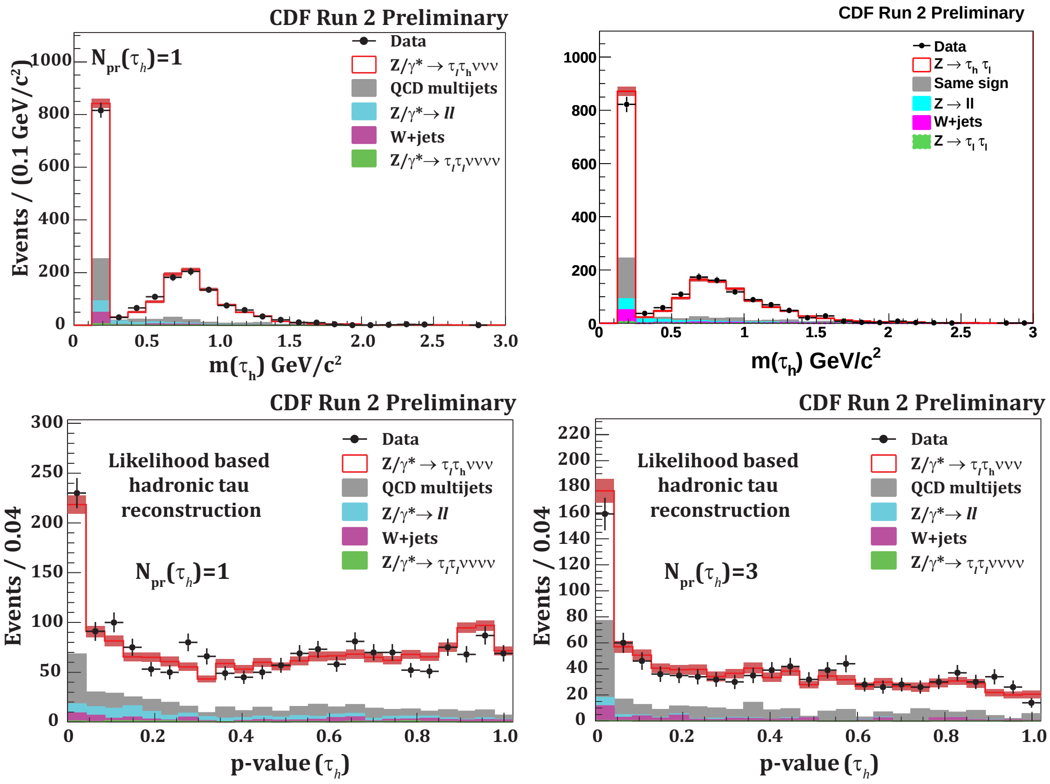

Fig. 13. Distribution of the invariant mass for reconstructed hadronic tau candidates and for events with a significant energy overlap using the clean tau sample after $Z \to \tau\tau \to l\tau_h\nu\nu\bar{\nu}$ ($l = e$ or μ). PPFA reconstruction(a) and standard CDF reconstruction(b). (c) and (d): Distribution of hadronic tau candidate p-value for events passing selection requirements in data (points) compared to the sum of background and signal predictions. Left : 1-prong taus. Right: 3-prong taus.

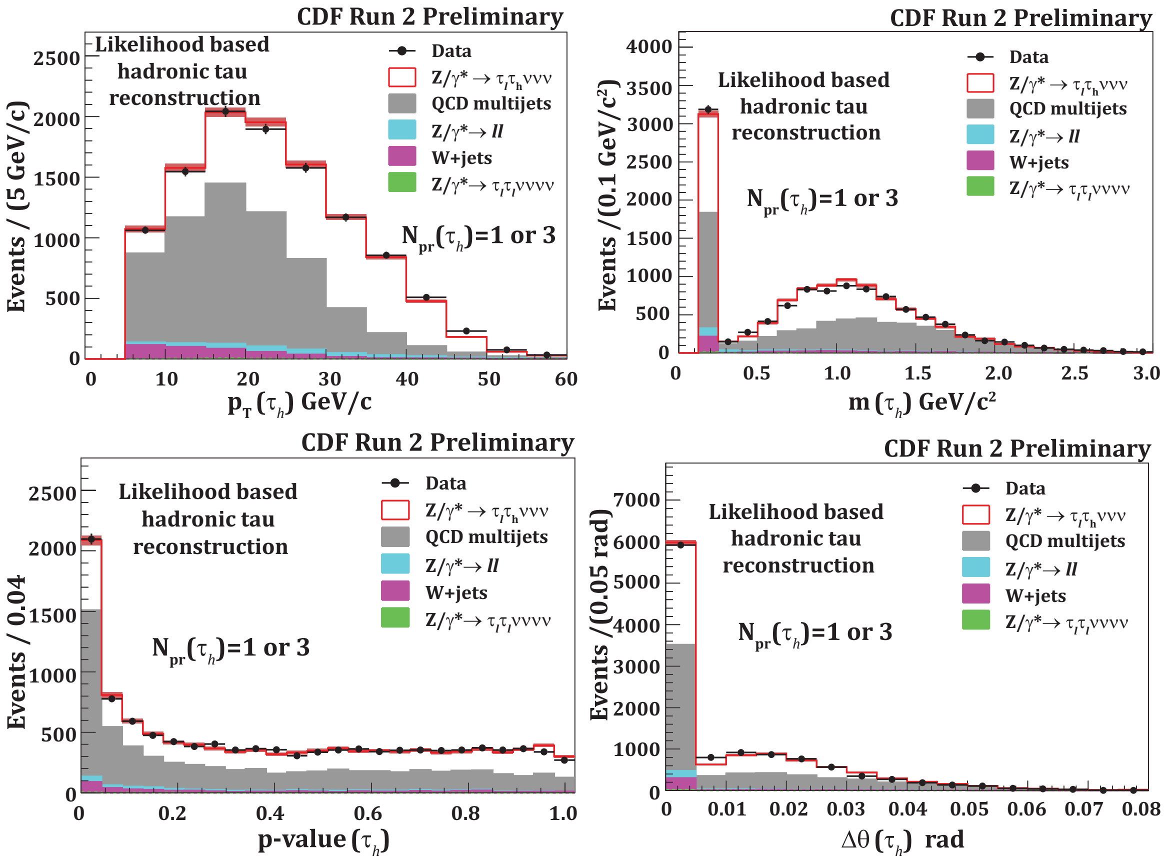

Fig. 14. 1- and 3-prong taus in the QCD enriched sample. Data (points) compared to the sum of background and signal predictions: (a) transverse momentum of visible decay products (b) hadronic tau visible invariant mass, (c) p-value distribution, (d) $\Delta\theta(\tau)$ distribution.

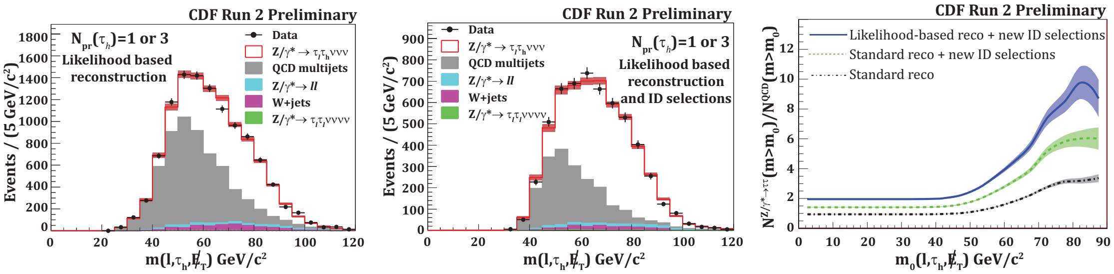

Fig. 15. Visible mass, $M(\tau, l, \not{E}_T)$. Events with 1 and 3-prong taus. Data (points) compared to the sum of background and signal predictions. (a): QCD enriched sample before likelihood-based identification requirements, (b): QCD enriched sample after likelihood-based identification is applied; (c) S/B ratio as a function of minimal threshold on $M(\tau, l, \not{E}_T)$. Black dash-dotted line corresponds to the standard CDF tau identification. Green dashed line - improved standard CDF tau identification. Blue full line - likelihood-based tau identification.

CHAPTER V

MISSING MASS CALCULATOR, A NEW TECHNIQUE FOR FULL MASS RECONSTRUCTION OF THE RESONANCES DECAYING TO $\tau\tau$

Invariant mass reconstruction is commonly used in experimental searches for new physics, such as for the Higgs or Z$'$ bosons, as well as in measurements of properties of known resonances. This technique is relatively straightforward for e^+e^-, $\mu^+\mu^-$, or di-jet final states. The accuracy of mass reconstruction in these channels is dominated by the detector resolution for lepton or jet momenta. The sensitivity of "mass bump-hunting" analyses depends critically on how narrow the signal invariant mass distribution is compared to the (usually broad) distributions in background processes. Unfortunately, this simple strategy is much less effective in searches for resonances decaying to a pair of τ leptons because the τ lepton energy associated with neutrinos escapes detection, and only visible products (leptons in the case of leptonic τ decays or low multiplicity jets in the case of hadronic τ decays) are observed in the detector.

Each τ lepton decay involves one or two neutrinos, depending on the final state: hadronic ($\tau \to \nu_\tau + hadrons$) or leptonic ($\tau \to \nu_\tau + l\bar{\nu}_l$, where l=e or μ). In pp or $p\bar{p}$ collisions, the full energy of neutrinos cannot be determined. Instead, one can only reconstruct a transverse energy imbalance in the detector (or missing transverse energy, $\not\!\!E_{\rm T}$), which is representative of the total transverse momentum of all neutrinos in the event. Therefore, when two or more neutrinos are produced in the same event, their individual transverse momenta and directions cannot be reconstructed. The situation in decays of heavy resonances into two τ leptons is even more complex. In these events, the two τ's are often produced "back-to-back" and the missing momentum associated with their neutrinos partially cancels out. As a result, the invariant mass of a resonance cannot be directly reconstructed from the $\not\!\!E_{\rm T}$ and visible decay products of τ leptons. Various techniques exist to partially reconstruct the mass of resonances in $\tau\tau$ final states. However, the reconstructed mass distributions for signal processes are rather broad (with long tails and typical core resolutions on the order of ∼20%), which makes it difficult to separate them from the background and considerably reduces the signal significance. This poses a major challenge for the Higgs boson searches in the $H \to \tau\tau$ channel, one of the most important channels for discovering a low-mass Higgs boson at the LHC [30, 31], whether in the context of the Standard Model or beyond (for example, in supersymmetric models). Another challenge in searching for a low-mass Higgs boson in the $\tau\tau$ channel is the large and irreducible background from $Z/\gamma^* \to \tau\tau$ events. This is because the Z/γ^* background is several orders of magnitude larger than any expected Higgs signal, and its broad partially reconstructed mass distribution completely dominates the signal region (for example, see Fig. 16 or reference [16]). Therefore, a major improvement in $\tau\tau$ invariant mass, $M_{\tau\tau}$, reconstruction techniques is needed in order to significantly enhance the sensitivity of $H \to \tau\tau$ searches at the Tevatron and LHC experiments.

We propose a new method[1], which substantially improves the accuracy of the $\tau\tau$ invariant mass reconstruction. We expect it will lead to a major improvement in the sensitivity of the Higgs boson searches in the $H \to \tau\tau$ channel at the Tevatron and LHC. In the next section, we briefly review currently used methods. Section B describes the new technique and illustrates its performance using a Monte Carlo simulation with a realistic detector resolution. In Sec. C, we report the results of tests on a clean sample of data $Z/\gamma^* \to \tau\tau$ events collected by the CDF experiment at the Tevatron.

---

[1] The material presented in this chapter is taken from Ref [28] where I am the author together with P. Murat, A. Pranko and A. Safonov. The permision from Elsevier is granted.

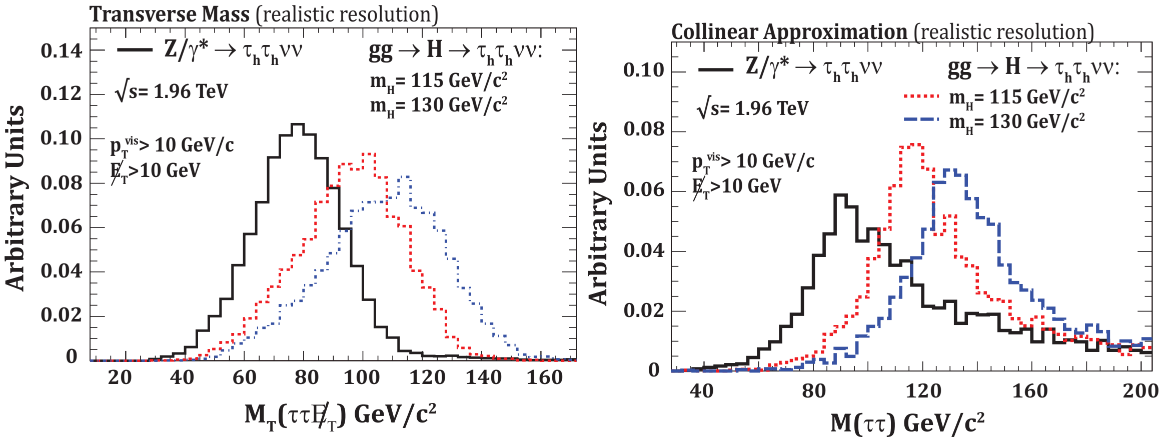

Fig. 16. Example of the transverse mass (left plot) defined as an invariant mass of $\not{E}_T$ and visible τ decay products, and the fully reconstructed mass (right plot) using the collinear approximation for three event samples: inclusive $Z/\gamma^* \to \tau\tau$ and $gg \to H \to \tau\tau$ with M_H=115 and 130 GeV/c^2. Results are obtained for the fully hadronic $\tau\tau$ decay mode. Events are simulated with a realistic detector resolution. All distributions are normalized to unit area.

A. Existing Methods for $\tau\tau$ Mass Reconstruction

The two methods frequently used at hadron colliders either rely on reconstructing a partial invariant mass or use the collinear approximation. In this section, we review these techniques and discuss their advantages and shortcomings.

1. The Transverse Mass Method

Neutrinos from the τ decays escape detection and make it impossible to determine the 4-momenta of τ leptons and thus $M_{\tau\tau}$. Therefore, one of the simplest and frequently used methods relies on a partial (or reduced) invariant mass reconstruction. Examples include either the invariant mass of visible decay products of the two τ leptons, the visible mass, or the invariant mass of the visible decay products and $\not\!\!E_{\mathrm{T}}$ in the event, the transverse mass. The latter is defined as follows:

$$M^2(\tau_{\mathrm{vis}_1}, \tau_{\mathrm{vis}_2}, \not\!\!E_{\mathrm{T}}) = P^{\mu} P_{\mu}, \tag{5.1}$$

$$P^{\mu} = P^{\mu}(\tau_{\mathrm{vis}_1}) + P^{\mu}(\tau_{\mathrm{vis}_2}) + P^{\mu}(\not\!\!E_{\mathrm{T}}),$$

where $P^{\mu}(\not\!\!E_{\mathrm{T}})$=$(\sqrt{\not\!\!E_{\mathrm{T}}{}_x^2 + \not\!\!E_{\mathrm{T}}{}_y^2}, \not\!\!E_{\mathrm{T}}{}_x, \not\!\!E_{\mathrm{T}}{}_y, 0)$ is a 4-momentum corresponding to missing transverse energy and $P^{\mu}(\tau_{\mathrm{vis}_1})$ and $P^{\mu}(\tau_{\mathrm{vis}_2})$ are the 4-momenta of the visible τ decay products. The transverse mass provides a somewhat better separation from the QCD multi-jet backgrounds with fake τ signatures than the visible mass, and it is often preferred in data analyses. The advantage of this technique is that the partial mass can be defined for all signal events, thus preserving the statistical power of the available data. However, ignoring or not fully accounting for the neutrino momenta biases and broadens the reconstructed $M_{\tau\tau}$ distributions, and therefore leads to a significantly reduced sensitivity in searches and measurements. This problem is particularly prominent in the low-mass $H \to \tau\tau$ search, where the signal cannot be separated from the much larger and very broad $Z \to \tau\tau$ background. To illustrate this we simulated inclusive $Z/\gamma^* \to \tau\tau$ and $gg \to H \to \tau\tau$ events produced in $p\bar{p}$ collisions at $\sqrt{s}$=1.96 TeV. Events are simulated with Pythia [27] supplemented with the TAUOLA package [32] for τ decays. We select events with both τ's decaying hadronically, $p_T(\tau_{\mathrm{vis}_{1,2}}) > 10$ GeV/c^2 and $\not\!\!E_{\mathrm{T}} > 10$ GeV. We assume 10% resolution for hadronic τ-jets and 5 GeV resolution for x- and y- components of $\vec{\not\!\!E}_{\mathrm{T}}$ (the realistic detector resolution is further discussed in Sec. 3). The left plot in Fig. 16 shows the transverse mass $M(\tau_{\mathrm{vis}_1}, \tau_{\mathrm{vis}_2}, \not\!\!E_{\mathrm{T}})$ distribution for the simulated events.

2. Collinear Approximation Technique

The collinear approximation is another frequently used technique [30, 31]. This method was first proposed in reference [33] to reconstruct the invariant mass in $\tau\tau$ decays of a Higgs boson produced in association with an energetic jet. It is based on two important assumptions: that the τ and all decay products are nearly collinear (i.e., $\phi_\nu \simeq \phi_{\rm vis}$ and $\theta_\nu \simeq \theta_{\rm vis}$); and that the $\not\!\!E_{\rm T}$ in the event is due only to neutrinos. In this case, the total invisible momentum carried away by neutrinos in each τ decay can be estimated by solving two equations:

$$\begin{aligned} \not\!\!E_{{\rm T}x} &= p_{{\rm mis}_1} \sin\theta_{{\rm vis}_1} \cos\phi_{{\rm vis}_1} + p_{{\rm mis}_2} \sin\theta_{{\rm vis}_2} \cos\phi_{{\rm vis}_2} \\ \not\!\!E_{{\rm T}y} &= p_{{\rm mis}_1} \sin\theta_{{\rm vis}_1} \sin\phi_{{\rm vis}_1} + p_{{\rm mis}_2} \sin\theta_{{\rm vis}_2} \sin\phi_{{\rm vis}_2}, \end{aligned} \tag{5.2}$$

where $\not\!\!E_{{\rm T}x}$ and $\not\!\!E_{{\rm T}y}$ are the x- and y-components of the $\not\!\!E_{\rm T}$ vector, $p_{{\rm mis}_1}$ and $p_{{\rm mis}_2}$ are the combined invisible momenta (there can be two ν's in a τ decay) of each τ decay, and $\theta_{{\rm vis}_{1,2}}$ and $\phi_{{\rm vis}_{1,2}}$ are the polar and azimuthal angles of the visible products of each τ decay. Then, the invariant mass of the $\tau\tau$-system can be calculated as $M_{\tau\tau} = m_{\rm vis}/\sqrt{x_1 x_2}$, where $m_{\rm vis}$ is the invariant mass of visible τ decay products, and $x_{1,2} = p_{{\rm vis}_{1,2}}/(p_{{\rm vis}_{1,2}} + p_{{\rm mis}_{1,2}})$ are momentum fractions carried away by visible τ decay products. Despite offering the great advantage of a fully reconstructed $\tau\tau$ mass ($M_{\tau\tau}$) instead of a partial visible mass, the collinear approximation still has significant shortcomings. The technique suits well for events where the $\tau\tau$ system is boosted, i.e., produced in association with a large E_T jet, and the visible τ decay products are not back-to-back in the plane transverse to the beam line. The last requirement is needed, because the system of Eqs. 5.2 becomes degenerate[2] if $\phi_{{\rm vis}_1} = \phi_{{\rm vis}_2} + \pi$ and solutions

---

[2] The system of Eqs. 5.2 is also degenerate when $\phi_{{\rm vis}_1} = \phi_{{\rm vis}_2}$, which is almost never achieved experimentally.

$p_{\mathrm{mis}_{1,2}} \sim \sin^{-1}(\phi_{\mathrm{vis}_1} - \phi_{\mathrm{vis}_2})$ diverge as $|\phi_{\mathrm{vis}_1} - \phi_{\mathrm{vis}_2}| \to \pi$. Unfortunately, the majority of $gg \to H \to \tau\tau$ events are produced with τ leptons in nearly the back-to-back topology. Therefore, this technique is applicable only to a relatively small fraction of such events. The collinear approximation is also very sensitive to the $\not{E}_{\mathrm{T}}$ resolution and tends to over-estimate the $\tau\tau$ mass, leading to long tails in the reconstructed mass distribution. This effect is especially undesirable for low-mass Higgs boson searches, where the tails of a much larger $Z \to \tau\tau$ background completely overwhelm the expected Higgs peak region. The right plot in Fig. 16 shows the distribution of $M_{\tau\tau}$ reconstructed with the collinear approximation for simulated inclusive $Z/\gamma^* \to \tau\tau$ and $gg \to H \to \tau\tau$ events with the same selection requirements as in the example of the transverse mass distribution.

B. The Missing Mass Calculator Technique

The new technique proposed in this paper, the Missing Mass Calculator (MMC) method, allows for a complete reconstruction of event kinematics in the $\tau\tau$ final states with significantly improved invariant mass and neutrino momentum resolutions. The MMC technique does not suffer from the limitations of the collinear approximation described in the previous section and can be applied not only to the events where $\tau\tau$ system is boosted but also to the events with back-to-back topologies without sacrificing the reconstructed mass resolution.

1. The Concept and Method Description

To facilitate the description of the method, we begin with assuming a perfect detector resolution and that there are no other sources for transverse missing energy apart from the neutrinos from the τ lepton decays. Under these assumptions, full reconstruction of the event topology requires solving for 6 to 8 unknowns: x-, y-, and z-components of the invisible momentum carried away by neutrino(s) for each of the two τ leptons in the event, and, if one or both τ's decay leptonically, the invariant mass of the neutrinos from each leptonic τ decay. However, there are only 4 equations connecting these unknowns:

$$\begin{aligned}
\not{E}_{\mathrm{T}x} &= p_{\mathrm{mis}_1} \sin\theta_{\mathrm{mis}_1} \cos\phi_{\mathrm{mis}_1} + p_{\mathrm{mis}_2} \sin\theta_{\mathrm{mis}_2} \cos\phi_{\mathrm{mis}_2} \\
\not{E}_{\mathrm{T}y} &= p_{\mathrm{mis}_1} \sin\theta_{\mathrm{mis}_1} \sin\phi_{\mathrm{mis}_1} + p_{\mathrm{mis}_2} \sin\theta_{\mathrm{mis}_2} \sin\phi_{\mathrm{mis}_2} \\
M_{\tau_1}^2 &= m_{\mathrm{mis}_1}^2 + m_{\mathrm{vis}_1}^2 + 2\sqrt{p_{\mathrm{vis}_1}^2 + m_{\mathrm{vis}_1}^2}\sqrt{p_{\mathrm{mis}_1}^2 + m_{\mathrm{mis}_1}^2} \\
&\qquad - 2p_{\mathrm{vis}_1} p_{\mathrm{mis}_1} \cos\Delta\theta_{vm_1} \\
M_{\tau_2}^2 &= m_{\mathrm{mis}_2}^2 + m_{\mathrm{vis}_2}^2 + 2\sqrt{p_{\mathrm{vis}_2}^2 + m_{\mathrm{vis}_2}^2}\sqrt{p_{\mathrm{mis}_2}^2 + m_{\mathrm{mis}_2}^2} \\
&\qquad - 2p_{\mathrm{vis}_2} p_{\mathrm{mis}_2} \cos\Delta\theta_{vm_2}
\end{aligned} \tag{5.3}$$

where $\not{E}_{\mathrm{T}x}$ and $\not{E}_{\mathrm{T}y}$ are the x- and y-components of the $\vec{\not{E}}_{\mathrm{T}}$ vector, $p_{\mathrm{vis}_{1,2}}$, $m_{\mathrm{vis}_{1,2}}$, $\theta_{\mathrm{vis}_{1,2}}$, $\phi_{\mathrm{vis}_{1,2}}$ are the momenta, invariant masses, polar and azimuthal angles of the visible τ decay products, and M_τ=1.777 GeV/c^2 is the τ lepton invariant mass. The rest of the variables constitute the "unknowns" which are the combined invisible ("missing") momenta $\vec{p}_{\mathrm{mis}_{1,2}}$ carried away by the neutrino (or neutrinos) for each of the two decaying τ leptons and the invariant mass of the neutrino(s) in the τ decay, $m_{\mathrm{mis}_{1,2}}$. Finally, $\Delta\theta_{vm_{1,2}}$ is the angle between the vectors $\vec{p}_{\mathrm{mis}}$ and $\vec{p}_{\mathrm{vis}}$ for each of the two τ leptons, and it can be expressed in terms of the other variables. For hadronic decays of τ's, the m_{mis} is set to 0 as there is only one neutrino involved in the decay. This reduces the number of unknowns.

