THE UNIVERSITY OF CHICAGO

MEASUREMENT OF THE TOP QUARK MASS AT CDF USING THE TEMPLATE METHOD IN THE LEPTON+JETS CHANNEL

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
JAHRED ADELMAN

CHICAGO, ILLINOIS
JUNE 2008

ACKNOWLEDGEMENTS

Wow. Countless people deserve praise on these pages. Apologies in advance to any friends or colleagues who might inadvertently have been left out. Thanks and appreciation must begin with my family. To my parents - I love you both very much. I am endebted to you for raising me, taking care of me after my surgery, and putting up with your pain-in-the-ass, know-it-all son. To my mom - thanks for supporting me in an endeavor that took your only son away from the East coast and now threatens to move him to a new continent. All your packages of goodies and phone calls of encouragement and love were much appreciated. To my dad - I did it. I finished. Now you can stop asking me when I will graduate. Thank you for teaching me the alphabet on the keyboard. Thank you for pushing me to find a job that keeps me happy. Thank you for help with my apartment. Just don't ask me to take up golf.

To my sister Ashley - many thanks for answering your phone and swapping back-and-forth complaints about life, mom or dad (note to mom and dad - you weren't supposed to read this far). Ips - I love you, and look forward to being the poor physicist brother of a famous actress. And to Emily and Cassidy - with me off to Europe, you will have to do a better job of answering my IMs so I can say I love you before you ignore me. Particular thanks to Emily for asking, many years ago at dinner in Harry Caray's, how many atoms there are in the universe, and then expressing shock when I didn't know the answer.

Sometime after learning the alphabet on a QWERTY keyboard and before my life as a CDFer, I learned to love physics in Mr. Tomes' high school classes. A few years later found Greg Landsberg offering me the opportunity to learn particle physics. Both of them deserve thanks.

As for the past 5-6 years: To all of my colleagues on CDF - it has been an honor and a pleasure to work with you all. It's still hard for me to imagine putting together hundreds of people on the same multi-purpose physics experiment and getting out any results. Yet somehow it works, at least most of the time. Thanks to all the top and top mass conveners over the years. Evelyn, Taka, Robin, Kristin, Florencia, Igor, Young-Kee, Eva, Un-Ki, Daniel, Nick, George and Erik - you made the groups work as a team, and somehow things got done, and done correctly, no less.

Many thanks also to my friends and co-workers on SVT - it was a blast maintaining, upgrading and understanding such a unique piece of hardware. I learned a tremendous amount, and had a ton of fun in the process. Alberto - I look forward to working with you as a colleague on ATLAS, though I can't promise to ever pick up Go as a hobby. Stefano - I would write the only word of Italian you taught me, but I'm afraid the UC dissertation office wouldn't be too pleased with me. In any case, I'm still planning on taking you up on your offer to spend time in Rome. Special praise is owed to Bill for doing his best to transfer some tiny fraction of his immense knowledge of the trigger system to a confused, untrained new graduate student. Many thanks also to Roberto, Benedetto, Paola, Paolo, Barbara, Taichi, Marco and Alex for being good friends and good co-workers.

In my tenure as a top mass metrologist I had the good luck to work with many talented individuals. The folks over at Toronto, Jean-Francois, Pekka and Sebastian, taught me a lot and were great people to work with. Similarly for Adam, over at Berkeley but now also at Toronto. Tom was a source of help, patience and sarcastic dry humor in our time working on the kinematic fitter, as well as every moment that I

bugged him over IM about event selection issues and backgrounds (which was often) and our colleagues (which was just as often). It would also be a shame for me to forget to thank the beams division for delivering the 2.5 fb^{-1} or so of $p\bar{p}$ collisions that made this analysis possible at all. Countless people at Fermilab directly and indirectly contributed to this analysis.

My time at the UofC has seen me work, collaborate and share cookies and lousy coffee with many smart, hard-working, friendly colleagues. The post-docs at Chicago are, without exception, excellent. Taka, Bruce, Bill, Kohei, Steve, Dan and Hyun Su were always good for a laugh and a smile. Ivan was good for advice and discussion of SVT, politics or basketball. And Vadim was and is an exceptional resource not only for physics, but also for how to live and enjoy life to its fullest, including the best places to eat and the best bottles to drink. Francesco, Pedro, Shawn and Sasha were great to have around in group meetings, in my office, and in and around the halls of HEP. Erin always managed to finish her problem sets before the rest of us would even look at our assignments. Un-Ki is a good friend and an exceptional resource, and did a great job leading our group through its early PRD. I can still remember the day Satomi came to HEP 213 to learn how to get started on CDF knowing absolutely zero C or C++, and having no idea how to get a list of files in a UNIX directory. A year or two later, and I'm still amazed at how hard she worked and at our wonderful top quark width limit. I'm also very lucky to have as friends two of the most important people to the UofC HEP group. I remember that when I arrived in Chicago, I was told that if I befriended Nobuko, I would have no problems in my graduate career. And I got very lucky to have an office next to Aspaspia so that I could chit-chat with her at random times of the day when ROOT was keeping me down. Many thanks to both of them. Thanks also to Mary for keeping the computer systems at Chicago up

and running, and for always remaining kind and helpful while dealing with endless, overwhelming requests from all of us. Thanks to Carla, both for being a friendly face as well as for interfacing between us and Mary when we were asking for too much. And thanks to Mircea, Harold, Tang, Mark and the other folks over in the e-shop for helping me out.