The number of unknowns (from 6 to 8, depending on the number of leptonic τ decays) exceeds the number of constraints. Therefore, the available information is not sufficient to find the exact solution. However, not all solutions of this underconstrained system are equally likely, and additional knowledge of τ decay kinematics can be used to distinguish more likely solutions from less likely ones. An example of such additional information is the expected angular distance between the neutrino(s) and the visible decays products of the τ lepton. Figure 17 shows the distribution for the distance $\Delta R = \sqrt{(\eta_{\text{vis}} - \eta_{\text{mis}})^2 + (\phi_{\text{vis}} - \phi_{\text{mis}})^2}$ between the directions of visible and invisible (missing) decay products[3] for the three distinct τ decay types: leptonic, 1-prong hadronic and 3-prong hadronic. We incorporate this additional knowledge of decay kinematics as probability density functions in a suitably defined global event likelihood as discussed in what follows to provide additional constraints and to obtain a better estimator of $M_{\tau\tau}$.

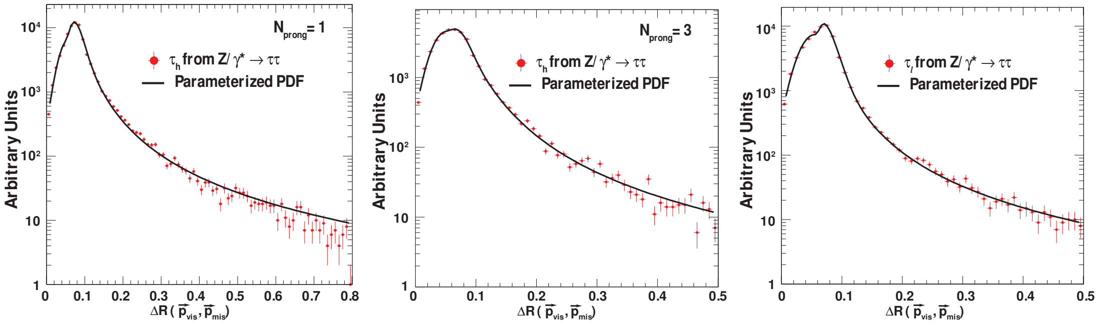

Fig. 17. Example of the probability distribution functions $\mathcal{P}(\Delta R, p_\tau)$ for a particular value of the original τ lepton momentum (p_τ). These functions are used in the calculation of the likelihood $\mathcal{L}$ for three cases: 1-prong τ (left plot), 3-prong τ (middle plot), and leptonic decays (right plot) of τ leptons. These distributions depend only on the decay type and initial momentum of the τ lepton.

---

[3] For simplicity, we use the ΔR parameterization, although a 3-dimensional angle between the decay products might be a more natural choice.

We first describe the method for the better constrained case, where both τ's decay hadronically, and then we explain how the machinery is adjusted for the case of leptonic decays. When both τ's decay hadronically, the system of Eqs. 5.3 can be solved exactly for any point in, for example, the (ϕ_{mis_1}, ϕ_{mis_2}) parameter space. For each point in that grid, the vectors $\vec{p}_{\mathrm{mis}_{1,2}}$ are fully defined and, therefore, one can calculate the distance $\Delta R_{1,2}$ between the vector $\vec{p}_{\mathrm{vis}_{1,2}}$ and the current assumed direction of $\vec{p}_{\mathrm{mis}_{1,2}}$. To evaluate the probability of such decay topology, we use ΔR distributions similar to those shown in Fig. 17, but we take into account the dependence of the distribution on the momentum of the initial τ lepton. If the τ lepton polarization is neglected, the ΔR distribution depends only on the τ momentum and decay type, but not on the source of τ's. Therefore, we use simulated $Z/\gamma^* \to \tau\tau$ events to obtain ΔR distributions for small bins (5 GeV/c) in the initial τ momentum, p_τ, in the range 10 GeV/c<p_τ<100 GeV/c (the range can be extended to both smaller and larger values). Events are simulated using Pythia supplemented with the TAUOLA package for τ decays. To simplify the calculations further, we parametrize the ΔR distributions by fitting them with a linear combination of Gaussian and Landau functions. Examples of such fits are shown as solid lines in Fig. 17. The p_τ-dependence of the mean, width and relative normalization of the Gaussian and Landau is then parametrized as $p_0/(p_\tau + p_1 p_\tau^2) + p_2 + p_3 p_\tau + p_4 p_\tau^2$ (where p_i are the coefficients of the parametrization), yielding fully parametrized distributions $\mathcal{P}(\Delta R, p_\tau)$, which can be used to evaluate the probability of a particular τ decay topology. To incorporate this information as an additional constraint, we define the logarithm of the event probability (or likelihood) as follows:

$$\mathcal{L} = -\log\left(\mathcal{P}(\Delta R_1, p_{\tau 1}) \times \mathcal{P}(\Delta R_2, p_{\tau 1})\right), \tag{5.4}$$

where functions $\mathcal{P}$ are chosen according to one of the decay types. To determine the best estimate for the $\tau\tau$ invariant mass in a given event, we produce an $M_{\tau\tau}$ distribution for all scanned points in the (ϕ_{mis_1}, ϕ_{mis_2}) grid weighed by a corresponding probability, $\mathcal{P}(\Delta R_1, p_{\tau 1}) \times \mathcal{P}(\Delta R_2, p_{\tau 1})$. The position of the maximum of the $M_{\tau\tau}$ distribution is used as the final estimator of $M_{\tau\tau}$ for a given event. An example of such a $M_{\tau\tau}$ histogram for randomly selected $H \to \tau\tau$ events of each category is shown in Fig. 18.

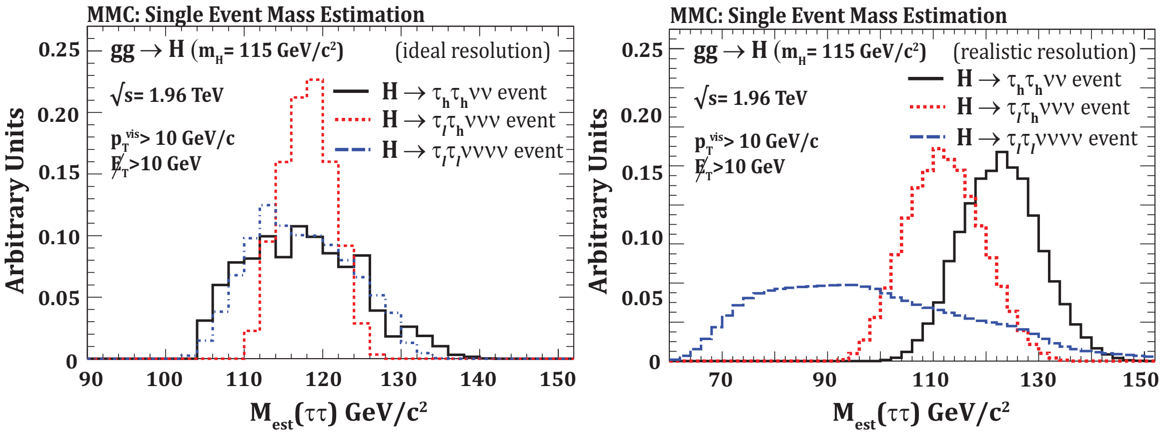

Fig. 18. Example of the $M_{\tau\tau}$ distribution filled for all grid points in one of the randomly selected $H \to \tau\tau$ events for each of the three decay modes. An entry for each point is weighted by its probability. The plot on the left shows $M_{\tau\tau}$ for the case of the ideal detector resolution and the plot on the right demonstrates $M_{\tau\tau}$ for the same three events in the case of a realistic detector resolution. All distributions are normalized to unit area.

For events where one or both τ leptons decay leptonically, the above procedure is adjusted to account for the unknown value of m_{mis} of the two neutrinos in each of the leptonically decaying τ's in the event. In this case, the scan is performed in a phase space of higher dimensionality: (ϕ_{mis_1},ϕ_{mis_2},m_{mis_1}), if only one of the two τ's decay leptonically; or (ϕ_{mis_1},ϕ_{mis_2},m_{mis_1},m_{mis_2}), if both decay τ's decay to leptons. As in the fully hadronic mode, one can unambiguously reconstruct the 4-momenta of both τ leptons for each point on the grid and calculate the event probability according to Eq. 5.4. For simplicity, we scan uniformly in the entire range of kinematically allowed values of m_{mis} assuming a flat probability function, but a scan performed according to the m_{mis} probability distribution function obtained from simulation may improve the algorithm performance.

2. Performance of the MMC Technique with Ideal Detector Resolution

To evaluate the performance of the MMC algorithm, we again use inclusive $Z/\gamma^* \to \tau\tau$ and $gg \to H \to \tau\tau$ (with M_H=115, 120, and 130 GeV/c^2) events produced by the Pythia MC generator supplemented with the TAUOLA package. All events are generated for $p\bar{p}$ collisions at $\sqrt{s}$ =1.96 TeV. However, the algorithm performance for events produced in pp collisions at the LHC is expected to be very similar to that for $\tau\tau$ events at the Tevatron. Unless it is otherwise noted, we select events where both visible τ leptons have p_T>10 GeV/c and $\not{E}_{\rm T}$>10 GeV ($\not{E}_{\rm T}$ is calculated as a combined transverse momentum of all neutrinos from both τ decays). The events are categorized according to the decay mode of each of the two τ leptons (leptonic, 1-prong or 3-prong hadronic), and the $\tau\tau$ mass is reconstructed using the appropriate version of the algorithm. Results for $H \to \tau\tau$ events with M_H =115 GeV/c^2 are shown in Fig. 19 for each of the three decay categories. In all cases, the peak position of the reconstructed $M_{\tau\tau}$ distribution is within ∼2% of the true mass, indicating that the assumptions used in the algorithm do not bias the reconstructed mass. The resolution of the reconstructed $M_{\tau\tau}$, defined as the RMS of the mass distribution in the range of $(1.0 \pm 0.4) \times M_{\tau\tau}^{true}$, changes from ∼8% for events with both τ leptons decaying hadronically to ∼13% when both τ leptons decay leptonically. The worse resolution in the leptonic modes is due to the weaker constraints on the system. The fraction of events where Eqs. 5.3 cannot be solved for any of the grid points ranges from ∼1% to 3%, which demonstrates the high reconstruction efficiency of the MMC algorithm. Figure 19 shows comparison of the reconstructed $\tau\tau$ mass in Z/γ^* and Higgs boson events with M_H =115 and 130 GeV/c^2 when both τ leptons decay hadronically.

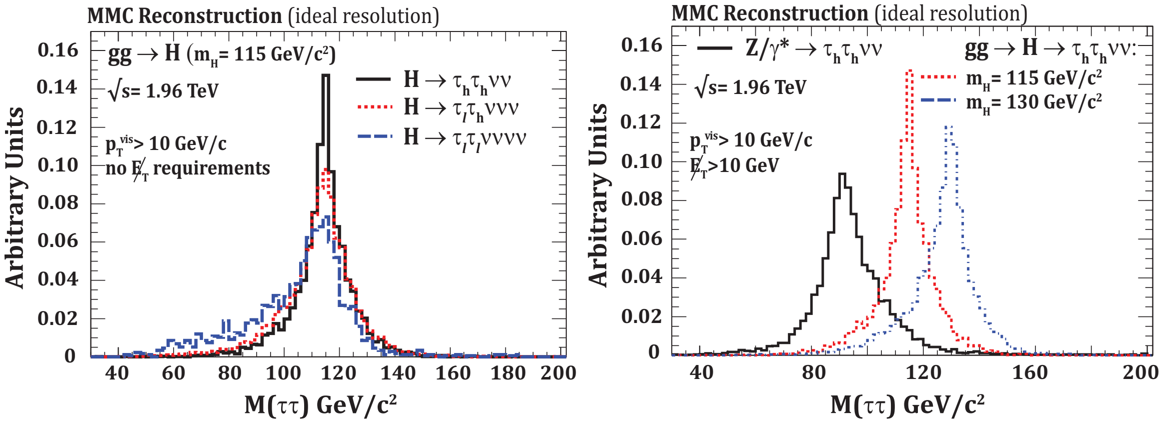

Fig. 19. Left plot demonstrates the reconstructed $M_{\tau\tau}$ in $H \rightarrow \tau\tau$ events with M_H =115 GeV/c^2 for each of the three decay categories: both τ's decay hadronically (solid line); one τ decays leptonically and the other one hadronically (dashed line); and both τ's decay leptonically (dashed-dotted line). Right plot shows the reconstructed mass in $Z/\gamma^* \rightarrow \tau\tau$ and $H \rightarrow \tau\tau$ events with M_H =115 and 130 GeV/c^2 in the fully hadronic decay mode. Results are obtained in the assumption of the ideal detector resolution. Each distribution is normalized to unit area.

3. Effects of Detector Resolution

To evaluate the importance of detector effects on MMC performance, we use the same inclusive $Z/\gamma^* \rightarrow \tau\tau$ and $gg \rightarrow H \rightarrow \tau\tau$ events and smear the $\not{E}_T$ and momenta of the visible τ decay products according to typical detector resolutions[4] at the LHC and Tevatron experiments [15, 25, 30, 31]. We assume 3% and 10% resolutions for momenta of light leptons and hadronic τ-jets, respectively. The $\not\!\!E_{\rm T}$ resolution for each of the two (x- and y-) components is taken to be σ_x=σ_y=σ=5 GeV [34]. Note that in a real experimental environment, the mismeasurements in lepton or hadronic τ-jet momenta also lead to an additional mismeasurement in $\not\!\!E_{\rm T}$. This effect is properly accounted for in our studies. Angular resolutions for visible τ decay products of typical detectors are usually accurate enough to have no noticeable effect on our calculations.

We find that mismeasurements of the momentum of τ lepton decay products alone have little effect on the performance of the algorithm. The $M_{\tau\tau}$ peak position and resolution are nearly unaffected and the efficiency is decreased by ~3-7% as a result of mismeasurements in the momenta of visible τ decay products, which are also propagated into $\not\!\!E_{\rm T}$. The stability of the peak position is related to a built-in self-correcting mechanism in the algorithm, which compensates slight under(over)-estimations in the measured momenta of visible decay products by over(under)-estimating the missing momentum, thus leading to the correctly reconstructed momentum of the original τ lepton.

One could expect the effects of finite $\not\!\!E_{\rm T}$ resolution to degrade the algorithm performance. We find that, if not taken into account, a 5 GeV resolution in $\not\!\!E_{\rm T}$ results in a 30-40% drop in reconstruction efficiency, long tails in the reconstructed $M_{\tau\tau}$ distribution, and a significant degradation in the $M_{\tau\tau}$ resolution (e.g., from ~8% to ~18% in fully hadronic $\tau\tau$ decay mode). In particular, a large reduction in the reconstruction efficiency occurs because mismeasurements in $\not\!\!E_{\rm T}$ break the key

---

[4] For simplicity, we assume Gaussian detector resolutions in this study.

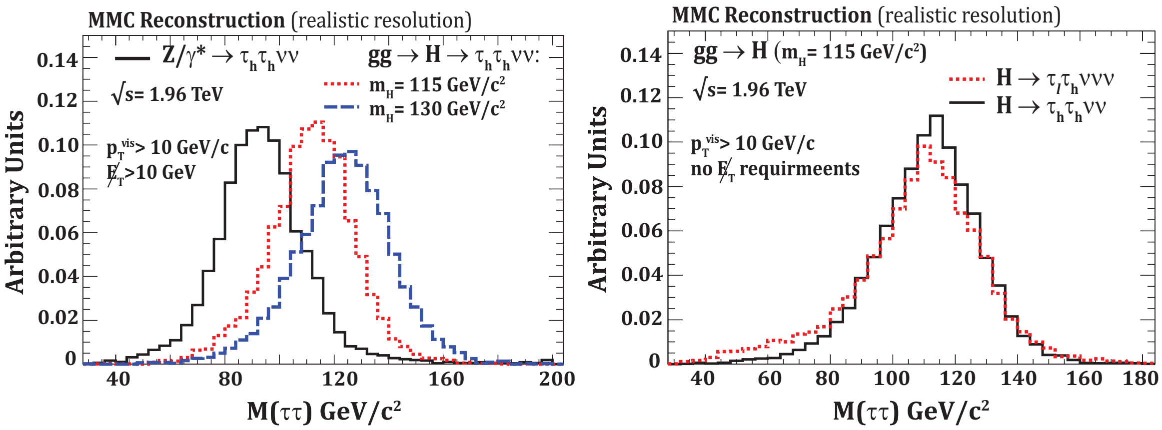

Fig. 20. Left plot illustrates the reconstructed $M_{\tau\tau}$ mass in $Z \rightarrow \tau\tau$ (solid line) and $H \rightarrow \tau\tau$ events with M_H =115 (dashed line) and 130 GeV/c^2 (dashed-dotted line) in the fully hadronic decay mode. Right plot demonstrates a comparison of the reconstructed mass in $H \rightarrow \tau\tau$ events with M_H =115 GeV/c^2 when both τ's decay hadronically (solid line) and when one τ decays leptonically and the other one hadronically (dashed line). All results are obtained by using the MMC technique in events simulated with a realistic detector resolution. Each distribution is normalized to unit area.

assumption that the neutrinos from the τ decays are the sole source of $\not{E}_{\mathrm{T}}$ in the event (see Sec. 1 and Eqs. 5.3). To mitigate these effects, the implementation of the MMC technique in a realistic experimental environment has to be adjusted to allow for possible mismeasurements in $\not{E}_{\mathrm{T}}$. It is achieved by increasing the dimensionality of the parameter space in which the scanning is performed to include the two components of the $\not{E}_{\mathrm{T}}$ resolution (for $\not{E}_{\mathrm{T}x}$ and $\not{E}_{\mathrm{T}y}$). In this case, the event likelihood, $\mathcal{L}$, has to be augmented with the corresponding resolution functions:

$$\mathcal{L} = -\log\left(\mathcal{P}(\Delta R_1, p_{\tau 1}) \times \mathcal{P}(\Delta R_2, p_{\tau 2}) \times \mathcal{P}(\Delta\not{E}_{\mathrm{T}x}) \times \mathcal{P}(\Delta\not{E}_{\mathrm{T}y})\right), \tag{5.5}$$

where the probability functions $\mathcal{P}(\Delta\not{E}_{\mathrm{T}_x})$ and $\mathcal{P}(\Delta\not{E}_{\mathrm{T}_y})$ are defined as:

$$\mathcal{P}(\not{E}_{\mathrm{T}_{x,y}}) = \exp\left(-\frac{(\Delta\not{E}_{\mathrm{T}_{x,y}})^2}{2\sigma^2}\right) \tag{5.6}$$

where σ is the resolution (which we take to be 5 GeV) and $\Delta\not{E}_{\mathrm{T}_{x,y}}$ are the differences between measured values of x- or y- components of $\not{E}_{\mathrm{T}}$ and the values in the parameter space while scanning over $\not{E}_{\mathrm{T}_x}$ and $\not{E}_{\mathrm{T}_y}$. In a real experimental setup, the $\not{E}_{\mathrm{T}}$ uncertainty can be larger in a particular direction, for example, if there is an energetic jet. In such cases, the uncertainty in the jet energy measurement will increase the uncertainty in $\not{E}_{\mathrm{T}}$ in the direction of the jet. These effects can be accounted for by suitably defining the x- and y- directions on an event-by-event basis and by choosing the appropriate σ_x and σ_y, which will not be equal to each other in general.

We evaluate the performance of the modified algorithm (with the $\not{E}_{\mathrm{T}}$ resolution scan) using $Z \to \tau\tau$ and $H \to \tau\tau$ events smeared with a realistic detector resolution as described above. Figure 20 shows the distribution of the reconstructed $M_{\tau\tau}$ in the fully hadronic decay mode for three samples: inclusive $Z/\gamma^* \to \tau\tau$ and $gg \to H \to \tau\tau$ with M_H=115 and 130 GeV/c^2. Right plot in the same Fig. 20 demonstrates a comparison of the reconstructed mass in $H \to \tau\tau$ events with M_H=115 GeV/c^2 in the case when both τ's decay hadronically (solid line) and in the case when one τ decays leptonically and the other one hadronically (dashed line). We find that the modified MMC algorithm recovers almost all lost efficiency (to the level of 97-99%) and significantly improves the relative $M_{\tau\tau}$ resolution (to the level of ~14%). The reconstructed mass peak position for each of the resonances is consistent with the corresponding true mass. We also observe that the mass resolution somewhat improves (at the level of 1-2%) for events with higher $\not{E}_{\mathrm{T}}$ and/or higher p_T of visible decay products.

4. Comparisons with Existing Methods

Figure 21 shows the reconstructed $M_{\tau\tau}$ distributions in $H \to \tau\tau$ events with M_h=115 GeV/c^2 obtained by using the MMC algorithm (black histogram) and the collinear approximation (red line). The events are simulated with realistic detector resolution effects as described in the previous section and the same selection requirements are applied. Two categories of $\tau\tau$ events are considered: both τ leptons decay hadronically (left plot in Fig. 21), and one leptonic and one hadronic τ decay (right plot in Fig. 21). The difference in normalizations of the MMC and collinear approximation results reflects a higher efficiency (by a factor of ~1.7) of the MMC method. This is because the substantial fraction of events have a moderate $\not\!\!E_{\mathrm{T}}$ or approximately back-to-back topology and are non-reconstructible by the collinear approximation technique. This happens when small mismeasurements in $\not\!\!E_{\mathrm{T}}$ lead to configurations for which Eqs. 5.2 have no solution. In contrast, the MMC method resolves this problem and has an average efficiency of 97-99%. In addition to a better resolution in the core of the $M_{\tau\tau}$ distribution, an important feature of the MMC technique is the absence of the long tail toward higher masses present in the distribution obtained using the collinear approximation. This tail is associated with the events of approximately back-to-back topology, where the collinear approximation diverges as $\cos\Delta\phi \to 1$. ($\Delta\phi$ is the angle between two visible τ decay products in the plane transverse to the beam line.) The reason for this divergency is discussed in Sec. 2. The effect is illustrated in Fig. 22, which shows a comparison of the ratio of the reconstructed and true mass as a function of $\cos\Delta\phi$ for the two methods. In contrast to the collinear approximation, the absence of long tails toward large masses in the MMC technique presents a significant improvement for low-mass Higgs boson searches in the $H \to \tau\tau$ channel by significantly reducing a large $Z \to \tau\tau$ background, which would otherwise completely overwhelm the Higgs search region.

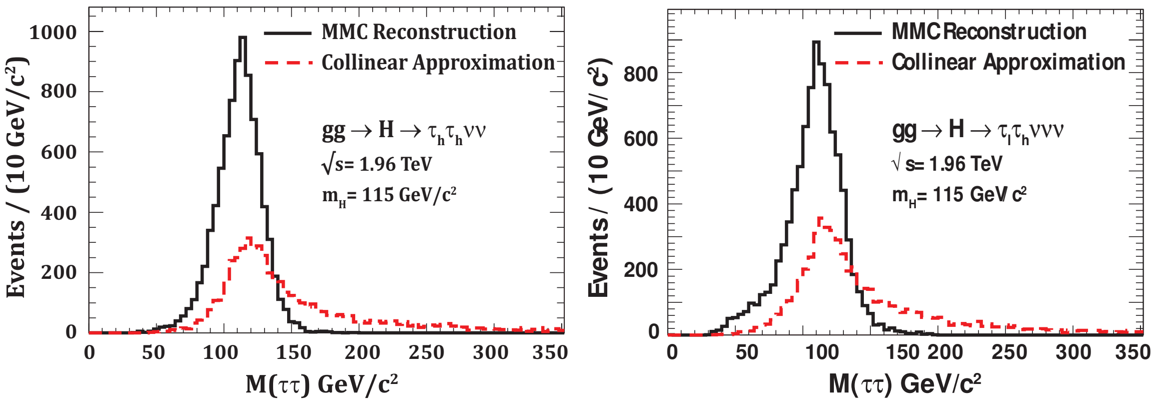

Fig. 21. Reconstructed mass of the $\tau\tau$ system for $gg \to H \to \tau\tau$ events with $M_H = 115$ GeV/c^2 simulated with realistic detector resolution effects. Results of the MMC technique (solid line) are compared to those based on the collinear approximation (dashed line). Two categories of $\tau\tau$ events are considered: when both τ leptons decay hadronically (left plot), and when one of the τ leptons decays to e or μ and the other τ decays hadronically (right plot). The difference in normalizations of the MMC and collinear approximation results reflects a higher efficiency of the MMC method. A long tail in the $M_{\tau\tau}$ distribution for the collinear approximation is due to the events where the two τ leptons have approximately back-to-back topology.

It is also important to point out that the algorithm efficiency and the shapes of likelihood $\mathcal{L}$ distributions are expected to be different for events with true τ leptons and those where jets are misidentified as hadronically decaying τ leptons. This may offer an additional handle on the backgrounds with the misidentified τ leptons, most notably W+jets and QCD multi-jet events, and it needs to be further investigated.