By sheer luck, I could not have asked for a better officemate at Chicago than Erik. Few people in our field have the ability to step back from a problem and find workable solutions, to think for themselves, or to explain themselves clearly. Erik does all of these things. He is also extremely patient, and shares my sense of joy and amusement at mocking the Tribune website. Erik - I wish you the best of luck wherever you end up after Chicago. Similarly, I can't imagine a better analysis partner-in-crime than Wojtek. We spent many a long and late night together in HEP, evenings that typically ended in me going home and yelling at him to get some sleep. We managed to bless several times without losing too many brain cells, and I'd like to think we even had a bit of fun in the process. Wojtek - I'll see you on the other side of the Atlantic.

The HEP program at the UofC is blessed with a group of dedicated, well respected, friendly faculty. It has been a pleasure to work with them all, in particular those on CDF. Henry has an unlimited supply of good stories to share - and he loves sharing. He also has an uncanny ability to ask the interesting questions, and hosts wonderful dinner parties. Young-Kee is, well, Young-Kee. It's not surprising that she is known on a first-name basis, sort of like Prince. Nobody I have ever met works as hard as she does. Most important of the faculty, though, has been my advisor, Mel, who initially paid my salary, even when I was completely useless. Over the years, if I've learned one important thing about this field, it's that Mel is always right. This is

not him saying this, and I don't mean it to imply that he is cocky or arrogant, but in my experience, Mel tends always to have the correct answer. I could not imagine a better advisor to back me up, and I thank him for giving me the opportunity to work with him.

My friends in Chicago kept me happy and even a bit social when work threatened to do otherwise. Collin has been a great drinking buddy, fellow student and friend over the years. Who could have guessed when he expressed interest in watching the All-Star game over beer way back in the summer of 2002 that we would almost end up owning property together? Meredith loves to bake and share cookies and has been a great friend, even if she has ridiculously heavy furniture and never used that sofa. Dave always seemed up for a night out instead of a night in. With him in the states and me overseas, the huge gap between us in baseball knowledge will grow ever-larger. Josh and Beth are two of the most generous and fun people that I know, and I still thank them for putting up with me as a houseguest for an entire month. I was thrilled when Joe and Amber moved to Chicago. It's been wonderful once again having my best friend from high school living nearby, at the very least so we can swap never-ending stories about our crazy families. I definitely plan on crashing at their place in Germany next year. Extra thanks to Joe and Amber for introducing me to Shawna. Joe, I am sorry I didn't do quite as good a job 10 years ago for our prom. Bryce comes rather close to being the nicest guy that I know. He also makes a mean meat sauce and bananas foster, though his Guitar Hero skills are a bit lacking. Many thanks to Mike and Arlise, and Colin and Nora for being good friends. Thanks to Keith, who is probably very cold right now, for being a good drinking partner. Thanks to Jenny for being a good roommate for a year. And thanks to Matt and Doug for being good buddies, both in Chicago and in Madison. I admit that the

combination of Doug and Amsterdam is both exciting and scary at the same time. To Heather - thanks for putting up with an overworked, pager-carrying, oft-scruffy graduate student. I wish you the best of luck in graduate school. And to Shawna - thanks for the use of your shower, as well as for helping me smile and laugh in my last year in Chicago. Knowing that you are staying on this side of the globe makes moving overseas that much harder. I will miss you.

I will end by additionally thanking Randy Mosher, Ray Daniels, Jeff Sparrow, Jim Koch, Sam Calagione, John and Greg Hall, Ron Jeffries, Michael Jackson, Todd and Jason Alstrom, Charlie Papazian, Nick Floyd, Larry Bell, and Dan and Deborah Carey. I'm sure none of them will ever read this, but thanks for keeping a graduate student sane every now and then.