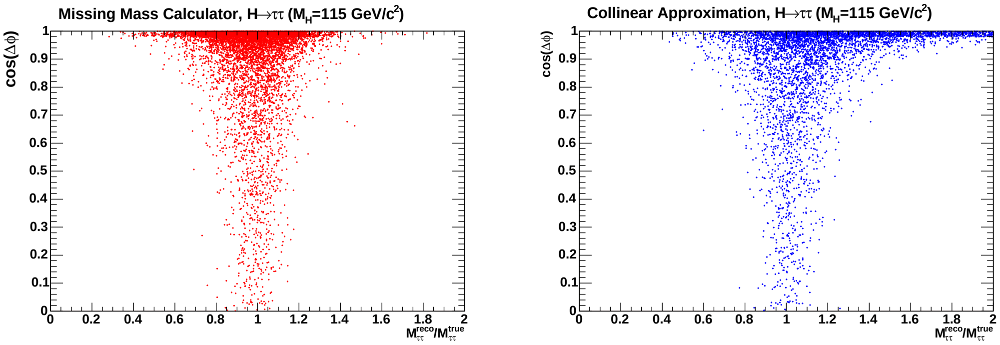

Fig. 22. Distribution of the ratio of the reconstructed invariant mass $M_{\tau\tau}$ versus $\cos\Delta\phi$, where $\Delta\phi$ is the azimuthal angle between visible decay products of the two τ leptons in $H \to \tau\tau$ events with M_h=115 GeV/c^2. Results of the MMC method (left plot) are compared to those of the collinear approximation (right plot). Note that the new method performs significantly better for nearly back-to-back topology ($\cos\Delta\phi \to$1), which constitutes the bulk of all $\tau\tau$ events.

C. Performance With Data and Monte Carlo After Full Detector Simulation

To illustrate the power of the proposed method using real data, we select a sample of clean $Z/\gamma^* \to \tau\tau$ events collected by the CDF experiment [35] in $p\bar{p}$ collisions at a center-of-mass energy $\sqrt{s}$=1.96 TeV at the Tevatron. We obtain a high purity sample of $\tau\tau$ events in the channel where one of the τ leptons decays into a light lepton (e or μ) while the other decays into one of the hadronic modes. The requirement of a well isolated muon or electron significantly reduces QCD multi-jet backgrounds in this channel. We then compare the observed $\tau\tau$ invariant mass spectrum of $Z/\gamma^* \to \tau\tau$ events reconstructed using the MMC technique with results obtained using the collinear approximation. Data are also compared with predictions obtained from

Monte Carlo (MC) simulation. Signal events and backgrounds coming from $Z/\gamma^* \rightarrow ee/\mu\mu$ and $W+jets$ processes are generated by Pythia Tune A with CTEQ5L parton distribution functions [29]. The detector response is simulated with the GEANT-3 package [26]. QCD multi-jet background is estimated from data by using events with lepton candidates of the same charge.

1. Data Selections

We ensure that the event selection is not sensitive to a potential Higgs boson signal to avoid biases in the $H \rightarrow \tau\tau$ analysis. This is achieved by using the same selections as in Chapter IV, which are extremely tight, thus effective for only a small fraction of $Z/\gamma^* \rightarrow \tau\tau$ and $H \rightarrow \tau\tau$ events. In fact, the signal acceptance is reduced by a factor of $\sim$3 compared to the search for $H \rightarrow \tau\tau$.

2. Mass Reconstruction using the MMC Technique

Although $Z/\gamma^* \rightarrow \tau\tau$ events in our data sample have no jets, a pair of τ leptons may be accompanied by one or more jets when different event selection requirements are applied. Therefore, we describe the $\not{E}_{\mathrm{T}}$ resolution parameterization for events with N_{jet}=0 and N_{jet}>0. For this purpose, we only count jets with E_T>15 GeV and $|\eta|$<3.6.

For N_{jet}=0 events, we perform scans for the x- and y-components of $\not{E}_{\mathrm{T}}$. The corresponding resolutions of each $\not{E}_{\mathrm{T}}$ component are parametrized by Gaussian distributions (Eq. 5.6) with width σ_{UE}, which is a function of unclustered energy[5] in the event: σ_{UE}=$p_0 + p_1\sqrt{\sum E_T}$. We use the same values of p_0 and p_1 as reported in the CDF publication [34].

---

[5] The unclustered energy is defined as the scalar sum of E_T for all calorimeter towers which are not included in electron, jet or hadronic τ reconstruction.

In events with $N_{jet}>0$, we consider the $\not\!\!E_{\mathrm{T}}$ resolution in the directions parallel ($\sigma_{\parallel}$) and perpendicular ($\sigma_{\perp}$) to the direction of a leading jet in the event. We take $\sigma_{\perp}=\sigma_{UE}$ and $\sigma_{\parallel}=\sqrt{\sigma_{UE}^2+\sigma_{jet}^2}$, where σ_{jet} is the jet energy resolution which is a function of the jet E_T and η. For σ_{jet}, we use a simplified version (assuming Gaussian jet energy resolution) of the parameterization reported in Ref. [34]. If there is more than one jet in the event, we project σ_{jet} for each additional jet onto axes parallel and perpendicular to the leading jet direction. These projections are then added in quadrature to $\sigma_{\parallel}$ and $\sigma_{\perp}$, respectively. Finally, we perform scans for $\not\!\!E_{\mathrm{T}}$ components parallel and perpendicular to the leading jet direction.

3. Reconstructed Mass Spectrum in Data

Figure 23 shows the $\tau\tau$ invariant mass distribution obtained with the MMC and collinear approximation methods for our data sample of $Z/\gamma^* \to \tau\tau$ events. The left plot shows the $\tau\tau$ mass calculated with the MMC technique and compares data with the sum of background and signal predictions. The first bin of the distribution contains events where no solution for $M_{\tau\tau}$ was found. We note that events unreconstructed by the MMC method are predominantly from background processes.

Excellent performance of the MMC technique and its advantage over the collinear approximation in terms of resolution and reconstruction efficiency is clearly demonstrated by differences in shape and normalization of the $M_{\tau\tau}$ distributions in the right plot of Fig. 23. To facilitate a comparison, the background predictions are subtracted from the $M_{\tau\tau}$ distributions in data. Events with the reconstructed mass $M_{\tau\tau}>160$ GeV/c^2 are outside the histogram range and are shown in the overflow bin. The fraction of such events is negligible ($\sim$0.3%) for the MMC method, while it is $\sim$18% for the collinear approximation. Shapes of the distributions agree well between data and simulation, therefore we use simulated $Z/\gamma^* \to \tau\tau$ events to estimate the

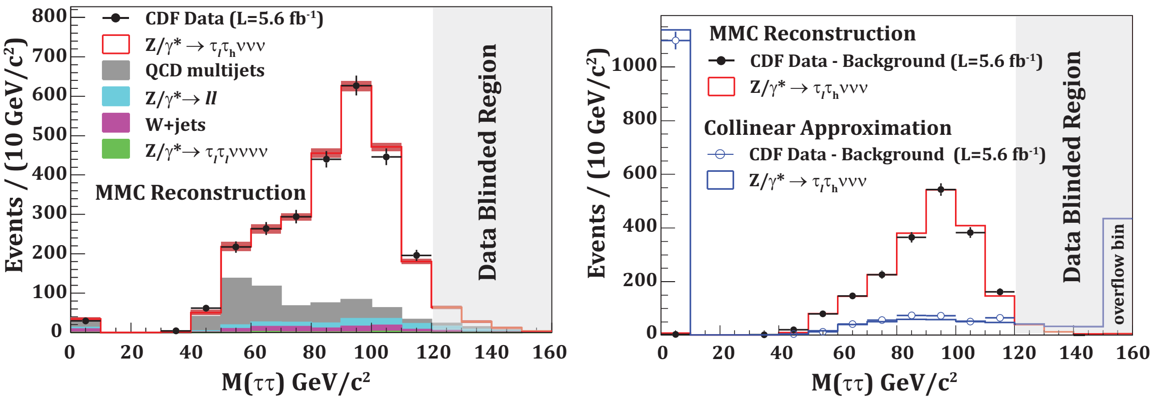

Fig. 23. Reconstructed mass of the $\tau\tau$ system in $Z/\gamma^* \to \tau\tau \to l\tau_h\nu\nu'\bar{\nu}$ ($l = e$ or μ) candidate events using the MMC and collinear approximation techniques: (a) $\tau\tau$ mass reconstructed with MMC technique, data (points) compared to the sum of background and signal predictions; (b) comparison of the MMC (filled circles are data and red line is the signal prediction) and collinear approximation (open circles are data and blue line is the signal prediction) results after subtracting the corresponding background predictions. Unreconstructed events are shown in the first histogram bin ($M_{\tau\tau}\sim 0$). Events with $M_{\tau\tau}>160$ GeV$/c^2$ are outside the histogram range and are shown in the overflow bin.

resolution and efficiency achieved by the MMC technique. We follow the definition introduced in Sec. B and define resolution as the RMS of the $M_{\tau\tau}/M_{\tau\tau}^{true}$ distribution in the 0.6-1.4 range, where $M_{\tau\tau}$ is the reconstructed mass and $M_{\tau\tau}^{true}$ is the generated mass. We find the resolution to be $\sim$16% and the reconstruction efficiency to be $\sim$99%, in good agreement with the results obtained using the simplified detector simulation model in Sec. 3. In contrast, the reconstruction efficiency of the collinear approximation method is found to be $\sim$42%. As explained in Sec. 3, events where the two τ leptons are back-to-back in the $x-y$ plane are particularly challenging for the collinear approximation; however, such events represent a major fraction of the $H \rightarrow \tau\tau$ signal. A reliable $\tau\tau$ mass reconstruction with the collinear approximation is possible only for a small fraction of boosted $\tau\tau$ events (with smaller angles between the τ's, $\Delta\phi(\tau\tau)$, and higher values of $\not{E}_{\mathrm{T}}$). The MMC method does not have such limitations, thus giving a substantial increase in the signal acceptance.

CHAPTER VI

DATA ANALYSIS

A. Introduction

In this chapter we describe the search for the Higgs boson using 7.8 fb^{-1} of proton-antiproton collisions collected by the CDF detector. The considered channel is $H \rightarrow \tau\tau$. The final states are categorized based on the flavour of the leptonic tau decay (electron or muon), type of the hadronic tau decay (1- or 3-prong) and the presense of an energetic jet (E_T>30 GeV).

1. Analysis Overview

We search for the Higgs boson decaying to a pair of tau leptons. Our search is sensitive to all four possible production mechanisms at the Tevatron: ggH, WH, ZH and VBF.

There are a number of final states in the $\tau\tau$ channel. Tau lepton is not a stable particle and it decays leptonically or hadronically. We use the following notations for tau decay modes: τ_e ($\tau \rightarrow e\nu_\tau\nu_e$) or τ_μ ($\tau \rightarrow \mu\nu_\tau\nu_\mu$) for leptonic decays and τ_h ($\tau \rightarrow N^{trk}\pi^{\pm}N^{neutr}\pi^0\nu_\tau$, where $N^{trk} = 1$ or 3 is the number of charged pions and $N^{neutr} \geq 0$ is the number of neutral pions) for hadronic decays. Table II shows the fraction of events for each $\tau\tau$ final state. While $\tau_h\tau_h$ has the largest branching fraction, it has overwhelming QCD background due to the large probability for a QCD jet to be misidentified as a τ_h candidate. Events where both tau leptons decay to electrons,$\tau_e\tau_e$, or to muons, $\tau_\mu\tau_\mu$ have relatively low contribution to the total number of events and suffer from large background due to Drell-Yan process [36]. Events with one tau decaying hadronically and another decaying to light leptons are the most valuable as they have large branching fraction and the presence of a well identified electron or muon allows significantly reduce QCD background. For this reason this analysis is optimized for this type of events. Yet efficient identification of hadronically decaying taus is important to minimize the rate of quark or gluon jets misidentified as hadronic tau jets.

Apart from the QCD multijet background, $Z \rightarrow \tau\tau$ events pose a challenge for $H \rightarrow \tau\tau$ search as Z boson has a large production rate and its mass is close to a possible mass of the Higgs boson. To separate Higgs signal from irreducible Z background we use a novel technique, Missing Mass Calculator (MMC), to reconstruct the full mass of $\tau\tau$ system [28].

Table II. $H \rightarrow \tau\tau$ final states

Final state	Fraction of events, %	Comment
$\tau_h\tau_h$	46	Large QCD multijet background.
$\tau_e\tau_e$	3	Large Drell-Yan background.
$\tau_\mu\tau_\mu$	3	Large Drell-Yan background.
$\tau_e\tau_\mu$	6	Clean signature. Small fraction.
$\tau_h\tau_e$	23	"Golden mode".
$\tau_h\tau_\mu$	23	"Golden mode".

To optimize sensitivity we split events into eight channels based on the tau decay type and presence of a jet in the event.

B. Event Selection

This analysis targets $\tau\tau$ events in the channel where one of the tau leptons decays leptonically and the other one decays hadronically. This particular mode has a high branching fraction $B(\tau\tau \to l\tau_h + \nu\nu\nu) \approx 46\%$ while the requirement of an isolated lepton allows suppressing the QCD multi-jet backgrounds to a reasonable level without large losses in signal efficiency.

The event selection requires presence of at least one light lepton (central electron, CMUP or CMX muon) with $p_T > 20$ GeV/c and at east one hadronic tau candidate with $p_T > 12.5$ GeV/c (for one prong tau candidates) or $p_T > 15$ GeV/c (for three-prong tau candidates). Both lepton and hadronic tau selections rely on fairly standard deifnitions providing high efficiency while delivering adequate suppression of QCD multi-jet backgrounds. After selecting the lepton candidates, a series of topology cuts are used to reduce contamination from Drell-Yan and W+jets processes as well as smaller diboson and $t\bar{t}$ contributions. After these selections, the dominant remaining contamination is due to the irreducible $Z \to \tau\tau$ background. Final discrimination between signal and $Z \to \tau\tau$ is performed by fitting the fully reconstructed di-tau resonance mass using the MMC technique searching for an enhancement of ditau pairs in the region above the $Z \to \tau\tau$ peak.

To further improve senstivity of the analysis, we define several sub-regions to better separate events with different signal to background ratio. We define four regions based on whether the light lepton is an electron or muon, and whether the tau candidate is one- or three-prong. Each of the four regions is further divided into two sub-regions based on the presence of at least one jet above a certain threshold. Analysis of the invariant mass spectrum is separately performed in each of the regions.

1. Selection of Electron Candidates

Electron identification and isolation allow a strong suppression of multi-jet events, which otherwise would dominate the sample of selected events. We require presence of at least one high quality electron candidate defined by closely following standard electron identification for CEM electrons. We only make a small modification of the isolation requirement: we do not apply calorimeter isolation and use track isolation instead. Electron ID requirements are listed in Table III.

Table III. Electron ID cuts

Cut
Track Quality 3×2×5
z_0<60 cm
E^{HAD}/E^{EM}<0.055 + 0.00045×E
E/p<2 or E>100 GeV
-3<$Q_e\Delta X_{CES}$<1.5 cm
$\|\Delta Z_{CES}\|$<3 cm
CES $\chi^2\|_Z$<10
L_{shr}<0.2
d_0<0.2 cm
$I_{trk}^{\Delta R<0.4}$<2.5 GeV/c if $\tau_h(1-prong)$
$I_{trk}^{\Delta R<0.4}$<2.0 GeV/c if $\tau_h(3-prong)$
E_T>20 GeV
Conversion veto

2. Selection of Muon Candidates

Muon identification and isolation requirements serve the same purpose as in the case of electrons. In the events collected using the inclusive muon trigger, we require presence of at least one high quality muon candidate. The specific selections are listed in Table IV, which again follow standard CDF muon identification.

Table IV. Muon ID cuts

Cut
CMUP or CMX
$\|\Delta x\|_{CMU}$<7 cm
$\|\Delta x\|_{CMP}$<5 cm
$\|\Delta x\|_{CMX}$<6 cm
Track Quality 3×2×5
d_0<0.2 cm if no SVX hits, 0.02 cm if SVX hits
$E^{EM}<$ 2.0 + max(0.0; 0.0115(p-100)) GeV
$E^{HAD}<$ 6.0 + max(0.0; 0.028(p-100)) GeV
$I_{trk}^{\Delta R<0.4}$<2.5 GeV/c if $\tau_h(1-prong)$
$I_{trk}^{\Delta R<0.4}$<2.0 GeV/c if $\tau_h(3-prong)$
p_T>20 GeV/c
Cosmic veto

3. Selection of Hadronically Decaying Tau Candidates

Once at least one high quality light lepton candidate is identified, the event is required to have at least one high quality hadronically decaying tau candidate not overlapping with at least one good light lepton candidate. This is equivalent to requiring a presence of at least one pair of high quality light lepton and hadronically decaying tau candidates in the event. Tau identification and isolation cuts used in defining a high quality candidate are listed in Table V

Table V. Tau ID cuts

Cut
$\lvert\eta\rvert<1$
Seed Track Quality 3×2×5
$p_T^{\tau-seed}>6$ GeV/c
$d_0^{\tau-seed}<0.2$ cm
$\lvert z_0^{\tau-seed}-z_0^{l-trk}\rvert<5$ cm
$9<\lvert z_{R_{CES}}^{\tau-seed}\rvert<230$ cm
$N_{trk}^{\Delta R<0.17}=1$ or 3
$N_{trk}^{0.17<\Delta R<0.52}=0$ (count tracks with $p_T>1$ GeV/c)
$I_{trk}^{0.17<\Delta R<0.52}<1$ GeV/c
$\frac{E^{EM}}{E^{EM}+E^{HAD}}<0.9$
$\frac{E^{calo}}{P^{trks}}>0.5$
$M^{trk}<1.8$ GeV/c^2
$0.6<M_{\tau_h}^{vis}(3-prong)<2.0$ GeV/c^2
$p_T^{vis}(1-prong)>12.5$ GeV/c
$p_T^{vis}(3-prong)>15$ GeV/c

4. Event Topology Cuts

Once events containing at least one pair of high quality light lepton and hadronically decaying tau candidates are selected, we apply a set of event topology selections. Given that the QCD multijet background is already significantly reduced at this point, these selections are primarily designed to reject background contamination owing to the $Z \to ll$ and W+jets processes. We reject events roughly consistent with $Z \to ee$ or $Z \to \mu\mu$ hypothesis and with the invariant mass close to the Z boson mass (see Table VI). We also apply a cut on mutual orientation of the two lepton candidates and $\not\!\!E_{\mathrm{T}}$ in the event (see Table VII), which is found to provide a very strong suppression against W+jets events while having very little inefficiency for events in signal topology (including $Z \to \tau\tau$ events as they have a similar topology to that of the signal events).

Table VI. Topological selections supressing Drell-Yan background (aka "$Z \to ee/Z \to \mu\mu$ veto").

	Calorimeter based ($\tau_e\tau_h$ only)
$\tau_e - \tau_h$ pair	$81 < M^{calo}(\tau_e, \tau_h) < 101$ GeV/c^2
τ_e plus second cluster	$76 < M(\tau_e, e) < 106$ GeV/c^2
Second cluster:	$E_T^{EM} > 8$ GeV, $E^{HAD}/E^{EM} < 0.12$
	Track based ($\tau_e\tau_h$ and $\tau_\mu\tau_h$)
τ_l plus second track	$66 < M^{trk}(\tau_l, trk) < 111$ GeV/c^2
Second track:	$p_T > 10$ GeV/c, $I_{trk}^{\Delta R<0.4} < 2$ GeV/c

Table VII. Topological selections suppressing $W + jets$ background events.

Cut
$\Delta\phi(e/\mu, \not\!\!E_{\rm T}) + \Delta\phi(\tau_h, \not\!\!E_{\rm T}) < 3.5$
$M_T(\tau_l, \not\!\!E_{\rm T}) < 65$ GeV/c^2

5. Definition of the Signal Regions

Once all lepton, tau and topological selections are applied, we define eight regions according to the flavor of the light lepton (electron or muon), the type of the hadronic tau candidate (one or three prong), and presence of at least one jet above a certain threshold. Such categorization separates events with different signal to background ratios improving the overall sensitivity of the analysis. The jet E_T threshold (E_T>30 GeV) is optimized to maximize the expected significance of the analysis using full fitting procedure.

C. Backgrounds

Here we explain our model used to predict backgrounds.

1. Overview

This analysis searches for excess of events with a large (essentially above the Z mass) invariant mass calculated using the leptonic and hadronic tau candidates. Background contributions owe to the following SM processes (in the order of significance) : $Z \to \tau\tau$ (the main irreducible background), QCD multijets and $W + jets$, $Z \to ll$, *di*-bosons,

$t\bar{t}$. The last three contributions are very small and have little influence on the results.

In evaluating background contributions, we rely on data driven approaches, either directly measuring backgrounds from data or, when the simulation is expected to be reliable, by defining scale factors between data and simulation. In the latter case we rely on a variety of measurements performed in the control regions enriched with events from one or more background processes (in which case we explore shape differences).

While the signal region definition requires the two lepton candidates to have opposite charge, events with the same electric charge provide an important control region for estimating some of the background contributions, most importantly the QCD multi-jet background. We therefore frequently compare the samples that only differ by the charge product of the two lepton candidates, which we refer to "opposite sign" (OS) events or "same sign" (SS) events. Generally, one can write the total rate of background events for the OS and SS samples with otherwise same selections as follows:

$$n_{os} = n_{os}^{Z\to\tau_h\tau_l} + n_{os}^{QCD} + n_{os}^{W+jets} + n_{os}^{Z\to ll} + n_{os}^{other} \tag{6.1}$$

$$n_{ss} = n_{os}^{Z\to\tau_h\tau_l} + n_{ss}QCD + n_{ss}^{W+jets} + n_{ss}^{Z\to ll} + n_{ss}^{other}, \tag{6.2}$$

where "other" refers to the combined sum of the "small backgrounds". For the QCD multi-jet backgrounds, $n_{os}(QCD) \approx n_{ss}(QCD)$, although the ratio is not exactly one to one[1]. Introducing $n_{os}(QCD) = (R_{os/ss}^{QCD} n_{ss}(QCD)$, one can rewrite Eq. (6.1) using

---

[1] Various measurements at CDF and CMS suggest that there slightly more (5–10%) opposite sign events depending on the exact selections. It is noteworthy that simulation, e.g. Pythia, suggests a similar size asymmetry. The effect is likely associated with $p\bar{p} \to q\bar{q}$ events in which final lepton candidates "remember" the charge of the parent parton. Typically, tighter isolation when the entire jet fluctuates into just a single track enhances this "memory".

Eq. (6.2) as:

$$n_{os} = n_{ss} + n_{os-ss}^{Z\to\tau_h\tau_l} + (R_{os/ss}^{QCD} - 1)n_{ss}^{QCD} + n_{os-ss}^{W+jets} + n_{os-ss}^{Z\to ll} + n_{os-ss}^{other}, \qquad (6.3)$$

where we use notation $n_{os-ss} = n_{os} - n_{ss}$.

Formula in Eq. (6.3) serves as a basis for the background estimation schema used in our analysis. Note that for many electroweak processes simulation predictions for n_{os-ss} are more robust as the difficult for simulation "fake rate" contributions leading to the SS events are already subtracted.

We use simulation to predict the shape and the rate of the small backgrounds, i.e. *di*-bosons and $t\bar{t}$. Upon validation, simulation is also used to obtain shapes for the background contributions due to $Z \to \tau_h\tau_l$, $W + jets$ and $Z \to ll$, but we rely on explicit comparisons of data and simulation in the control region to determine the absolute normalization.

Because the main signal region is split into sub-regions with and without jets, in some cases applying the above formalism directly for each sub-region can lead to large systematic uncertainties on the backgrounds due to the limited statistics of events in the control regions. Therefore, where appropriate, we first determine the absolute normalizations for the inclusive (with and without jets) samples and then introduce probabilities f_{0jets} and f_{1+jet}, which can be measured in data using other control regions. One example is the measurement of f_{1+jet} for $Z \to \tau_h\tau_l$ events, which can be accurately measured using $Z \to ee$ events selected with the "equivalent" kinematical selections.

2. Normalization of the W+jets MC sample.

To estimate n_{os-ss}^{W+jets} in each of the signal sub-regions, we first determine the normalization for the pairs of sub-regions with and without jets. For each of such "inclusive" sub-regions, we define control regions dominated by W+jets events with very small other background contributions. Selection cuts for these four control regions (from two light lepton types and two choices for the number of prongs in the hadronic tau candidate) are listed in Table VIII. Note that compared to the signal region, the control region uses the reverted $\delta\phi$ requirement (which depends on event kinematics and not on jet fragmentation details, and therefore is reliably predicted by simulation), while other selections are highly efficient for W+jet events therefore not introducing any unexpected biases.

Table VIII. $W + jets$ normalization region

Cut
$\Delta\phi(e/\mu, \not{E}_{\mathrm{T}}) + \Delta\phi(\tau_h, \not{E}_{\mathrm{T}}) > 3.5$
$\not{E}_{\mathrm{T}}$>30 GeV
$M_T(\tau_l, \not{E}_{\mathrm{T}})$>40 GeV/$c^2$

Note that the rate of W+jets events is not expected to be perfectly predicted by the simulation because this rate depends on the probability for the recoil jet to be misidentified as either one or three prong hadronic tau candidate, which are due to rare fluctuations in jet fragmentation. Similarly, the ratio of events with one and three-prong tau candidates is not to be trusted as the degree to which simulation can be incorrect in evaluating the probabilities of such fluctations can be (and porobably is) different. One also does not expect MC to perfectly model the OS/SS ratio, as those sub-samples have different fractions of quark and gluon recoil jets, which fragment differently. To avoid over-relying on simulation, we separately normalize opposite sign and same sign events using scale factors K_{os}^{W+jets} and K_{ss}^{W+jets}:

$$K_{os(ss)}^{W+jets} = \frac{n_{os(ss)}^{data,W+jets}}{n_{os(ss)}^{MC,W+jets}} \tag{6.4}$$

measured in the control regions for OS and SS events and corrected for (small) contamination from other backgrounds:

$$n_{os(ss)}^{data,W+jets} = n_{os(ss)} - n_{os(ss)}^{Z\to ll} - n_{os(ss)}^{other}. \tag{6.5}$$

The scale factor is defined relative to the simulation prediction for $n_{os/ss}^{MC}(W + jets)$, which includes normalization to luminosity and the NLO K-factor of 1.4.

Figures 24 and 25 show hadronic tau candidate p_T distribution for each of the sub-regions with the normalization correction applied. Table IX shows normalization factors obtained for each channel relative to the $W + jets$ MC predictions. As we mentioned earlier, these normalization factors are expected to be different due to the several separate effects that are likely not described well by for the simulation. However, one can use these normalization scale factors to construct “invariants” where possible “mistakes” by the simulation would cancel out. For example, the rato of scale factors for one and three-prong OS events should be the same for samples in the electron and muon channel. the same should be true for the SS events. Table X shows that these ratios (only the numbers in the same column are to be compared) agree within the uncertainties, which serves as a cross-check for the procedure. Table XI demonstrates further validation of the normalization using other similar invariants.

a. Jet Counting Scale Factors

To estimate the fraction of the events falling into the 0-jet and 1+-jet sub-regions, we note that these fractions are determined by the probability of producing an extra

Table IX. $K(W+jets) \pm \sigma_{K(W+jets)}(\epsilon_{K(W+jets)}(\%))$, N_{jet}>=0

	$\tau_h\tau_e$	$\tau_h\tau_\mu$
τ_h(1-prong) OS	0.590 ± 0.034 (5.8)	0.393 ± 0.027 (6.9)
τ_h(1-prong) SS	0.893 ± 0.010 (10.7)	0.530 ± 0.075 (14.2)
OS/SS	0.66 ± 0.08 (12)	0.74 ± 0.12 (16)
τ_h(3-prong) OS	0.668 ± 0.036 (5.4)	0.512 ± 0.033 (6.4)
τ_h(3-prong) SS	0.667 ± 0.057 (8.6)	0.481 ± 0.051 (10.6)
OS/SS	1.00 ± 0.08 (10)	1.07 ± 0.13 (12)

recoil jet in an event with a W and a tau-like jet. To measure that probability with acceptable accuracy, we define another control region for $W + jet$ events, which is different from the main $W + jet$ control region only by relaxing the isolation requirements on the τ_h side. This allows increasing the statistics of the data sample and the measured f's are listed in Table XII.

b. Final Scale Factors and the Uncertainties

Table XIII shows the uncertainty on the normalization of the $W + jet$ background contribution used in setting the final limits. The shape of the $W + jet$ events is taken from the simulation with relaxed cuts on the isolation of the hadronic tau leg to increase the statistics.

Table X. Validation of the normalization factors. There should be agreement between numbers within each column.

l	$\frac{K_{os}^{l-1prong}}{K_{os}^{l-3prong}}$	$\frac{K_{ss}^{l-1prong}}{K_{ss}^{l-3prong}}$
e	0.817892 ± 0.0730119 (8.92684)	1.17157 ± 0.177176 (15.123)
μ	0.665597 ± 0.0711544 (10.6903)	1.10688 ± 0.219597 (19.8393)

3. $Z \to ll$

With the background estimation schema used in our analysis, the Drell-Yan events present a fairly small background contribution. There are two categories of $Z \to ee$ and $Z \to \mu\mu$ events passing the selection requirements of this analysis. The first one is due to events $Z + (jet \to “\tau_h”)$, where one of the leptons in $ee/\mu\mu$ pair becomes the $\tau_{e\mu}$ candidate while the recoil jet fakes a τ_h candidate. The second lepton in the pair is typically lost as otherwise the event would not pass the $Z \to ee/Z \to \mu\mu$ veto. Due to smaller acceptance of the muon system, muons fail reconstruction more frequently than electrons and the relative fraction of events for this category is higher for the $Z \to \mu\mu$ samples. Note that because jet fakes have no charge correlation with the charge of the surviving lepton, the n_{os-ss} quantity used in our background estimation schema vanishes for this category of events. This category of background events is therefore automatically subtracted.

The second type of $Z \to ee$ and $Z \to \mu\mu$ events owes to events where one of the leptons is misidentified as a hadronic tau candidate, $e/\mu \to \tau_h$, either due to an overlap with another particle (e.g. a photon from the underlying event), an abnormally large

Table XI. Validation of the normalization factors.

X = (sign, n-prong)	$\frac{K^e_X}{K^\mu_X}$
os, 1-prong	1.55398 ± 0.161072 (10.3651)
ss, 1-prong	1.65606 ± 0.323628 (19.542)
os, 3-prong	1.26462 ± 0.117641 (9.30246)
ss, 3-prong	1.56463 ± 0.242599 (15.5052)

calorimeter energy deposition (for the muon case) or a large leakage into the hadron calorimeter (for an electron). Lepton and hadronic tau candidate in these events are typiclaly oppositely charged. We use simulation predictions corrected with a scale factor measured in a control region defined by inverting the Z veto requirement. For the electron and muon sub-regions, we additionally invert the requirement on the electromagnetic fraction of the selected hadronic tau candidate and the calorimeter to track energy ratio, respectively. Both control regions have high purity of $Z \to ee$ and $Z \to \mu\mu$ events as illustrated in Fig. 26 showing the invariant mass of the $l - \tau_h$ pair. Note that the lepton leg in these events that fakes a τ_h candidate is reconstructed using hadronic tau reconstruction sequence. in calculating the invariant mass, we use the 4-momenta of the tracks associated with the tau candidate to reconstruct the $\mu\mu$ invariant mass, for the ee case we use the 4-momentum of the calorimeter cluster associated with the tau candidate. The latter does not include any of the standard electron energy corrections leading to a small difference in resolution in data and simulation. This is expected and has no effect on the analysis. Table XIV shows

Table XII. Fraction of events in the =0-jet and >0-jet channels measured from $W + jets$ data.

channel	$F_{=0}$	$F_{>0}$
$\tau_e\tau_h$(1-prong) OS	0.93384 ± 0.0370581 (3.97)	0.0661597 ± 0.00732394 (11.07)
$\tau_e\tau_h$(1-prong) SS	0.909091 ± 0.0585655 (6.44)	0.0909091 ± 0.0139998 (15.40)
$\tau_e\tau_h$(3-prong) OS	0.926934 ± 0.0357697 (3.86)	0.0730659 ± 0.00749424 (10.26)
$\tau_e\tau_h$(3-prong) SS	0.924183 ± 0.0482138 (5.22)	0.075817 ± 0.0103258 (13.62)
$\tau_\mu\tau_h$(1-prong) OS	0.926279 ± 0.0393383 (4.25)	0.0737207 ± 0.00828564 (11.24)
$\tau_\mu\tau_h$(1-prong) SS	0.920705 ± 0.0624112 (6.78)	0.0792952 ± 0.0137298 (17.31)
$\tau_\mu\tau_h$(3-prong) OS	0.931571 ± 0.0374061 (4.02)	0.0684292 ± 0.00754003 (11.02)
$\tau_\mu\tau_h$(3-prong) SS	0.924266 ± 0.0524299 (5.67)	0.0757342 ± 0.0112214 (14.82)

normalization scale factors obtained for the inclusive jet $Z \to ll$ samples.

a. Jet Counting Scale Factors

To obtain the normalization scale corrections accounting for differences in data in simulation for the distribution of the inclusive events between the sub-regions with and without jets, we use the data from the same $Z \to ee$ control region to measure the fraction of events with and without jets, $f_{0\,jets}$ and $f_{1+\,jets}$ (see Table XV and Fig. 27). The same value (as the probability of observing a jet in Z events does not depend on the channel) is then used for each of the four channels as the scale correction for the normalization of events in pairs of the jet sub-regions.

b. Final Scale Factors and the Uncertainties

Table XVI shows the final scale factors and the associated uncertainties used in limit setting.

Table XIII. Final uncertainties on the total normalizations of the $W + jets$ add-on background.

$= 0$-jet	$\tau_h\tau_e$	$\tau_h\tau_\mu$
τ_h(1-prong)	12	13
τ_h(3-prong)	14	14

> 0-jet	$\tau_h\tau_e$	$\tau_h\tau_\mu$
τ_h(1-prong)	29	22
τ_h(3-prong)	25	26

4. Estimation of the $Z \rightarrow \tau_h\tau_l$ and QCD Multi-Jet Background Contaminations

To first order, QCD multi-jet contribution can be estmated using the rate and the shape of SS events in data. The more accurate calculation requires correcting the normalization to take into account that the ratio $R^{QCD}_{os/ss}$ is slightly deviated from unity. In the case of the inclusive $Z \rightarrow \tau\tau$ events, the shapes of various distributions are known to be well described by the simulation, but the overall normalization is not known due to fairly significant trigger inefficiencies and also smaller differences in the isolation efficiencies between data and simulation. With all other backgrounds accounted for, that leaves two unknown normalization scale factors (times the number of signal regions). To determine them, we use the distribution of the MMC-based invariant mass $M_{\tau\tau}$ shown in Fig. 28 for each of the four regions ($e\tau_h$ or $\mu\tau_h$ for one and 3-prong taus). While the contributions from $Z \rightarrow \tau\tau$ and QCD multi-jet events

Table XIV. $Z \rightarrow ll$ normalization factors with the absolute and relative uncertainties for the inclusive N_{jet} >=0 channel, $K(Z \rightarrow ll) \pm \sigma_{K(Z\rightarrow ll)}(\epsilon_{K(Z\rightarrow ll)}, \%)$

	ee	$\mu\mu$
τ_h(1-prong)	0.880 ± 0.005 (0.5)	0.887 ± 0.033 (3.7)
τ_h(3-prong)	0.763 ± 0.024 (3.1)	N/A

Table XV. Fraction of events in the = 0-jet and > 0-jet channels measured from $Z \rightarrow ee$ data.

$F_{=0}$	$F_{>0}$
0.983147 ± 0.00440547 (0.448099)	0.0168526 ± 0.000413018 (2.45076)

have been scaled in the plots to obtain reasonable agreement with the data, it is clear that the shapes of the two remaining backgrounds are very different, allowing one to fit this distribution for the two normalization scale factors. Given the apparently very good agreement of the data and expectations, an equivalent but simpler method is to divide the entire distribution into two sub-regions with $65<M(\tau\tau)<100$ GeV/c^2 and $0<M(\tau\tau)<65$ GeV/c^2, dominated by $Z \rightarrow \tau_h\tau_l$ and multi-jet events, respectively. We then count events in each of the sub-regions, and write down an equation similar to

Table XVI. Final uncertainties on the total normalizations of the $Z \to ll$ background.

$= 0$-jet	ee	$\mu\mu$
$e/\mu \to \tau_h$(1-prong)	0.7	3.7
$e/\mu \to \tau_h$(3-prong)	3.1	N/A

> 0-jet	ee	$\mu\mu$
$e/\mu \to \tau_h$(1-prong)	2.5	4.4
$e/\mu \to \tau_h$(3-prong)	4.0	N/A

Eq. (6.1 for each sub-regions.Then we solve the system of two equations for $K(Z \to \tau_h\tau_l)$ and $R^{QCD}_{os/ss}$, using a simple iterative procedure. Results are shown in Tables XVII and XVIII. Note that there is an excellent agreement between the one- and three-prong sub-channels. This is expected as the differences between data and simulation are driven by the effects related to the lepton leg. While we quote all four numbers, based on a very good agreement between electrons or muon channels we use two factors obtained by weiging the scale factors with one and three prong tau candidates and further reduce the uncertainty on $Z \to \tau\tau$ scale.

As for potential concerns over the possible Higgs signal leaking into the region below $M(\tau\tau) = 100$ GeV/c^2 (one can see that this indeed happens in Fig. 28), one should note that even for the least populous channel, the number of events in the 65<$M(\tau\tau)$<100 GeV/c^2 region is of the order of a thousand events. Therefore, affecting the $Z \to \tau\tau$ scale at the level of 1 sigma would require about 30 Higgs events with $M(\tau\tau) < 100$ GeV/c^2. In this case, the region above $M(\tau\tau) < 100$ GeV/c^2 would have a very strong enhancement of Higgs events over the (unchanged shape and 3% increased scale) $Z \rightarrow \tau\tau$ background, leading to only a slight underestimation of the Higgs signal significance. For Higgs cross-sections of the order of the SM expectations, one expects no visible effects on the limit.

Table XVII. $Z \rightarrow \tau\tau$ normalization factors with the absolute and relative uncertainties for the inclusive N_{jet} >=0 channel, $K(Z \rightarrow \tau_h\tau_l) \pm \sigma_{K(Z\rightarrow\tau_h\tau_l)}(\epsilon_{K(Z\rightarrow\tau_h\tau_l)}, \%)$

	$\tau_h\tau_e$	$\tau_h\tau_\mu$
τ_h(1-prong)	0.867 ± 0.042 (4.8)	0.629 ± 0.028 (4.5)
τ_h(3-prong)	0.801 ± 0.057 (7.1)	0.639 ± 0.034 (5.4)
1- and 3-prong	0.844 ± 0.034 (4.0)	0.633 ± 0.022 (3.5)

a. Jet Counting Scale Factors

To obtain the normalization scale corrections accounting for differences in data in simulation for the distribution of the inclusive events between the sub-regions with and without jets, we use the data from the same $Z \rightarrow ee$ control region to measure the fraction of events with and without jets, $f_{0\,jets}$ and $f_{1+\,jets}$. The same value (as the probability of observing a jet in Z events does not depent on the channel or the number of prongs) is then used for each of the four channels as the scale correction for the normalization of events in pairs of the jet sub-regions.

Because we re-calculate the QCD contribution on the fly for each sub-region, we

Table XVIII. $R^{QCD}_{os/ss}$ factors with the absolute and relative uncertainties for the inclusive N_{jet} >=0 channel.

	$\tau_h\tau_e$	$\tau_h\tau_\mu$
τ_h(1-prong)	1.10 ± 0.05 (4.5)	1.40 ± 0.18 (12.9)
τ_h(3-prong)	0.97 ± 0.10 (10.3)	0.93± 0.10 (10.8)

only need to measure scales $f_{0\,jets}$ and $f_{1+\,jets} = 1 - f_{0\,jets}$ only for $Z \to \tau\tau$ events. Given that such measurement in a super clean sample of $Z \to \tau\tau$ events will be limited by statistics, the simplest approach is to measure the efficiency of the jet requirement in $Z \to ee$ events. One only has to ensure that the topology of selected $Z \to ee$ events is similar to that of the selected $Z \to \tau\tau$ events to avoid biases. Because signal region selections are highly efficient for $Z \to \tau\tau$ events, the distribution for the number of jets should come out the same as for "normal" $Z \to ee$ events as long as no peculiar cuts are applied and lepton acceptance ranges are similar. We therefore select $Z \to ee$ events in the same way as in the Sec.4.3 (reconstructing one of the electrons as a tau candidate ensures that the acceptance coverage in η is exactly the same) except that we drop the topology cuts designed to reject W+jets (Table VII) because applying such cuts to $Z \to ee$ events signifcantly modifies the topology of the "normal" $Z \to ee$ events. This is because $\not\!\!E_{\mathrm{T}}$-related requirements, when applied to $Z \to ee$ enhance selection of events with jets (there is no true $\not\!\!E_{\mathrm{T}}$ in $Z \to ee$ events, but experimentally measurable imbalance can come from a mismeasured jet). In the case of $Z \to \tau\tau$ events, the same selections are not modifying the topology of the events because of true $\not\!\!E_{\mathrm{T}}$ present in these events. As a quantitative proof of these observations, we measure f's in thus selected $Z \to ee$ and $Z \to \tau\tau$ simulated events and observe a good agreement (see Table XIX and Fig. 29). We also verified that W+jets rejection cuts applied to $Z \to \tau\tau$ MC do not change jets distribution (see Fig. 30). Small difference in f's between $Z \to ee$ and $Z \to \tau\tau$ is due to higher full p_T of the leptons in $Z \to \tau\tau$ selection. Only visible decay products are reconstructed for tau leptons and the same cuts on visible p_T applied to $Z \to \tau\tau$ and $Z \to ee$ on average select events with higher boost of the Z in the former case. Thus, the probability to have a jet in the event is slightly higher for $Z \to \tau\tau$ events. We rely on MC to correct for this effect and calculate f's as follows: $f = f^{Z \to ee(DATA)} \cdot \frac{f^{Z \to \tau\tau(MC)}}{f^{Z \to ee(MC)}}$.

Once all these checks are done, we measure the fraction $f_{0\,jets}$, the results are shown in Table XIX.

Table XIX. Fraction of events in the =0-jet and >0-jet channels used for the $Z \to \tau\tau$ background.

	$F_{=0}$	$F_{>0}$
Data $Z \to ee$	0.958768 ± 0.00406156 (0.423623)	0.0412319 ± 0.000614094 (1.48937)
MC $Z \to ee$	0.958258 ± 0.00542452 (0.566081)	0.0417425 ± 0.000825761 (1.97823)
MC $Z \to \tau\tau$	0.956273 ± 0.00839393 (0.877775)	0.0437272 ± 0.00131108 (2.99832)
Final	0.956782 ± 0.0107841(0.0112712)	0.0431923 ± 0.00167959(0.0388864)

b. Final Scale Factors and the Uncertainties

We obtain the final scale factors for $Z \to \tau\tau$ as a product of the “inclusive” scale factor K for that channel times the measured fraction of events with and without extra jets. Final normalization factors are shown in Table XX.

Table XX. Final uncertainties on the total normalizations of the $Z \rightarrow \tau\tau$ background.

$= 0$-jet	$\tau_h\tau_e$	$\tau_h\tau_\mu$
τ_h(1- or 3-prong)	4.18	3.65
> 0-jet	$\tau_h\tau_e$	$\tau_h\tau_\mu$
τ_h(1- or 3-prong)	5.60	5.22

5. Shape of the QCD Multi-Jet Background

Only small fraction of QCD multi-jet events enter our signal region $M_{\tau\tau} > 100$ GeV/c^2. While low background contribution is favorable for better sensitivity in a simple counting experiment, low statics of the QCD events causes large uncertainties on the shape. The shape of the MMC reconstructed mass for the QCD bakcground does not depend on the isolation requirements on hadronic tau or lepton. Therefore we construct a region enriched with the background coming from QCD events by modifying isolation requirments as shown in Table XXI. While relaxed isolation on the tau side allows for more $jet \rightarrow \tau_h$ fakes, the anti-isolation requirement on the lepton side ensures that other backgrounds (mostly $W + jets$) are very low. We fit $M_{\tau\tau}$ distribution for SS events with exponential and use this fit in the signal region. Figures 31 and 32 show fits for all eight channels.

Table XXI. Isolation requirements for the QCD enriched sample.

Cut
$I_{trk}^{0.17<\Delta R<0.52}(\tau_h)<8$ GeV/c
$I_{trk}^{\Delta R<0.4}(l)>1$ GeV/c

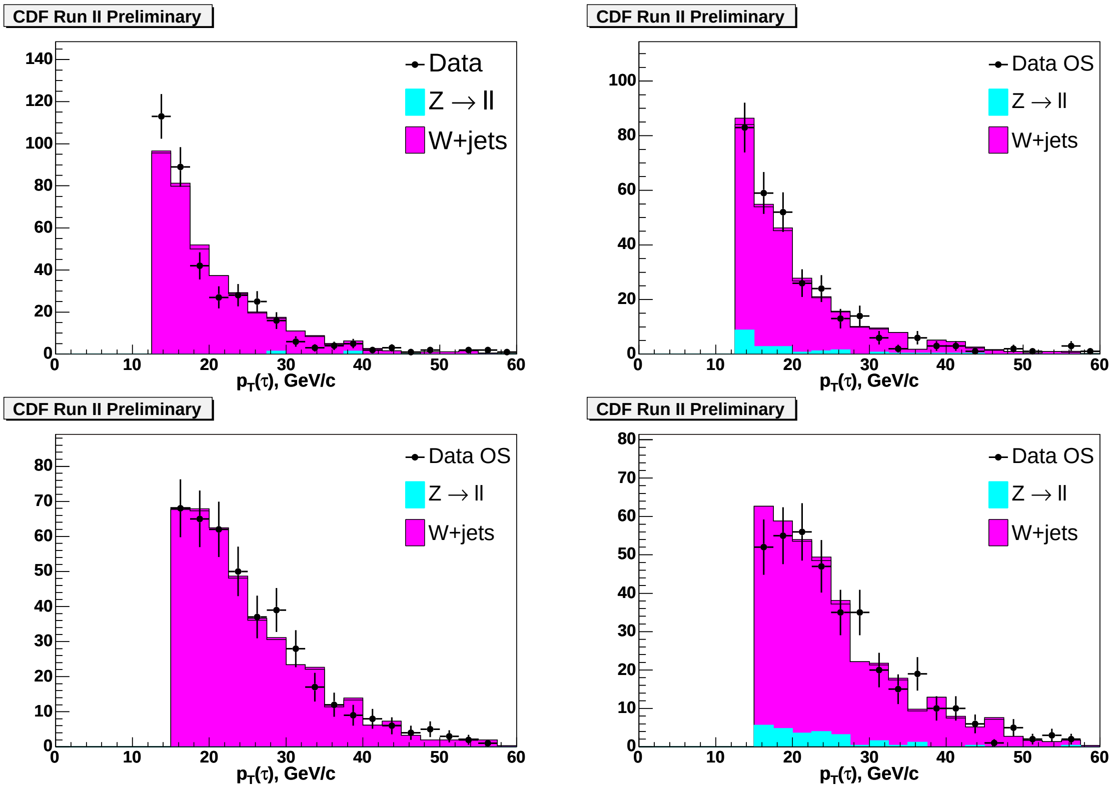

Fig. 24. Transverse momentum of hadronically decaying tau leptons, $p_T(\tau)$. W+jets Opposite Sign events. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

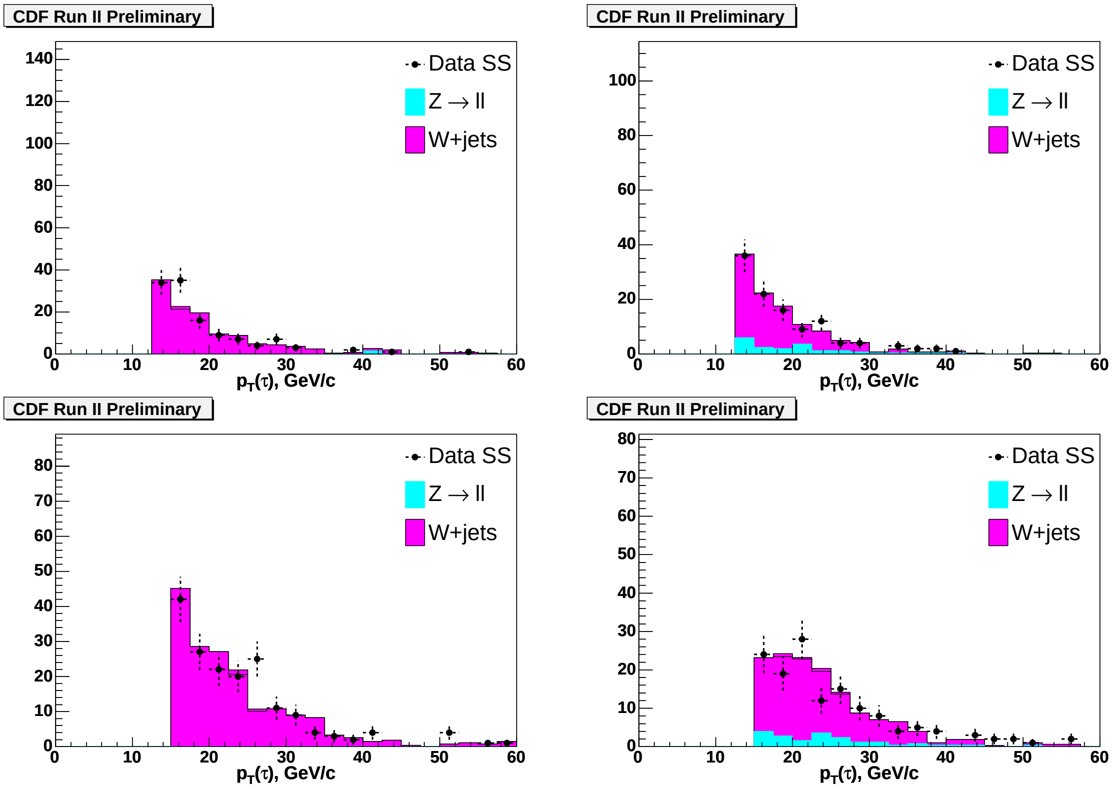

Fig. 25. Transverse momentum of hadronically decaying tau leptons, $p_T(\tau)$. W+jets Same Sign events. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

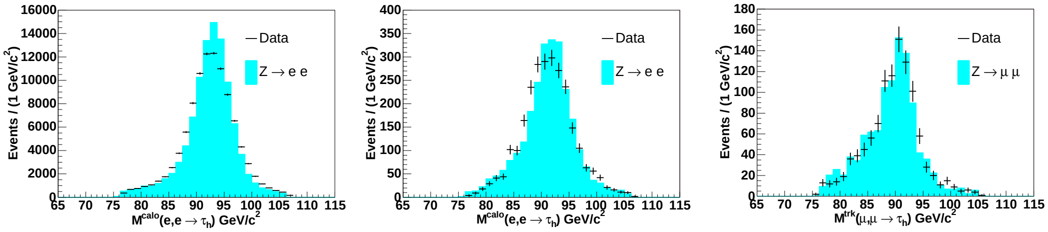

Fig. 26. $Z \longrightarrow ee$ and $Z \longrightarrow \mu\mu$ events after normalization. Left: $e \longrightarrow \tau_h$(1-prong). Middle: $e \longrightarrow \tau_h$(3-prong). Right: $\mu \longrightarrow \tau_h$(1-prong).