ABSTRACT

A measurement of the top quark mass in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV is presented. The analysis uses a template method, in which the overconstrained kinematics of the Lepton+Jets channel of the $t\bar{t}$ system are used to measure a single quantity, the reconstructed top quark mass, that is strongly correlated with the true top quark mass. In addition, the dijet mass of the hadronically decaying W boson is used to constrain *in situ* the uncertain jet energy scale in the CDF detector. Two-dimensional probability density functions are derived using a kernel density estimate-based machinery. Using 1.9 fb^{-1} of data, the top quark mass is measured to be $171.8\ ^{+1.9}_{-1.9}$ (stat.) $\pm$ 1.0 (syst.) GeV/c^2.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

To first order, every piece of matter around us is made up of just 3 fundamental particles - the up and down quarks that form protons and neutrons, and electrons. Upon closer examination, however, this description is found to be insufficient. The electroweak force-carrying $W^{\pm}$ and Z bosons and the photon are necessary to describe nuclear decays and scattering, as well as all of electromagnetism. An electron neutrino is emitted in the β-decay of heavy nucleii, and gluons are required to hold together the up and down quarks that make up protons and neutrons, as well to combine the nucleons into nucleii. Yet even this picture is incomplete. Four additional quarks, four additional leptons, and four additional neutrinos are essential pieces of the Standard Model (SM) of particle physics. And ever then the SM is incomplete, requiring an additional, as-of-yet undiscovered but much-sought-after particle, the Higgs boson, to complete the theory.

The top quark is by far the most massive of all these fundamental particles, some 5 orders of magnitude more massive than the electron and 4 orders of magnitude more massive than the up and down quarks that constitute the physical world that we know. This unique property of top quarks - a huge mass that sets it apart from all other point particles - leads to an important role for the top quark in theoretical predictions. In particular, the masses of the top quark, the $W^{\pm}$ bosons and the Higgs boson are not three independent parameters of the SM, and the relationship between the three masses is predicted by the theory.

High-energy experimentalists love to push the SM for ever-more accurate predictions with the hope that at some point it breaks, indicating the discovery of new physics. Precision measurements of the mass of the top quark are thus a priority of the Tevatron program. This analysis is based on techniques for extracting the top quark mass first used in Run I measurements at the Tevatron, but includes many new tricks and features that we've learned along the way to mine as much information as possible from the data. In particular, we use the *in situ* hadronic W resonance to calibrate the largest systematic in the detector, the uncertain calibration of the calorimeter response to hadronic particles. New to this analysis is a set of non-parametric techniques that allow for fully two-dimensional probability density functions without making *ad hoc* assumptions about the shape of distributions. And of course, we also make use of the very large datasets available to Run II measurements at the Tevatron.

CHAPTER 2

THEORY

2.1 The Standard Model

Remarkably successful in its predictive ability and power to endure countless fleeting signs of new physics, the SM represents physicists' complete understanding to date of all the fundamental particles observed in nature. The SM accounts for the fermions and their interactions, which are governed by the force-carrying bosons. Each of the three generations of fermions (spin-1/2 particles) are divided into two classes - leptons and quarks. Each lepton generation consists of an electrically charged particle (such as the electron) and its associated neutrino (such as the electron neutrino). Quarks generations also come in doublets, with one unit of electrical charge separating the fractional charges of the quarks (such as $+\frac{2}{3}e$ and $-\frac{1}{3}e$ for the up and down quarks, respectively). The strong interaction of the quarks is described by quantum chromodynamics (QCD) and mediated by the spin-1 force-carrying gluons. Other interactions are governed by the electroweak force-carrying gauge bosons, the charged $W^{\pm}$, the neutral Z, and the neutral photon (γ). Equations 2.1.1 and 2.1.2 show the 3 generations of leptons and quarks. Not shown are their antiparticles.

$$\begin{pmatrix} \nu_e & \nu_\mu & \nu_\tau \\ e & \mu & \tau \end{pmatrix}; \tag{2.1.1}$$