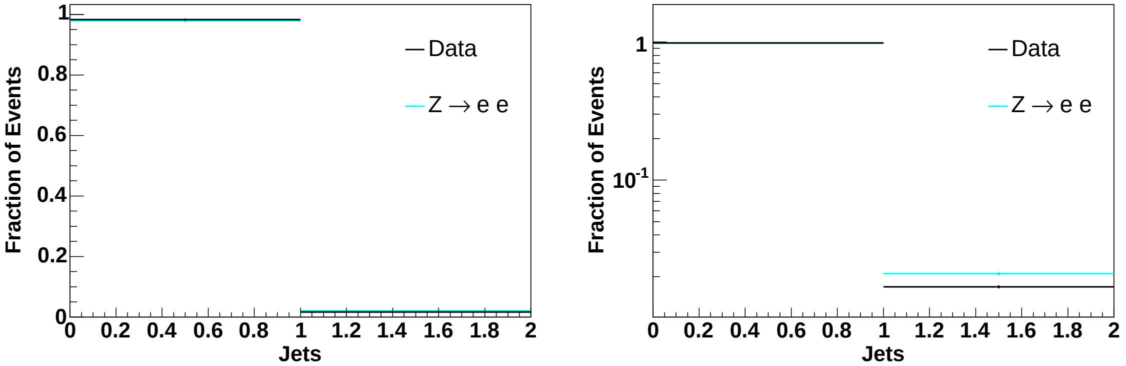

Fig. 27. Jet count histogram. $Z \longrightarrow ee$ simulation is compared to data. All analysis cuts applied.

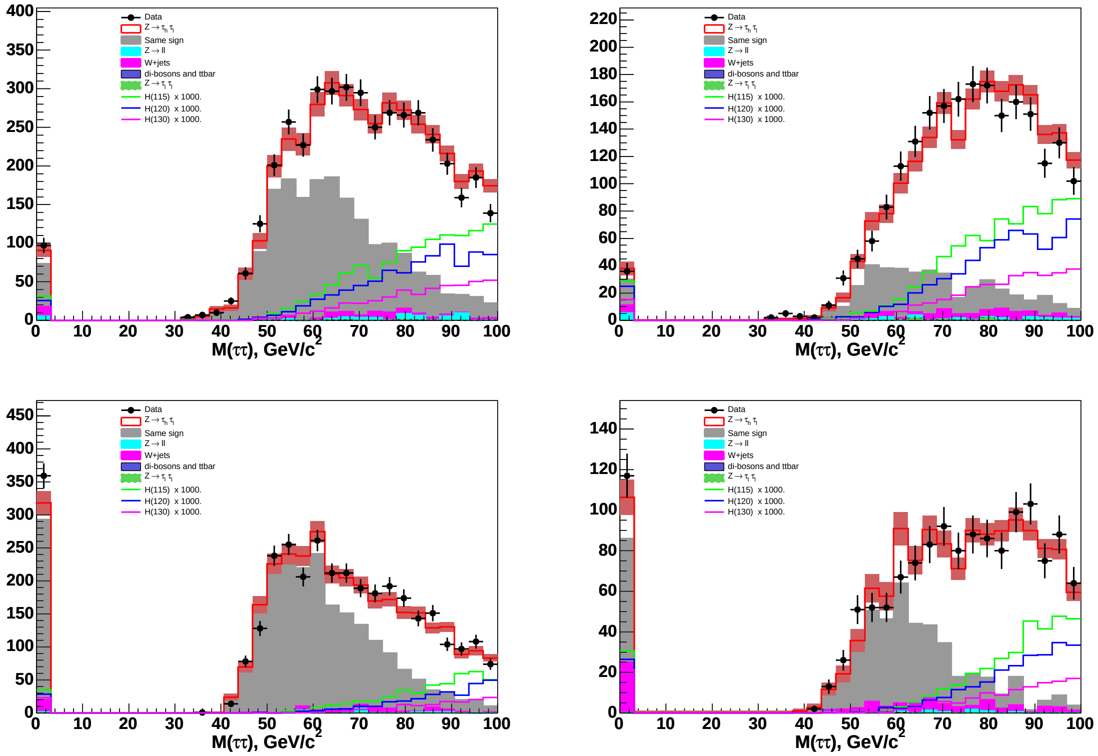

Fig. 28. MMC-based invariant mass distribution of the $\tau\tau$ system for $M(\tau\tau) < 100$ GeV/c^2 for the four signal regions before splitting into the $N_{jets} = 0$ and $S_{jets} > 1$ sub-regions. Left: $e + \tau_h$ channel, right: $\mu + \tau_h$ channel. Top: 1-prong hadronic taus, bottom: 3-prong hadronic taus. Data is compared to background estimations and the expected Higgs signal contributions from all production modes combined (multiplied by 1000).

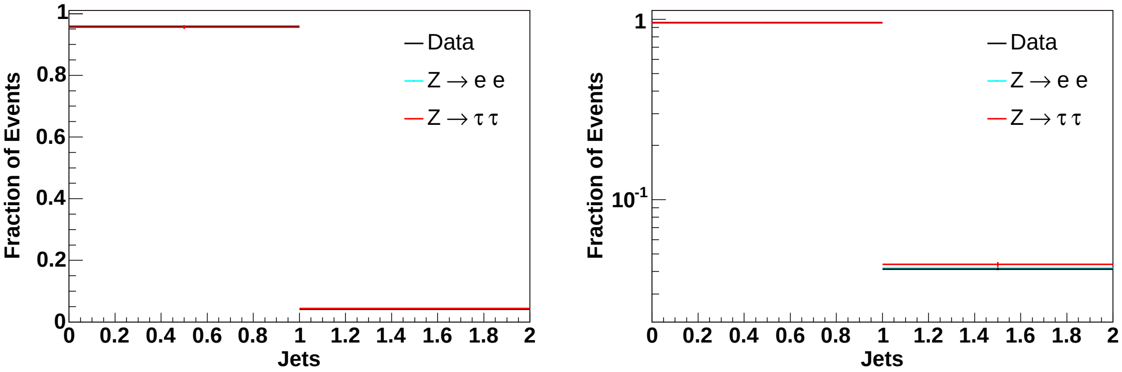

Fig. 29. Jet count histogram. $Z \to ee$ simulation is compared to data. No W+jets rejection cuts.

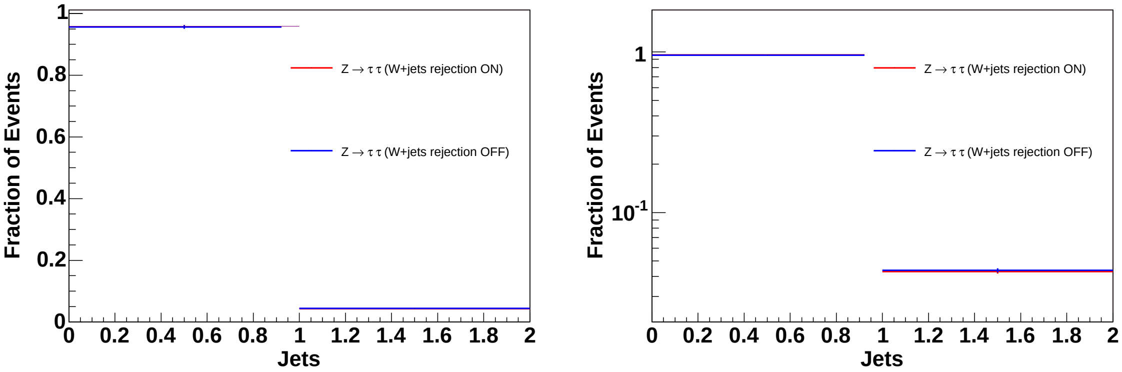

Fig. 30. Jet count histogram for $Z \to \tau\tau$ simulation with and without W+jets rejection cuts.

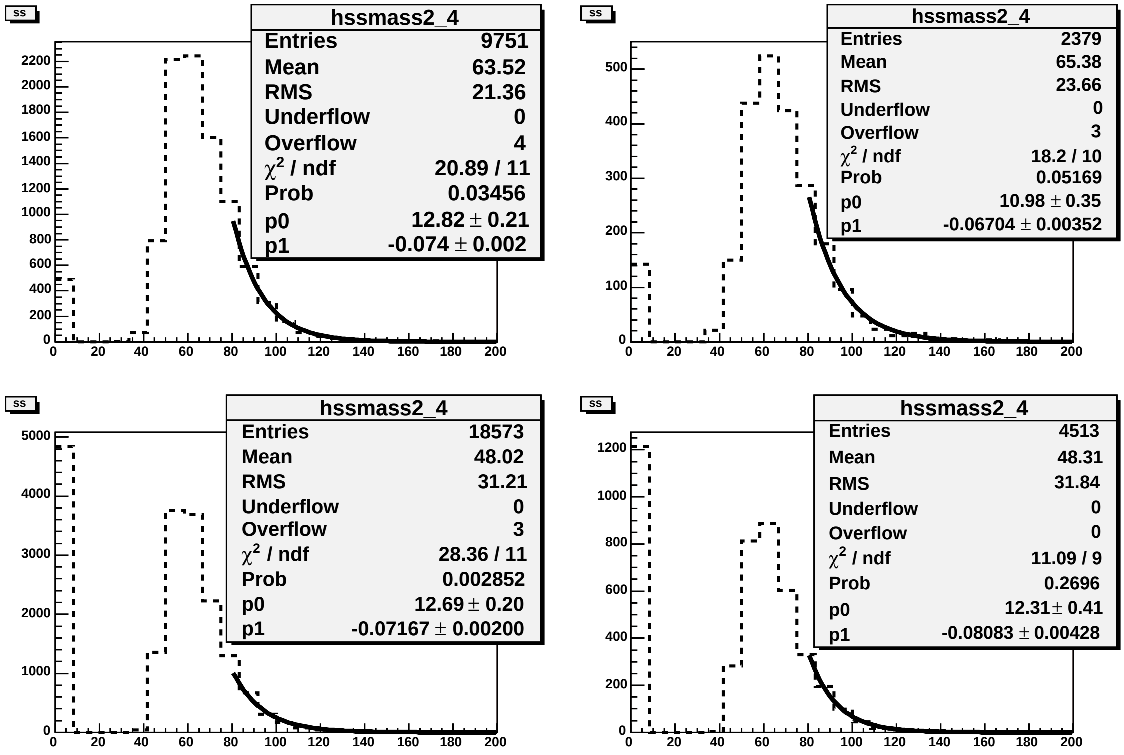

Fig. 31. Fit of the QCD multi-jet $M_{\tau\tau}$ shape distribution using SS events in the QCD enriched sample. = 0jet channels. Left: electrons, right muons. Top: 1-prong, bottom: 3-prong.

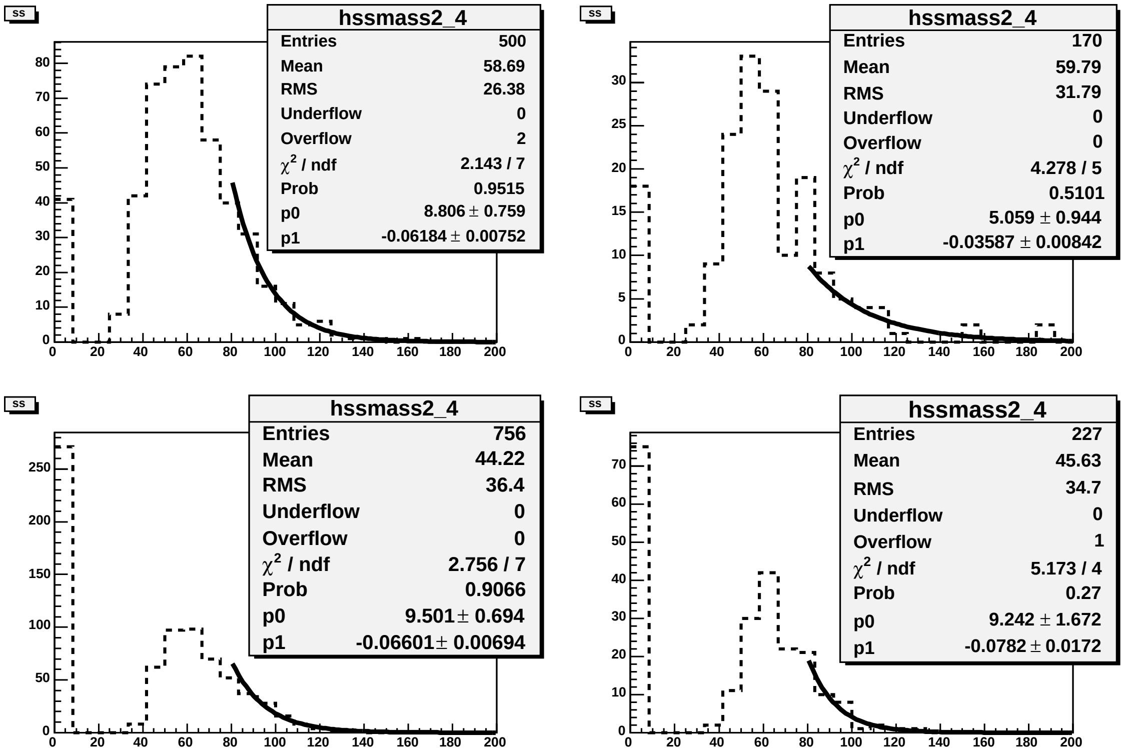

Fig. 32. Fit of the QCD multi-jet $M_{\tau\tau}$ shape distribution using SS events in the QCD enriched sample. $>$ 0jet channels. Left: electrons, right muons. Top: 1-prong, bottom: 3-prong.

D. Validation of the Background Prediction: Kinematic Distributions and Event Yield

We first verify that our prediction for the distributions of various kinematic parameters ($p_T(\tau_h)$, $p_T(l)$, $\not{E}_{\mathrm{T}}$, etc.) agrees with the observed data. We than perform event counting in the control region ($M_{\tau\tau}$<100 GeV/c^2). By comparing the number of predicted background events and the number of events in data we ensure that our background model is in good agreement with data in the region where no significant contribution from the signal is expected.

1. Kinematic Distributions

Figures 33 - 56 show kinematic distributions for each channel demonstrating good agreement between data and the background prediction.

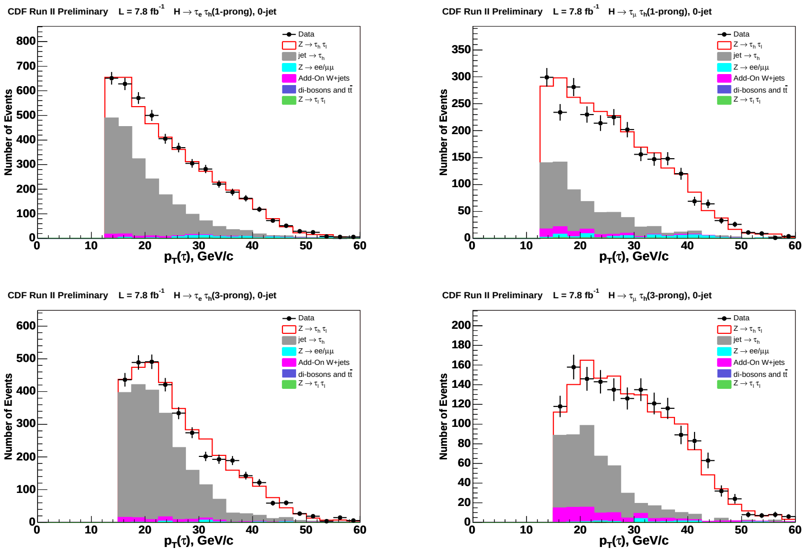

Fig. 33. Transverse momenum of hadronically decaying taus, $p_T(\tau)$. Events with N_{jet}=0. Data (points) compared to the background prediction. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

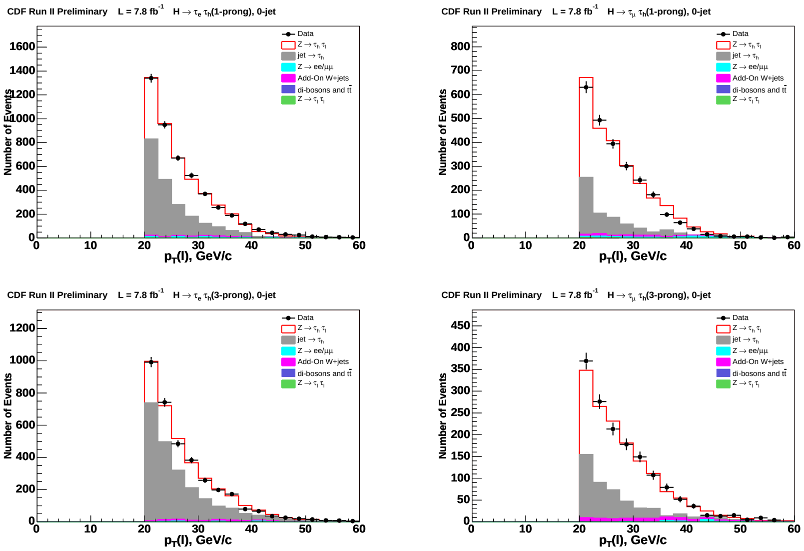

Fig. 34. Transverse momentum of leptonically decaying taus, $p_T(l)$. Events with N_{jet}=0. Data (points) compared to the background prediction. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

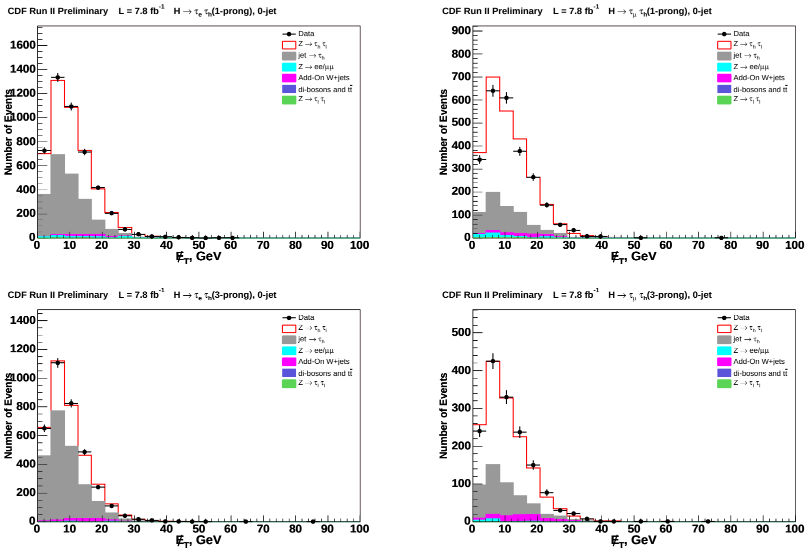

Fig. 35. Transverse missing energy, $\not{E}_{\mathrm{T}}$. Events with N_{jet}=0. Data (points) compared to the background prediction. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

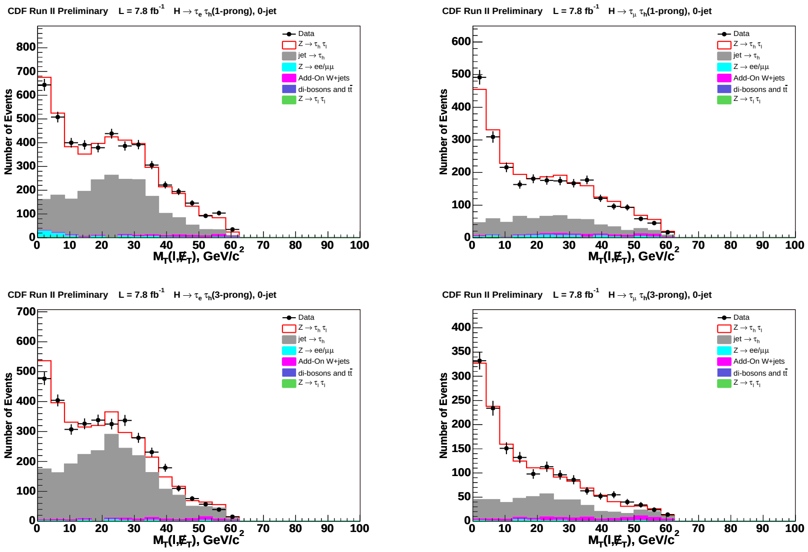

Fig. 36. Transverse mass $M_T(l, \not{E}_{\mathrm{T}})$. Events with N_{jet}=0. Data (points) compared to the background prediction. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

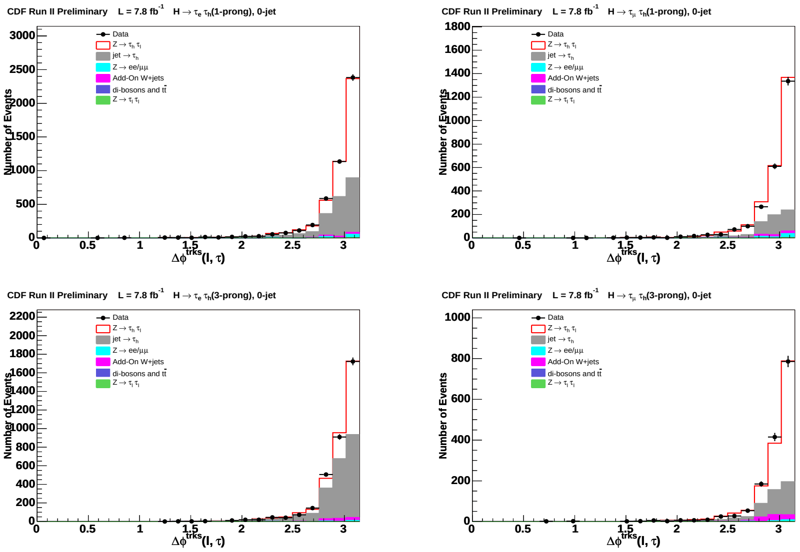

Fig. 37. $\Delta\phi(l, \tau)$. Events with N_{jet}=0. Data (points) compared to the background prediction. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

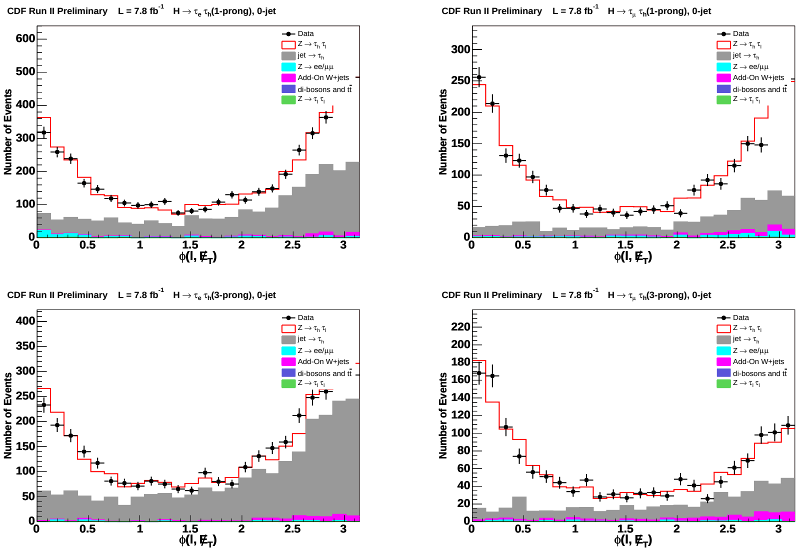

Fig. 38. $\Delta\phi(l, \not\!\!E_{\mathrm{T}})$. Events with N_{jet}=0 Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

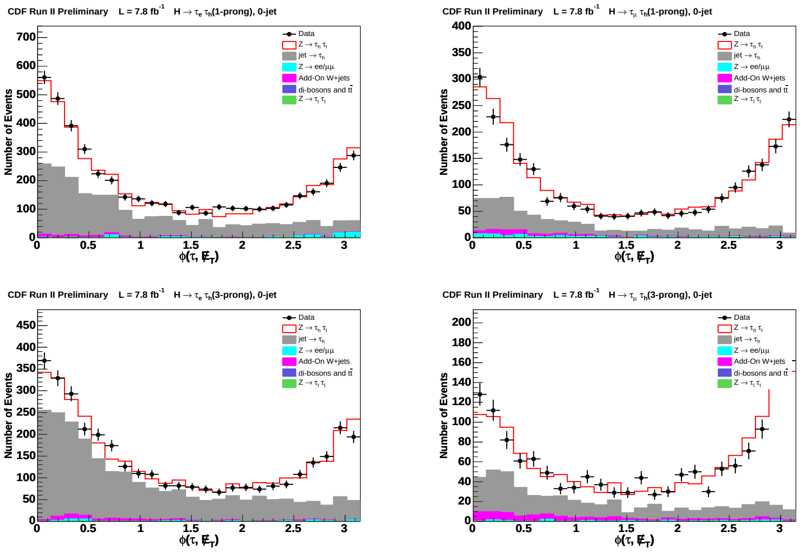

Fig. 39. $\Delta\phi(\tau, \not{E}_T)$. Events with N_{jet}=0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

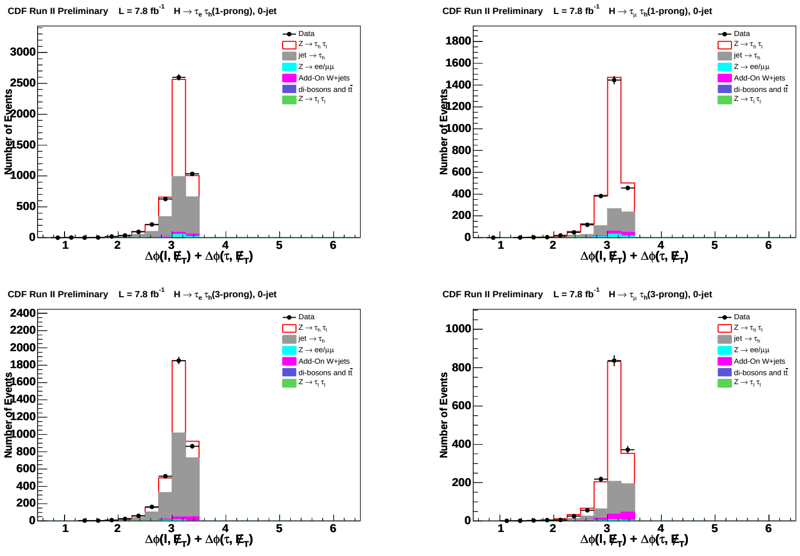

Fig. 40. $\Delta\phi(l, \tau) + \Delta\phi(\tau, \not{E}_T)$. Events with N_{jet}=0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

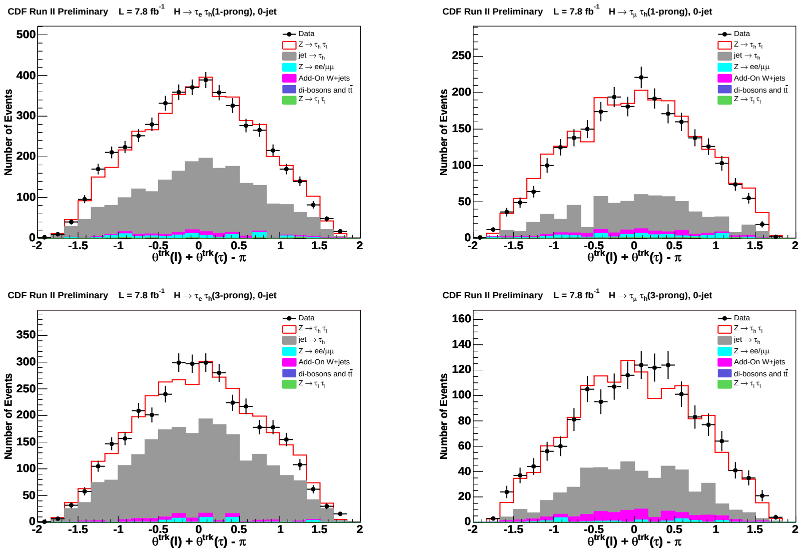

Fig. 41. $\theta(l) + \theta(\tau) - \pi$. Events with N_{jet}=0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

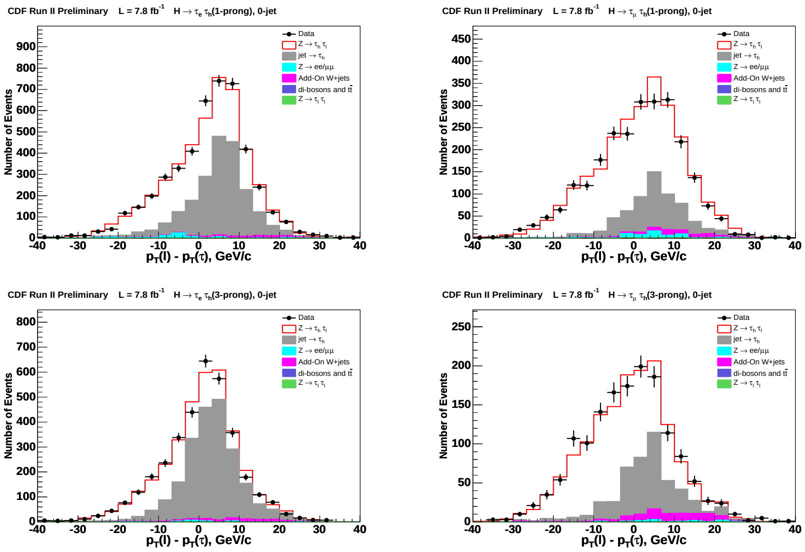

Fig. 42. $p_T(l) - p_T(\tau)$. Events with N_{jet}=0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

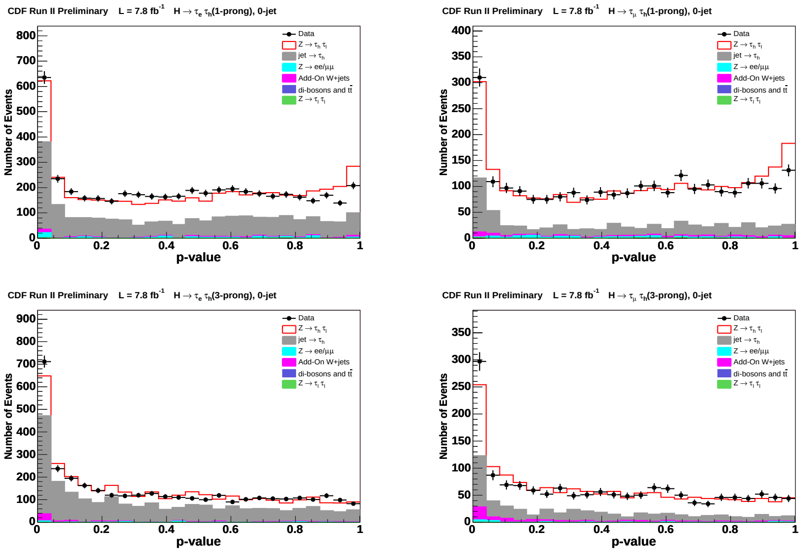

Fig. 43. p-value. Events with N_{jet}=0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

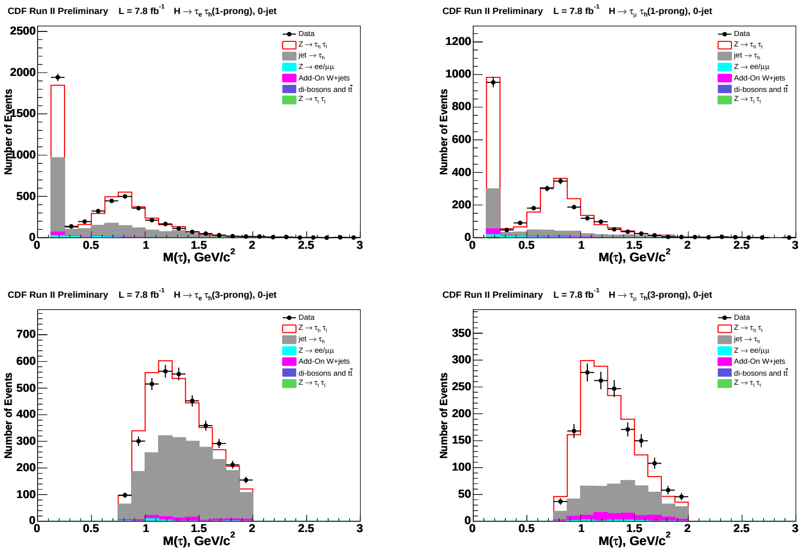

Fig. 44. Invariant mass of hadronically decaying taus, $M(\tau)$. Events with N_{jet}=0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

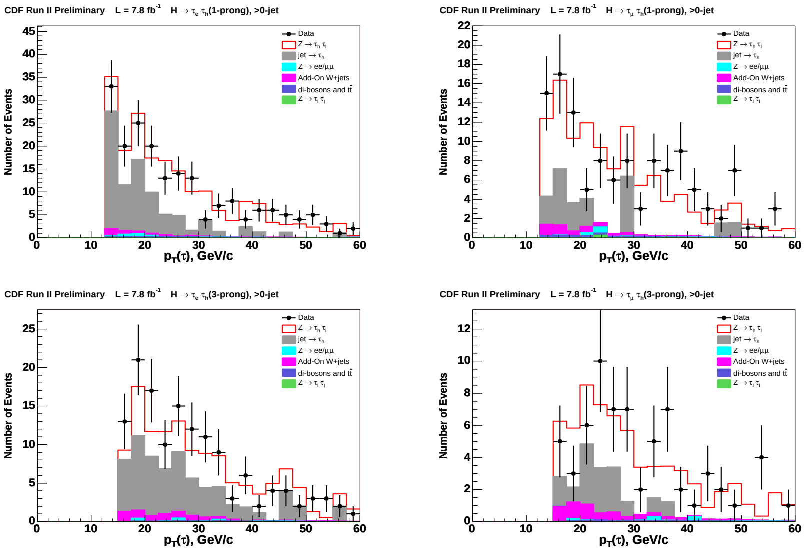

Fig. 45. Transverse momenum of hadronically decaying taus, $p_T(\tau)$. Events with N_{jet}>0. Data (points) compared to the background prediction. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

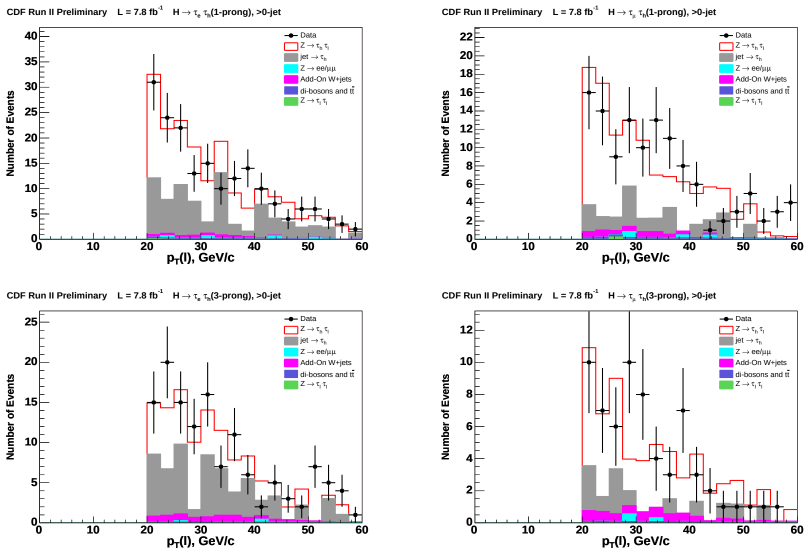

Fig. 46. Transverse momentum of leptonically decaying taus, $p_T(l)$. Events with N_{jet}>0. Data (points) compared to the background prediction. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

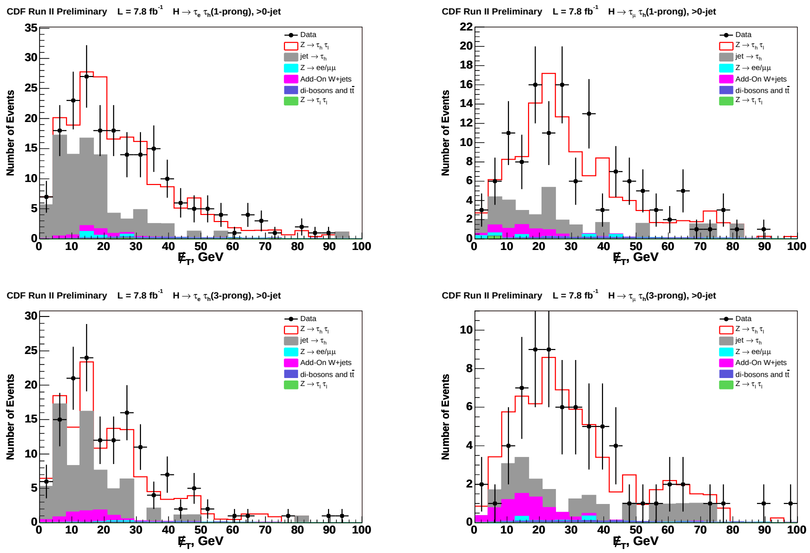

Fig. 47. Transverse missing energy, $\not{E}_{\mathrm{T}}$. Events with N_{jet}>0. Data (points) compared to the background prediction. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

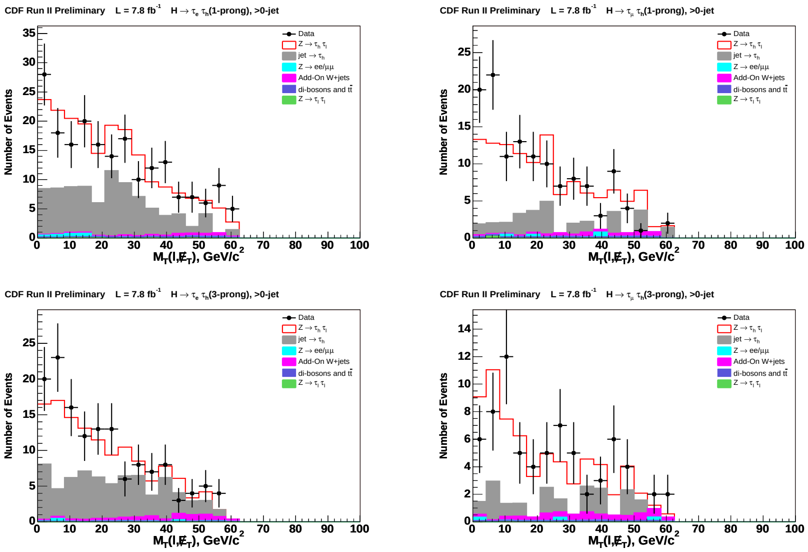

Fig. 48. Transverse mass $M_T(l,\not{E}_T)$. Events with N_{jet}>0. Data (points) compared to the background prediction. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

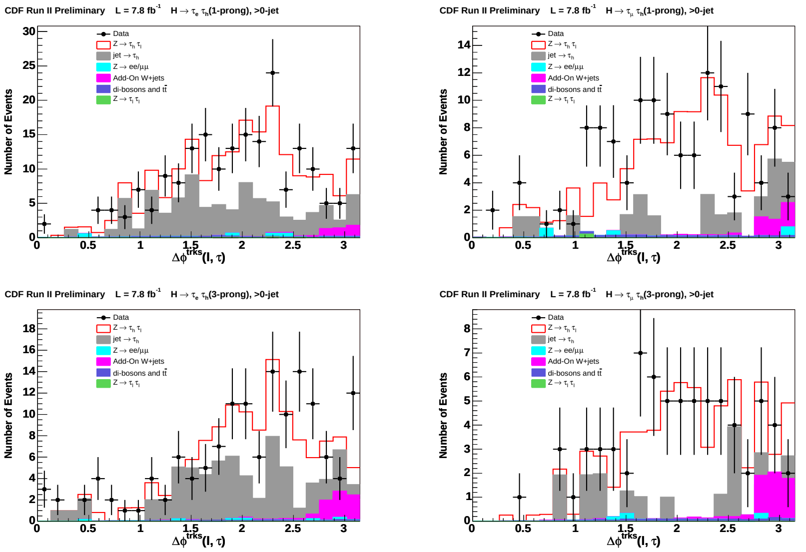

Fig. 49. $\Delta\phi(l,\tau)$. Events with N_{jet}>0. Data (points) compared to the background prediction. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

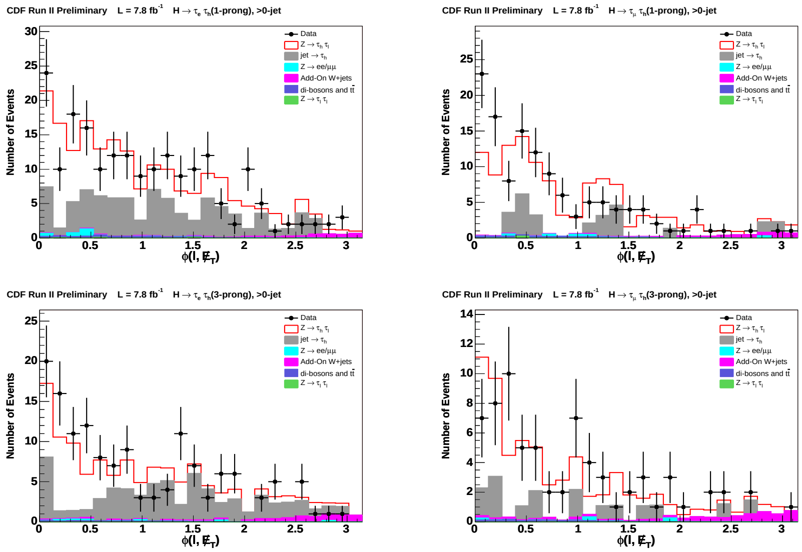

Fig. 50. $\Delta\phi(l, \not{E}_{\mathrm{T}})$. Events with N_{jet}>0 Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

2. Event Yield in the Control Region

We estimate the number of events in the control region based on our background model described in the previous section. The number of the signal events is obtained from the simulation using production cross sections and $\tau\tau$ decay branching ratio reported in [37] and summarized in Table XXII.

The distribution for $M_{\tau\tau}$ in the control region is shown in Fig. 57-58. In Tables XXIII-XXX we report the number of expected and observed events for each channel in the control region.

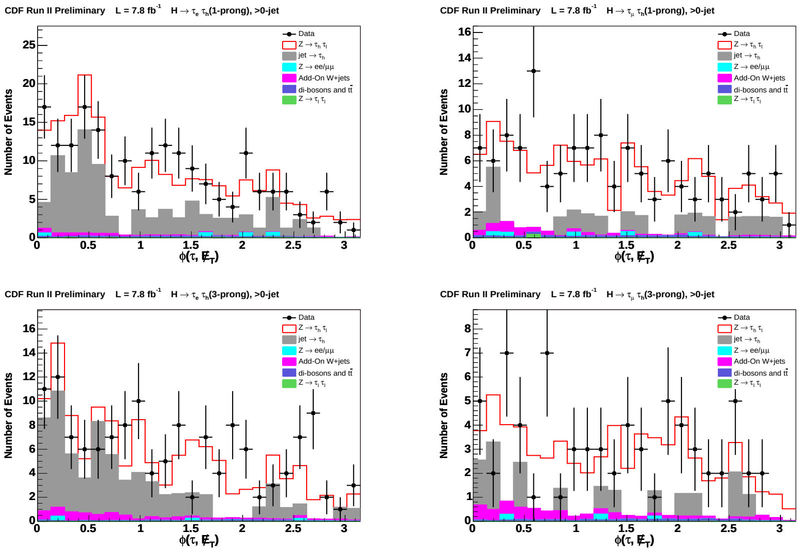

Fig. 51. $\Delta\phi(\tau, \not{E}_{\mathrm{T}})$. Events with N_{jet}>0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

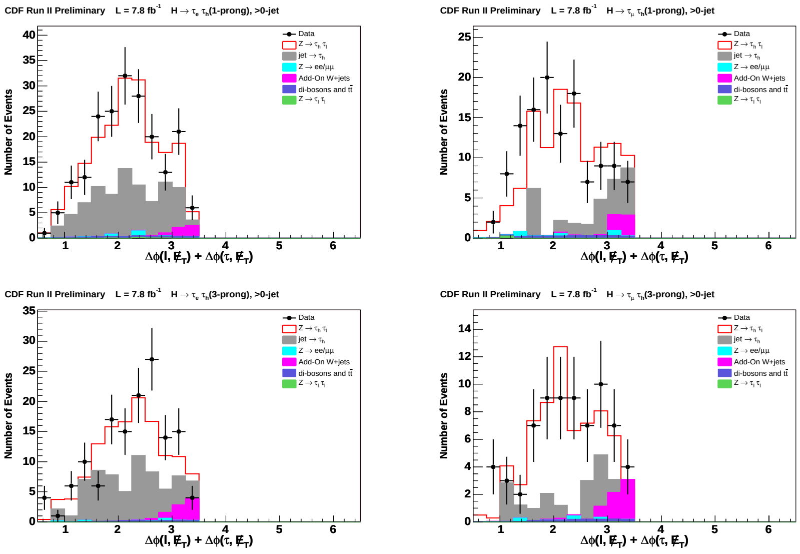

Fig. 52. $\Delta\phi(l, \tau) + \Delta\phi(\tau, \not{E}_{\mathrm{T}})$. Events with N_{jet}>0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

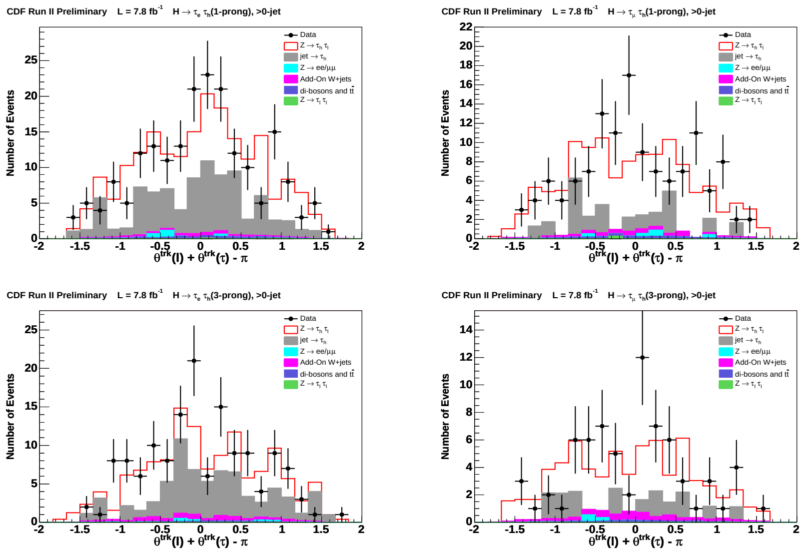

Fig. 53. $\theta(l) + \theta(\tau) - \pi$. Events with N_{jet}>0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

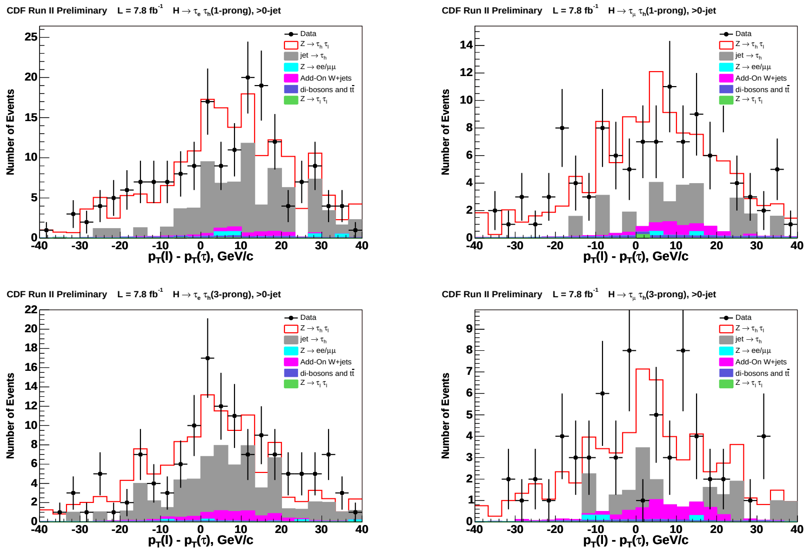

Fig. 54. $p_T(l) - p_T(\tau)$. Events with N_{jet}>0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

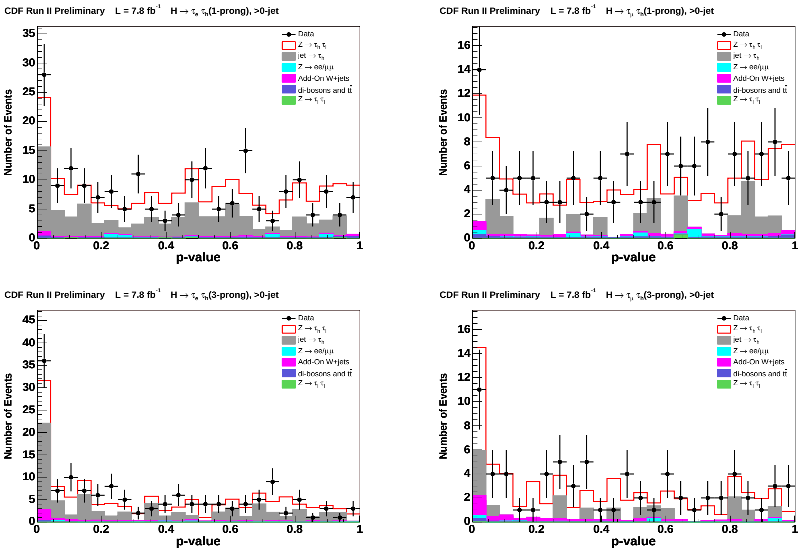

Fig. 55. p-value. Events with N_{jet}>0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

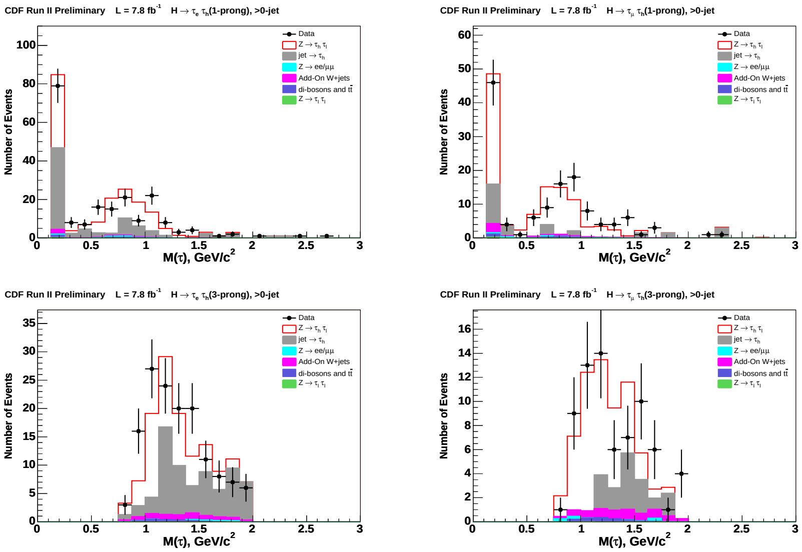

Fig. 56. Invariant mass of hadronically decaying taus, $M(\tau)$. Events with N_{jet}>0. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

Table XXII. Higgs boson production cross-section and $\tau\tau$ decay branching ratio.

M_H	ggH	WH	ZH	VBF	BR(%)
115	1215.9	174.50	103.9	78.6	7.288
120	1072.3	150.10	90.2	72.7	6.789
130	842.9	112.00	68.5	62.1	5.305
140	670.6	84.60	52.7	53.2	3.472
150	539.1	64.40	40.8	45.8	1.778

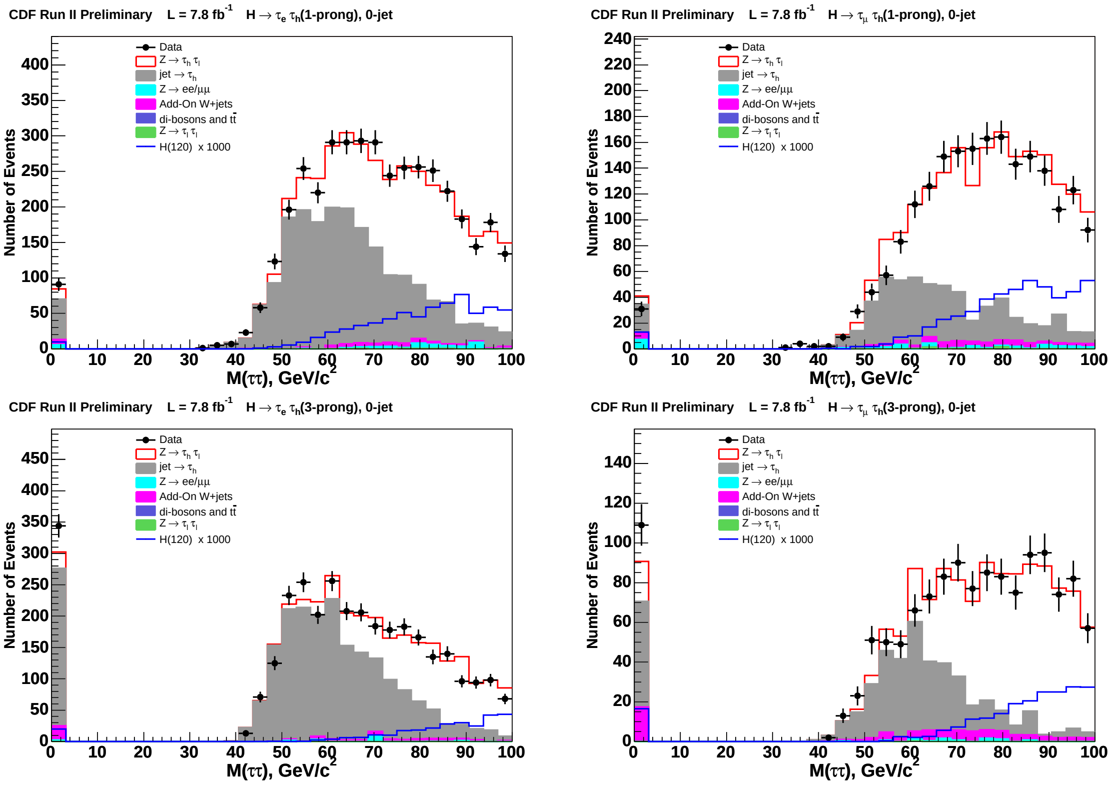

Fig. 57. Di-tau invariant mass, $M_{\tau\tau}$. Events with N_{jet}=0. Control region. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

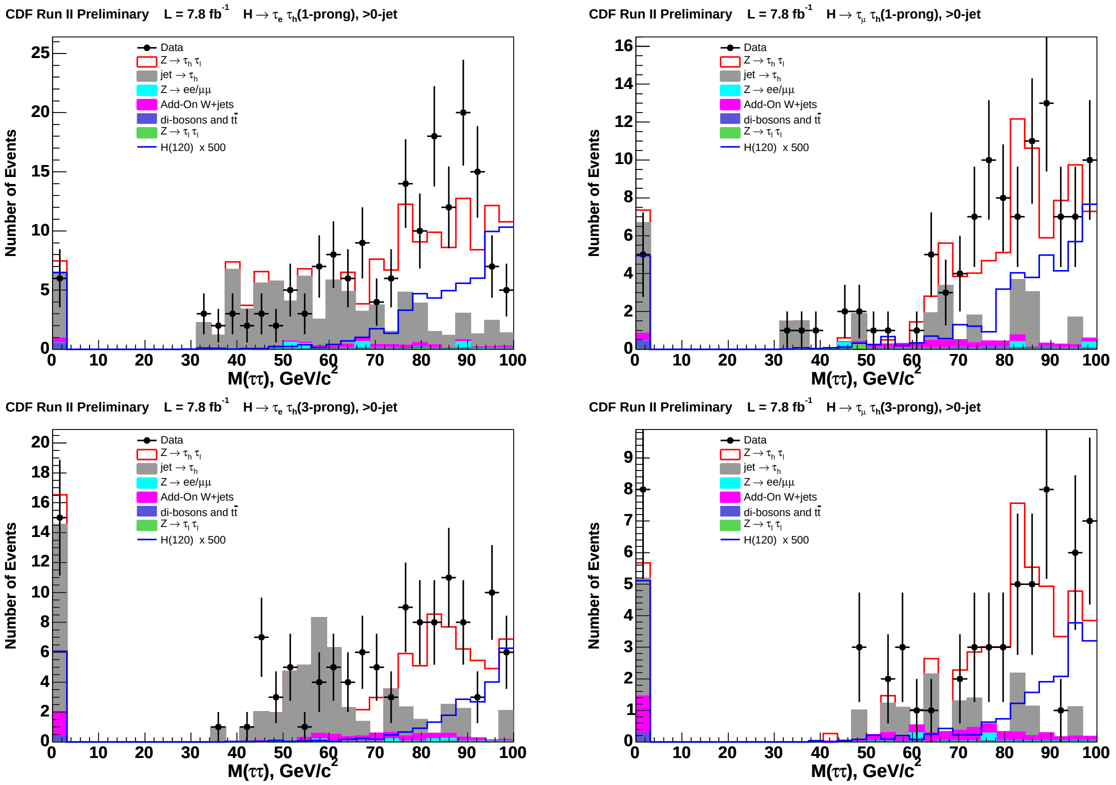

Fig. 58. Di-tau invariant mass, $M_{\tau\tau}$. Events with N_{jet}>0. Control region. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

Table XXIII. Event yield in the control region $M_{\tau\tau}$ <100 GeV: $\tau_e\tau_h$(1-prong), N_{jet} >0

Background source	Events (uncertainty, %)
$Z \to \tau_h\tau_l$	79.4 ± 4.45 (5.6)
fakes from SS	74.8 ± 9.32 (12.5)
add-on W+jets	4.06 ± 1.21 (29.8)
$Z \to ll$	1.3 ± 0.0325 (2.5)
di-bosons	0.299 ± 0.0299 (10)
$t\bar{t}$	2.2 ± 0.22 (10)
$Z \to \tau_l\tau_l$	0.0551 ± 0.00338 (6.13)
Total background	161 ± 10.4 (6.47)
OS data	170

Signal expectations

M_H	ggH	WH	ZH	VBF	
115	0.0654	0.0504	0.0335	0.0266	0.176
120	0.0526	0.0334	0.0189	0.0198	0.125
130	0.021	0.0169	0.00841	0.00979	0.0561

Table XXIV. Event yield in the control region $M_{\tau\tau}$ <100 GeV: $\tau_e\tau_h$(1-prong), N_{jet} =0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h\tau_l$	1894 ± 79.2 (4.18)
fakes from SS	1965 ± 102 (5.17)
add-on W+jets	72.9 ± 9.07 (12.4)
$Z \rightarrow ll$	36.1 ± 0.242 (0.671)
di-bosons	2.63 ± 0.263 (10)
$t\bar{t}$	0.227 ± 0.0227 (10)
$Z \rightarrow \tau_l\tau_l$	5.69 ± 0.241 (4.24)
Total background	3941 ± 129 (3.28)
OS data	4011

Signal expectations					
M_H	ggH	WH	ZH	VBF	
115	0.808	0.0349	0.017	0.0112	0.871
120	0.619	0.0275	0.0122	0.00891	0.668
130	0.341	0.0146	0.00642	0.0043	0.366

Table XXV. Event yield in the control region $M_{\tau\tau}$ <100 GeV: $\tau_e\tau_h$(3-prong), N_{jet} >0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h\tau_l$	44.3 ± 2.48 (5.6)
fakes from SS	56 ± 7.83 (14)
add-on W+jets	6.96 ± 1.74 (25)
$Z \rightarrow ll$	0.128 ± 0.00505 (3.95)
di-bosons	0.144 ± 0.0144 (10)
$t\bar{t}$	0.98 ± 0.098 (10)
$Z \rightarrow \tau_l\tau_l$	0
Total background	108 ± 8.39 (7.74)
OS data	123

Signal expectations					
M_H	ggH	WH	ZH	VBF	
115	0.0314	0.0241	0.0137	0.0133	0.0824
120	0.0222	0.0158	0.00871	0.00965	0.0563
130	0.0128	0.0073	0.0039	0.0045	0.0285

Table XXVI. Event yield in the control region $M_{\tau\tau}$ <100 GeV: $\tau_e\tau_h$(3-prong), N_{jet} =0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h\tau_l$	1071 ± 44.8 (4.18)
fakes from SS	2106 ± 97.7 (4.64)
add-on W+jets	91.1 ± 12.9 (14.1)
$Z \rightarrow ll$	1.95 ± 0.061 (3.13)
di-bosons	1.47 ± 0.147 (10)
$t\bar{t}$	0.0978 ± 0.00978 (10)
$Z \rightarrow \tau_l\tau_l$	0
Total background	3270 ± 108 (3.31)
OS data	3254

Signal expectations

M_H	ggH	WH	ZH	VBF	
115	0.406	0.017	0.00764	0.00563	0.436
120	0.266	0.0125	0.00509	0.00381	0.288
130	0.143	0.00624	0.00323	0.00218	0.154

Table XXVII. Event yield in the control region $M_{\tau\tau}$ <100 GeV: $\tau_\mu\tau_h$(1-prong), N_{jet} >0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h\tau_l$	63.9 ± 3.33 (5.22)
fakes from SS	23.4 ± 5.6 (23.9)
add-on W+jets	4.86 ± 1.05 (21.7)
$Z \rightarrow ll$	0.613 ± 0.0272 (4.44)
di-bosons	0.245 ± 0.0245 (10)
$t\bar{t}$	1.7 ± 0.17 (10)
$Z \rightarrow \tau_l\tau_l$	0.268 ± 0.0184 (6.85)
Total background	94.4 ± 6.61 (7)
OS data	107

		Signal expectations			
M_H	ggH	WH	ZH	VBF	
115	0.0467	0.0396	0.0222	0.0197	0.128
120	0.0342	0.0277	0.0152	0.0135	0.0906
130	0.0162	0.0148	0.00775	0.00716	0.0459

Table XXVIII. Event yield in the control region $M_{\tau\tau}$ <100 GeV: $\tau_\mu\tau_h$(1-prong), N_{jet} =0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h\tau_l$	1473 ± 53.8 (3.65)
fakes from SS	516 ± 66.4 (12.9)
add-on W+jets	63.1 ± 8.24 (13.1)
$Z \rightarrow ll$	29.7 ± 1.11 (3.73)
di-bosons	2.14 ± 0.214 (10)
$t\bar{t}$	0.137 ± 0.0137 (10)
$Z \rightarrow \tau_l\tau_l$	5.47 ± 0.286 (5.22)
Total background	2060 ± 85.9 (4.17)
OS data	2037

		Signal expectations			
M_H	ggH	WH	ZH	VBF	
115	0.594	0.0261	0.0137	0.00842	0.643
120	0.466	0.0185	0.00873	0.0064	0.5
130	0.249	0.0104	0.00489	0.00426	0.268

Table XXIX. Event yield in the control region $M_{\tau\tau}$ <100 GeV: $\tau_\mu\tau_h$(3-prong), N_{jet} >0

Background source	Events (uncertainty, %)
$Z \to \tau_h\tau_l$	34.5 ± 1.8 (5.22)
fakes from SS	14 ± 4.02 (28.7)
add-on W+jets	5.37 ± 1.4 (26.1)
$Z \to ll$	0.144 ± 0.0148 (10.3)
di-bosons	0.127 ± 0.0127 (10)
$t\bar{t}$	0.737 ± 0.0737 (10)
$Z \to \tau_l\tau_l$	0.0103 ± 0.00119 (11.5)
Total background	54.8 ± 4.63 (8.44)
OS data	61

		Signal expectations			
M_H	ggH	WH	ZH	VBF	
115	0.0203	0.0186	0.0105	0.00871	0.0582
120	0.0169	0.0139	0.00736	0.00604	0.0442
130	0.00698	0.00667	0.00312	0.00321	0.02

Table XXX. Event yield in the control region $M_{\tau\tau}$ <100 GeV: $\tau_\mu\tau_h$(3-prong), N_{jet} =0

Background source	Events (uncertainty, %)
$Z \to \tau_h\tau_l$	821 ± 30 (3.65)
fakes from SS	405 ± 47.1 (11.6)
add-on W+jets	78.8 ± 11.3 (14.3)
$Z \to ll$	3.15 ± 0.315 (10)
di-bosons	1.1 ± 0.11 (10)
$t\bar{t}$	0.0674 ± 0.00674 (10)
$Z \to \tau_l\tau_l$	0.229 ± 0.0244 (10.7)
Total background	1306 ± 57 (4.36)
OS data	1331

		Signal expectations			
M_H	ggH	WH	ZH	VBF	
115	0.276	0.0139	0.00689	0.0041	0.301
120	0.2	0.0105	0.00489	0.00283	0.219
130	0.112	0.00447	0.00223	0.00174	0.121

E. Event Yield in the Signal Region

After extensive cross checks of the background model in the control region we look at the signal region. Figures 59-62 shows $M_{\tau\tau}$ distribution in the signal region and in the mass range from 0 to 200 GeV/c^2. In Tables XXXI-XXXVIII we report the number of expected and observed events for each channel in the signal region. The number of the observed events is in agreement with background only hypothesis.

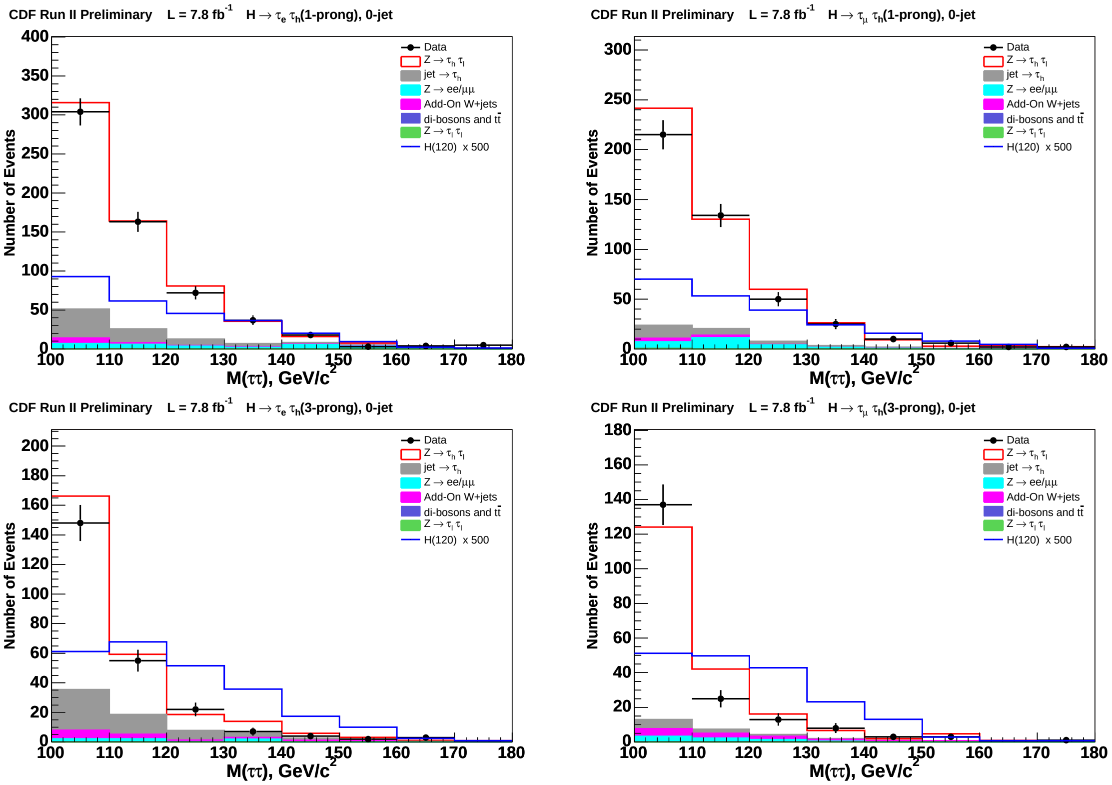

Fig. 59. Di-tau invariant mass, $M_{\tau\tau}$. Events with N_{jet}=0. Signal region. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

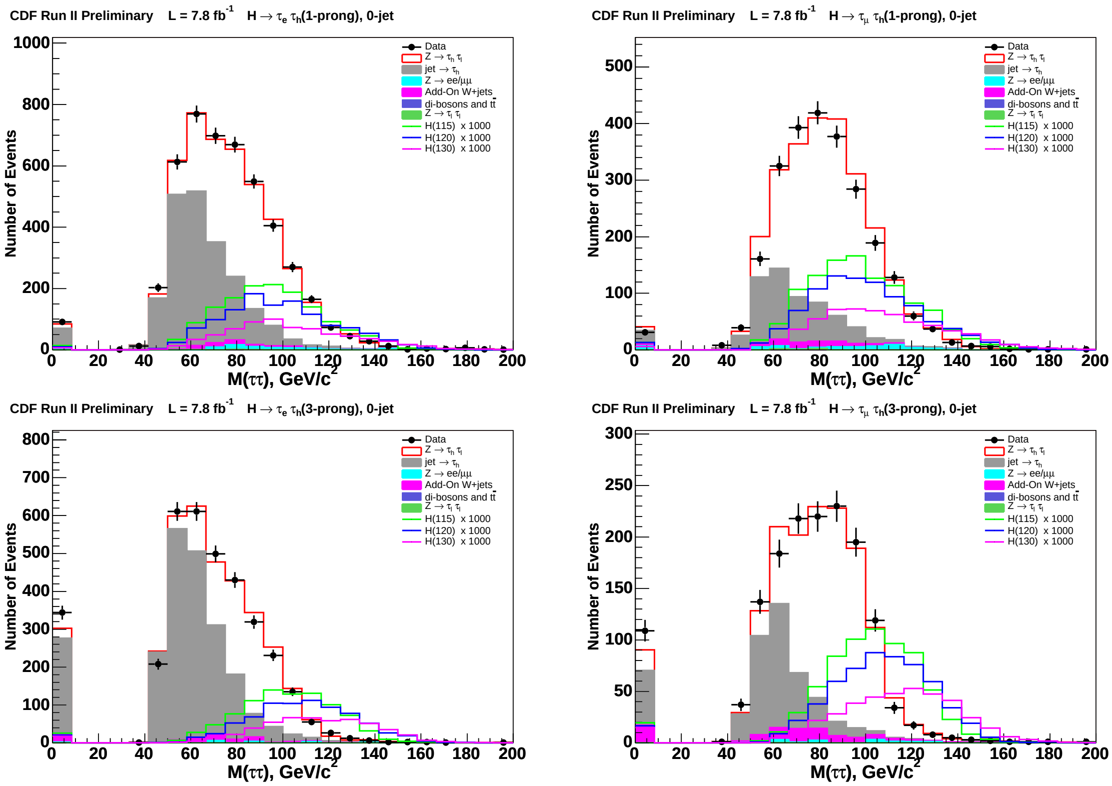

Fig. 60. Di-tau invariant mass, $M_{\tau\tau}$. Events with N_{jet}=0. Full mass range. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

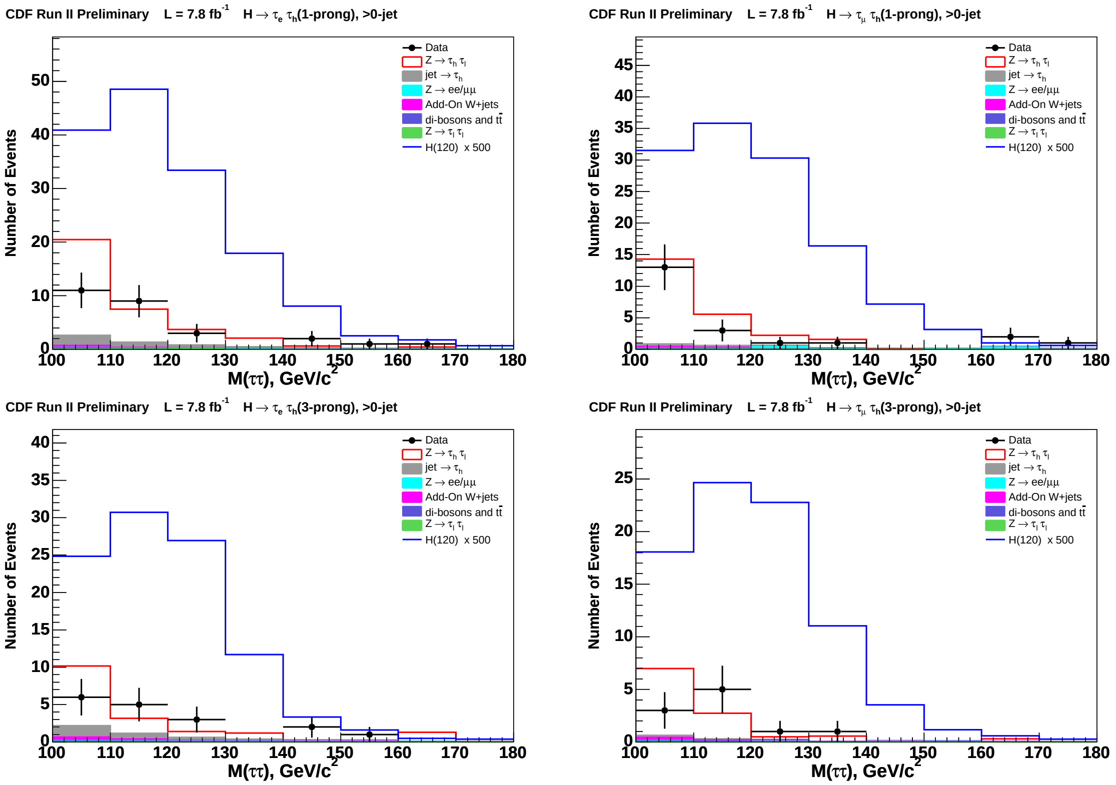

Fig. 61. Di-tau invariant mass, $M_{\tau\tau}$. Events with N_{jet}>0. Signal region. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

Table XXXI. Event yield in the signal region $M_{\tau\tau}$ >100 GeV: $\tau_e\tau_h$(1-prong), N_{jet} >0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h\tau_l$	29.2 ± 1.63 (5.6)
fakes from SS	3.35 ± 1.84 (54.8)
add-on W+jets	0.692 ± 0.181 (26.1)
$Z \rightarrow ll$	0
di-bosons	0.113 ± 0.0113 (10)
$t\bar{t}$	0.798 ± 0.0798 (10)
$Z \rightarrow \tau_l\tau_l$	0.22 ± 0.0135 (6.13)
Total background	34.3 ± 2.47 (7.18)
OS data	27

		Signal expectations			
M_H	ggH	WH	ZH	VBF	
115	0.139	0.0871	0.0481	0.0629	0.337
120	0.125	0.0769	0.0457	0.0597	0.307
130	0.0926	0.0477	0.0299	0.0417	0.212

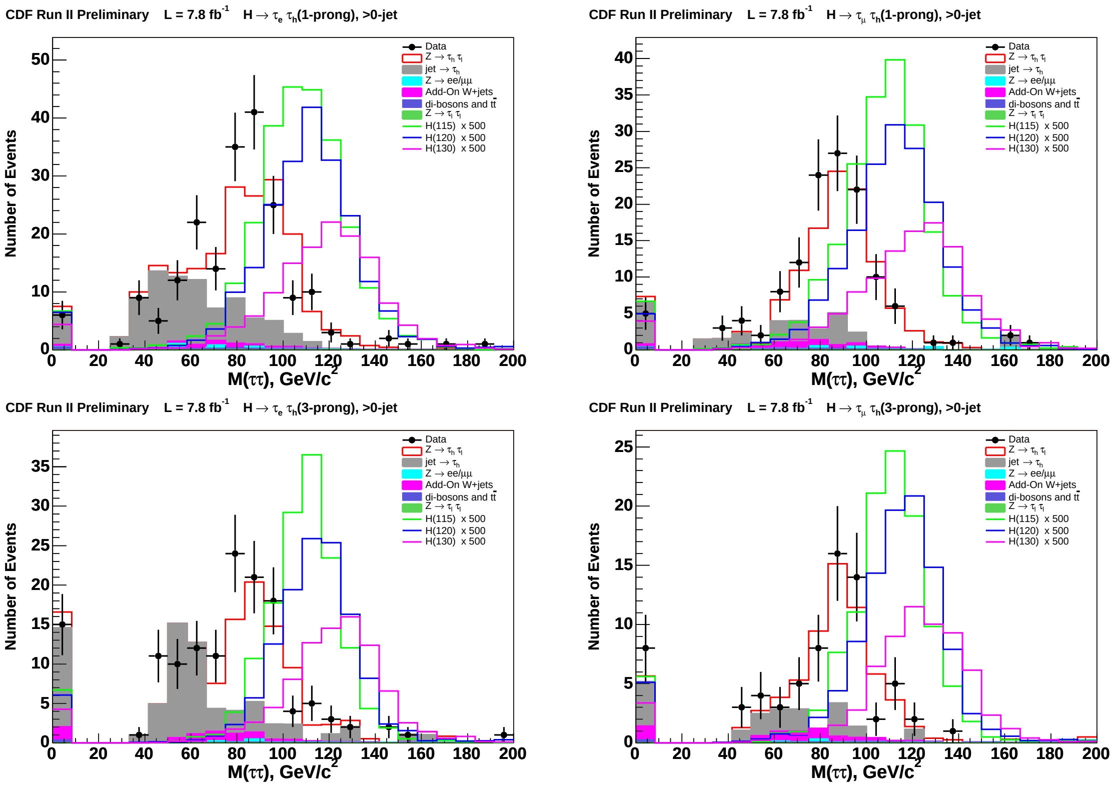

Fig. 62. Di-tau invariant mass, $M_{\tau\tau}$. Events with N_{jet}>0. Full mass range. Left: electron channel, right: muon channel. Top: 1-prong taus, bottom: 3-prong taus.

Table XXXII. Event yield in the signal region $M_{\tau\tau}$ >100 GeV: $\tau_e\tau_h$(1-prong), N_{jet} =0

Background source	Events (uncertainty, %)
$Z \to \tau_h\tau_l$	512 ± 21.4 (4.18)
fakes from SS	45.8 ± 7.09 (15.5)
add-on W+jets	11.9 ± 1.37 (11.4)
$Z \to ll$	28.7 ± 0.193 (0.671)
di-bosons	0.651 ± 0.0651 (10)
$t\bar{t}$	0.0634 ± 0.00634 (10)
$Z \to \tau_l\tau_l$	0.407 ± 0.0172 (4.24)
Total background	570 ± 22.6 (3.96)
OS data	606

		Signal expectations			
M_H	ggH	WH	ZH	VBF	
115	0.53	0.0222	0.00868	0.0144	0.576
120	0.5	0.0186	0.00805	0.0133	0.54
130	0.314	0.011	0.00555	0.00885	0.339

Table XXXIII. Event yield in the signal region $M_{\tau\tau}$ >100 GeV: $\tau_e\tau_h$(3-prong), N_{jet} >0

Background source	Events (uncertainty, %)
$Z \to \tau_h\tau_l$	12.8 ± 0.715 (5.6)
fakes from SS	6.76 ± 2.61 (38.7)
add-on W+jets	0.736 ± 0.152 (20.6)
$Z \to ll$	0
di-bosons	0.0876 ± 0.00876 (10)
$t\bar{t}$	0.732 ± 0.0732 (10)
$Z \to \tau_l\tau_l$	0
Total background	21.1 ± 2.72 (12.9)
OS data	17

		Signal expectations			
M_H	ggH	WH	ZH	VBF	
115	0.0941	0.0513	0.0331	0.0386	0.217
120	0.0873	0.044	0.029	0.0397	0.2
130	0.0665	0.0329	0.0214	0.0298	0.151

Table XXXIV. Event yield in the signal region $M_{\tau\tau}$ >100 GeV: $\tau_e \tau_h$(3-prong), N_{jet} =0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h \tau_l$	196 ± 8.19 (4.18)
fakes from SS	34.8 ± 6.07 (17.4)
add-on W+jets	9.55 ± 1.12 (11.7)
$Z \rightarrow ll$	5.51 ± 0.173 (3.13)
di-bosons	0.804 ± 0.0804 (10)
$t\bar{t}$	0.037 ± 0.0037 (10)
$Z \rightarrow \tau_l \tau_l$	0
Total background	241 ± 10.3 (4.26)
OS data	241

Signal expectations

M_H	ggH	WH	ZH	VBF	
115	0.449	0.0191	0.00701	0.00874	0.484
120	0.462	0.0155	0.00637	0.00974	0.494
130	0.346	0.0118	0.00542	0.0079	0.371

Table XXXV. Event yield in the signal region $M_{\tau\tau}$ >100 GeV: $\tau_\mu\tau_h$(1-prong), N_{jet} >0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h\tau_l$	21.5 ± 1.12 (5.22)
fakes from SS	1.46 ± 1.22 (83.6)
add-on W+jets	0.595 ± 0.156 (26.3)
$Z \rightarrow ll$	0.311 ± 0.0138 (4.44)
di-bosons	0.11 ± 0.011 (10)
$t\bar{t}$	0.735 ± 0.0735 (10)
$Z \rightarrow \tau_l\tau_l$	0
Total background	24.4 ± 1.67 (6.84)
OS data	21

		Signal expectations			
M_H	ggH	WH	ZH	VBF	
115	0.117	0.0669	0.0405	0.0483	0.273
120	0.111	0.0605	0.0348	0.0459	0.252
130	0.081	0.0386	0.0224	0.0315	0.173

Table XXXVI. Event yield in the signal region $M_{\tau\tau}$ >100 GeV: $\tau_\mu\tau_h$(1-prong), N_{jet} =0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h\tau_l$	417 ± 15.2 (3.65)
fakes from SS	21.9 ± 5.38 (24.5)
add-on W+jets	7.85 ± 1.23 (15.7)
$Z \rightarrow ll$	24.2 ± 0.902 (3.73)
di-bosons	0.638 ± 0.0638 (10)
$t\bar{t}$	0.0555 ± 0.00555 (10)
$Z \rightarrow \tau_l\tau_l$	0.476 ± 0.0248 (5.22)
Total background	448 ± 16.2 (3.62)
OS data	444

		Signal expectations			
M_H	ggH	WH	ZH	VBF	
115	0.431	0.0171	0.00777	0.0105	0.466
120	0.4	0.016	0.00684	0.00961	0.432
130	0.3	0.00874	0.00471	0.00697	0.32

Table XXXVII. Event yield in the signal region $M_{\tau\tau}$ >100 GeV: $\tau_\mu\tau_h$(3-prong), N_{jet} >0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h\tau_l$	9.8 ± 0.511 (5.22)
fakes from SS	0.934 ± 0.972 (104)
add-on W+jets	0.767 ± 0.228 (29.7)
$Z \rightarrow ll$	0 ± 0.0216 (10.3)
di-bosons	0.093 ± 0.0093 (10)
$t\bar{t}$	0.511 ± 0.0511 (10)
$Z \rightarrow \tau_l\tau_l$	0
Total background	12.1 ± 1.12 (9.27)
OS data	10

Signal expectations

M_H	ggH	WH	ZH	VBF	
115	0.0714	0.0385	0.0236	0.0312	0.165
120	0.0749	0.0376	0.0217	0.0299	0.164
130	0.0507	0.0251	0.0153	0.022	0.113

Table XXXVIII. Event yield in the signal region $M_{\tau\tau}$ >100 GeV: $\tau_\mu\tau_h$(3-prong), N_{jet} =0

Background source	Events (uncertainty, %)
$Z \rightarrow \tau_h\tau_l$	169 ± 6.16 (3.65)
fakes from SS	6.54 ± 2.65 (40.5)
add-on W+jets	11.4 ± 1.83 (16)
$Z \rightarrow ll$	6.82 ± 0.683 (10)
di-bosons	0.652 ± 0.0652 (10)
$t\bar{t}$	0.0713 ± 0.00713 (10)
$Z \rightarrow \tau_l\tau_l$	0
Total background	187 ± 6.98 (3.73)
OS data	190

Signal expectations

M_H	ggH	WH	ZH	VBF	
115	0.355	0.0139	0.00604	0.00799	0.383
120	0.342	0.0127	0.00607	0.00741	0.368
130	0.265	0.00853	0.00406	0.00628	0.284

F. Systematics Uncertainties

For the backgrounds coming from $Z \to \tau\tau$, $W + jets$ and $Z \to ll$ we derive normalization directly from data. Therefore we use uncertainties on the normalization for each channel calculated in Sec.4 as rate systematics uncertainties (see Tables XIII, XVI, XX). Systematics for the QCD background is taken into account as bin to bin uncertainties from the histogram which correspond QCD background distribution in the signal region. The sources for the rate systematics uncertainties and their values for the $t\bar{t}$ and di-boson backgrounds are given in Table XXXIX. Rate systematics for the signal is shown in Table XL.

Table XXXIX. Systematic uncertainties for the $t\bar{t}$ and di-boson backgrounds.

Source	$t\bar{t}$	di-boson
JES (= 0-jet)	-17	-13
JES (> 0-jet)	-7	8
cross-section	10.0	10.0
PDF model	1.0	1.0

The shape systematics is driven by the Jet Energy Scale (JES). We compare the shapes of the $M_{\tau\tau}$ distributions for the $Z \to \tau\tau$, ggH, WH, ZH and VBF for the nominal value of JES and after applying $\pm 1\sigma$ variation in JES. The shape variation is within statistical uncertainties in each bin of the distributions and have very little effect on the final limits.

Table XL. Systematic uncertainties for the signal.

Source	ggH	WH	ZH	VBF
Luminosity	5.9	5.9	5.9	5.9
JES (=0-jet)	-3.3	-21	-26	-36
JES (>0-jet)	22	10	8	10
σ (=0-jet)	10	5.0	5.0	10.0
σ (>0-jet)	20	5.0	5.0	10.0
PDF model	4.9	1.2	0.9	2.2
ISR/FSR (=0-jet)	-6.3	-2.5	1.4	-5.2
ISR/FSR (>0-jet)	7.3	-2.8	-1.7	-4.3

G. Results

We use MCLimit package [38] to calculate the expected 95% CL upper limits on the Higgs boson production cross section times branching ratio. As input to MCLimit, we use di-tau invariant mass distributions in the signal region, $M_{\tau\tau}$, shown in Figs. 59 and 61. Limits for the channels without jets and for the channels with at least one jet are reported in Tables XLI and XLII and shown in Fig. 63. The combined limit is reported in Table XLIII and shown in Fig. 64.

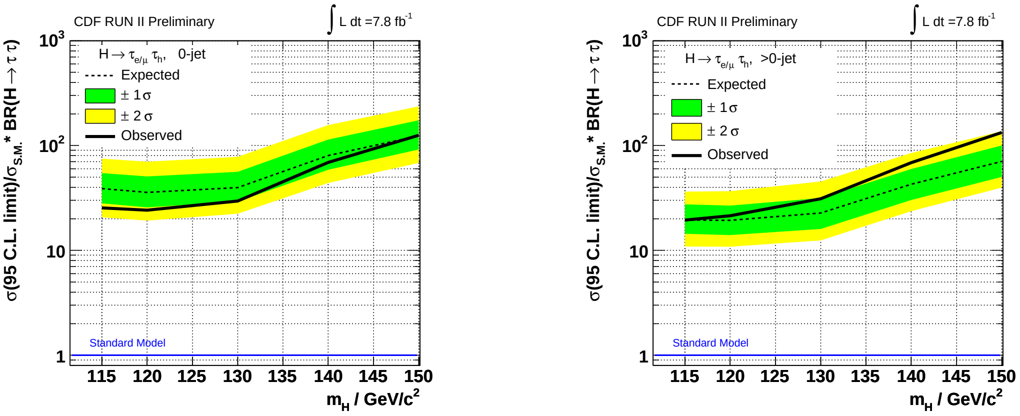

Fig. 63. Expected and observed upper limits on the Higgs boson production cross section times branching ratio in the units of the standard model prediction as a function of the Higgs mass. Left: =0-jet channels, right: >0-jet channels.

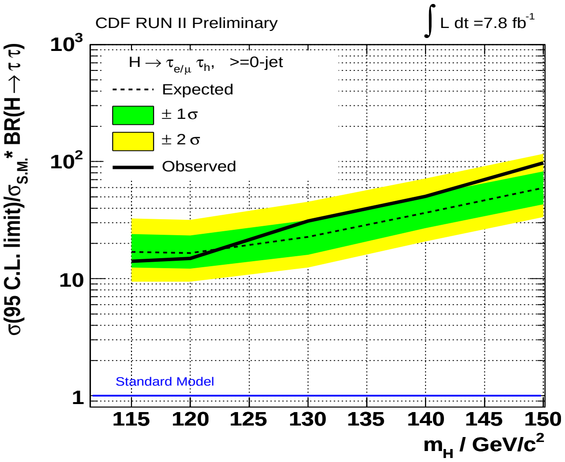

Fig. 64. Expected and observed upper limits on the Higgs boson production cross section times branching ratio in the units of the standard model prediction as a function of the Higgs mass. All channels combined.

Table XLI. Expected and observed upper limits on the Higgs boson production cross section times branching ratio in the units of the standard model prediction. =0-jet channels.

$\tau_{e/\mu}\tau_h$(1- and 3-prong), $N_{jet} = 0$						
M_H	Expected 95% C.L. Limit					Observed 95% C.L. Limit
	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	
115	20.7	28.1	38.8	54.7	75.0	25.4
120	19.2	25.7	35.7	50.8	70.3	24.2
130	22.4	29.2	39.7	56.0	78.0	29.6
140	43.9	58.5	80.2	113.9	156.7	68.9
150	68.3	91.9	123.9	174.3	237.1	125.2

Table XLII. Expected and observed upper limits on the Higgs boson production cross section times branching ratio in the units of the standard model prediction. >0-jet channels.

$\tau_{e/\mu}\tau_h$(1- and 3-prong), $N_{jet} > 0$						
M_H	Expected 95% C.L. Limit					Observed 95% C.L. Limit
	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	
115	10.9	14.4	19.5	27.5	36.4	19.5
120	10.8	14.0	19.4	26.7	36.8	21.3
130	12.5	16.0	22.8	31.7	45.4	31.0
140	23.6	30.3	42.8	59.6	84.8	68.8
150	39.7	50.2	70.7	100.2	136.8	133.0

Table XLIII. Expected and observed upper limits on the Higgs boson production cross section times branching ratio in the units of the standard model prediction. All channels combined.

$\tau_{e/\mu}\tau_h$(1- and 3-prong), $N_{jet} >= 0$						
M_H	Expected 95% C.L. Limit					Observed 95% C.L. Limit
	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	
115	9.4	12.4	16.9	24.0	32.7	14.1
120	9.4	12.2	16.5	23.4	31.7	14.9
130	11.2	14.3	19.3	26.7	37.7	21.5
140	20.8	27.1	36.6	51.8	71.8	50.4
150	33.6	43.2	59.5	82.4	116.4	97.5

CHAPTER VII

CONCLUSIONS

We performed a search for the Higgs boson at the Tevatron collider using CDF data in the channel where the Higgs is produced via gluon fusion, associated production and vector boson fusion and decays to a pair of tau leptons. This channel contributes to the combined sensitivity of the global search for the Higgs in the most difficult, low mass region. While there are many indications that the mass of the Higgs is $115<M_H<150$ GeV/c^2, no single measurement is capable to discover or rule out the Higgs boson in this mass range. However combined sensitivity of several measurements improves with every single channel added to the combination.

A significant improvement to the sensitivity of the $H \rightarrow \tau\tau$ channel has been achieved in this study by introducing two novel techniques and better event categorization. At least a 30% gain in the sensitivity is achieved by using the PPFA method for tau energy measurements and MMC algorithm for di-tau mass reconstruction. This level of improvements in the sensitivity using conventional methods[1] would only be achieved by accumulating and analyzing a ≈1.7 times larger dataset.

The observed data from CDF are found to be consistent with the background only hypothesis. Therefore we set an upper limit on the Higgs boson production cross section times branching ratio for Higgs decay to two tau leptons. The observed and expected limits are evaluated in the mass range from 115 GeV/c^2 to 150 GeV/c^2. We express limit in the units of the SM prediction for the cross section times branching ratio. At M_H=120 GeV/c^2 the observed limit is $14.9\times\sigma_{SM} \times Br(H \rightarrow \tau\tau)$.

---

[1] "Cut-based" analysis using visible $\tau\tau$ mass is assumed. Multivariative techniques were proven to be an alternative option to achieve similar improvements in $H \rightarrow \tau\tau$ channel relying on kinematics [39].

REFERENCES

[1] S. L. Glashow, Nucl. Phys. B **22**, 579 (1961).

[2] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).

[3] A. Salam, *Elementary particle theory: relativistic groups and analyticity.* Nobel symposium No.8, page 367 (1968).

[4] P. Higgs, Phys. Rev. Lett. **12**, 132 (1964).

[5] The LEP Electroweak Working Group, m_H fit result (Accessed on October 20, 2011) [http://lepewwg.web.cern.ch/LEPEWWG].

[6] Tevatron New Phenomena & Higgs Working Group, Summer 2011 combined CDF and D0 results (Accessed on October 20, 2011) [http://tevnphwg.fnal.gov].

[7] ATLAS Collaboration, Report No. ATLAS-CONF-2011-112, 2011.

[8] CMS Collaboration, Report No. CMS-PAS-HIG-11-011, 2011.

[9] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group, and the SLD electroweak and heavy flavour groups, Report No. CERN-PH-EP-2010-095, 2010.

[10] T. Aaltonen *et al.*, (CDF Collaboration), Phys. Rev. Lett. **103**, 101802 (2009).

[11] T. Aaltonen *et al.*, (CDF Collaboration), Phys. Rev. Lett. **105**, 251802 (2010).

[12] T. Aaltonen *et al.*, (CDF Collaboration), Phys. Rev. Lett. **104**, 141801 (2010).

[13] CDF Collaboration, Report No. CDF-NOTE-10485, 2011.

[14] D0 Collaboration, Report No. D0-NOTE-6177-CONF, 2011.

[15] A. Abulencia *et al.*, (CDF Collaboration), Phys. Rev. D **75**, 092004 (2007).

[16] A. Aaltonen *et al.*, (CDF Collaboration), Phys. Rev. Lett. **103**, 201801 (2009).

[17] Fermilab Accelerator Division, Concept Rokie Book (Accessed on October 20, 2011) [http://www-bdnew.fnal.gov/operations/rookie_books/rbooks.html].

[18] R. Blair *et al.*, (CDF Collaboration), Report No. FERMILAB-PUB-96/390-E, 1996.

[19] A. Abulencia *et al.*, (CDF Collaboration), J. Phys. G **34**, 2457 (2007).

[20] F. Abe *et al.* (CDF Collaboration), Report No. FERMILAB-PUP-88-25-E, 1988.

[21] CDF Collaboration, Physics Groups (Accessed on October 20, 2011) [http://www-cdf.fnal.gov/CDForg/Physics_Groups.html].

[22] D. Buskulic *et al.*, (ALEPH Collaboration), Nucl. Instrum. Methods A **360**, 481 (1995).

[23] J. C. Brient and H. Videau, arXiv:hep-ex/0202004v1 (2002).

[24] A. Connolly, Report No. FERMILAB-THESIS-2003-45, 2003.

[25] CMS Collaboration, Report No. CMS-PAS-PFT-08-001, 2008.

[26] R. Brun, R. Hagelberg, M. Hansroul and J.C. Lassalle, Report No. CERN-DD-78-2-REV, 1978.

[27] T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, G. Miu, S. Mrenna, and E. Norrbin, Comput. Phys. Commun. **135**, 238 (2001).

[28] A. Elagin, P. Murat, A. Pranko, A. Safonov, Nucl. Instrum. Methods A **654**, 481 (2011).

[29] H.L. Lai *et al.*, Eur. Phys. J. C **12**, 375 (2000).

[30] G. Aad *et al.*, (ATLAS Collaboration), Report No. CERN-OPEN-2008-020, 2008.

[31] G.L. Bayatian *et al.*, (CMS Collaboration), J. Phys. G **34**, 995 (2007).

[32] R. Decker, S. Jadach, M. Jezabek, J.H. Kuhn, and Z. Was, Comput. Phys. Commun. **76**, 361 (1993).

[33] R.K. Ellis, I. Hinchliffe, M. Soldate and J.J. Van der Bij, Nucl. Phys. B **297**, 221 (1988).

[34] T. Aaltonen *et al.*, (CDF Collaboration), Phys. Rev. D **82**, 052005 (2010).

[35] D. Acosta *et al.*, (CDF Collaboration), Phys. Rev. D **71**, 032001 (2005).

[36] S. Drell, T.-M. Yan Phys. Rev. Lett. **25**, 316 (1970).

[37] E. James, T. Junk, Report No. CDF-NOTE-10474, 2011 (unpublished).

[38] T. Junk, Report No. CDF-NOTE-8128, 2006 (unpublished).

[39] CDF Collaboration, Report No. CDF-NOTE-10625, 2011.

VITA

Andrey Lvovich Elagin [redacted] [redacted] He attended Moscow Institute of Physics and Technology from 1999 to 2005. He received BS in physics in 2003 and MS in physics in 2005. From 2002 to 2006 he was employed as a juniour reseacher with Joint Institute for Nuclear Research in Dubna where he conducted research within ATLAS, C2GT and FCAL collaborations and with the HARP-CDP group. He joined Texas A&M University in 2006 where he began working within the CDF collaboration. He received PhD in physics in 2011. His permanent mailing address is Fermilab-CDF-MS318, P.O. Box 500, Batavia, IL 60510.

The typist for this dissertation was Andrey Lvovich Elagin.