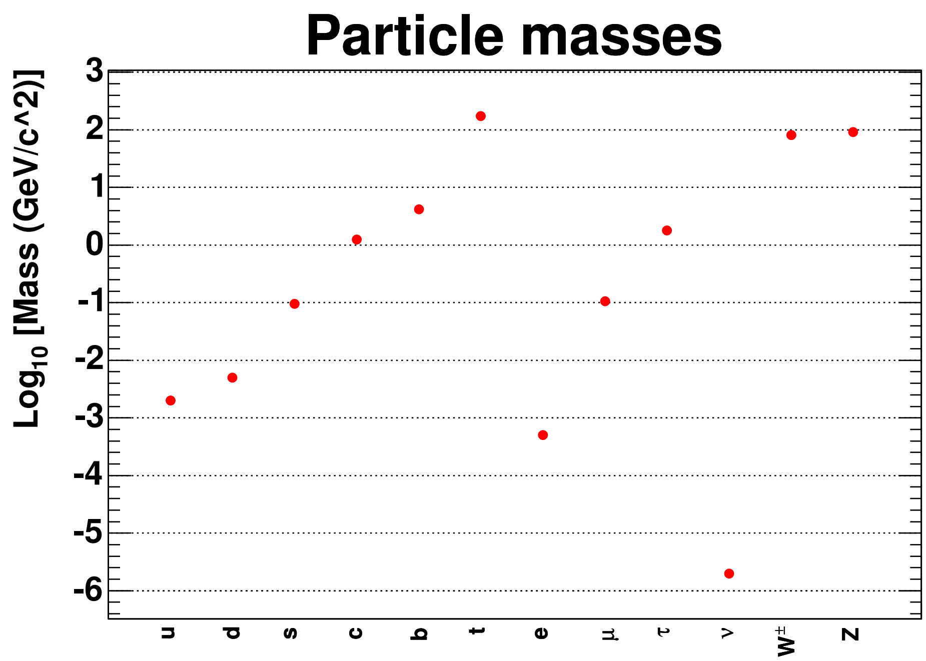

Figure 2.1: Masses of the known fundamental particles. The mass of the neutrino shown is the upper limit on ν masses. The masses of the light quarks are not well measured. Not shown are the massless photon and gluon.

$$\begin{pmatrix} u & c & t \\ d & s & b \end{pmatrix}. \tag{2.1.2}$$

2.2 The Top Quark

With a mass of roughly 172.0 GeV/c^2 [1], the top quark (t) is by far the most massive fundamental object observed to date in nature, as can be seen in Figure 2.1. It is almost twice as heavy as the second-most massive point particle (the Z boson), and 40 times as massive as its isospin partner, the b quark (b). At least 4 orders of magnitude separate the mass of the top (M_{top}) and the mass of the corresponding

3rd generation, weak isospin +1/2 lepton (ν_τ). Perhaps most amazingly, the top quark is almost as massive as a single gold nucleus, a complex object containing 79 protons and 118 neutrons tantalyzingly close in size to a mesoscopic length scale.

Though predicted to exist almost immediately after the discovery of the bottom quark in 1977 [2], the top quark was not discovered for 17 more years. The CDF and D0 collaborations jointly announced discovery of the top quark in 1995 [3, 4], but it was not until the recent high luminosity and large datasets of Run II at the Tevatron that precision measurements of the top quark were possible.

Many review articles of top quark physics exist in the literature. Two particularly good ones are [5, 6].

2.3 The Higgs boson

The last remaining piece of the SM is the as-of-yet undiscovered Higgs boson, which provides an explanation for how particles obtain mass. A very useful, short introduction to the interaction and interplay between the top quark and the Higgs boson can be found in [6]. The SM Lagrangian contains an $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry that forbids any explicit mass terms. Such a term for a particle of mass m connects left-handed and right-handed fermion fields, and would look like:

$$\mathcal{L}_{\mathrm{mass}} = -m\overline{f_L} f_R + \mathrm{h.c.} \tag{2.3.1}$$

This term is not allowed because the left-handed and right-handed fermion fields have different quantum numbers under the SU(2) and U(1) symmetries. The hypothetical Higgs field (ϕ) breaks the SU(2) symmetry, and in the process generates masses [7] for fermions and the weak gauge bosons, while keeping the photon massless. The Higgs field is a complex scalar doublet, and has a Lagrangian that looks like:

$$\mathcal{L}_{\text{Higgs}} = (D^{\mu}\phi)^{\dagger} D_{\mu}\phi + \mu^2 \phi^{\dagger}\phi - \lambda(\phi^{\dagger}\phi)^2. \tag{2.3.2}$$

The first term in Equation 2.3.2 contains the kinetic energy and gauge interaction terms of the Higgs field. The second and third terms give rise to a potential that looks like the one in Figure 2.2. The Higgs potential does not have a minimum when the field is zero, but instead at:

$$< \phi >= \left(\begin{array}{c} 0 \\ \frac{\mu}{\sqrt{2\lambda}} \end{array} \right) \equiv \left(\begin{array}{c} 0 \\ \frac{v}{\sqrt{2}} \end{array} \right), \tag{2.3.3}$$

where v is called the vacuum expectation value of the field.[1] The gauge interactions in the first term of Equation 2.3.2 provide the mass terms of the gauge bosons, as well as those of the fermions. Writing the SU(2) gauge coupling as g and the U(1) coupling as g', the gauge boson masses become:

$$m_W = g\frac{v}{2}; \tag{2.3.4}$$

$$m_Z = \sqrt{g^2 + g'^2}\frac{v}{2}, \tag{2.3.5}$$

---

1. With the exception of the $\phi^{\dagger}\phi$ Higgs potential term, the SM Lagranian consists only of dimension-4 operators. The parameter in front of this $\phi^{\dagger}\phi$ term (μ) is the only coupling in the SM with any dimensions, and thus it should set the scale for all masses in the SM [6].

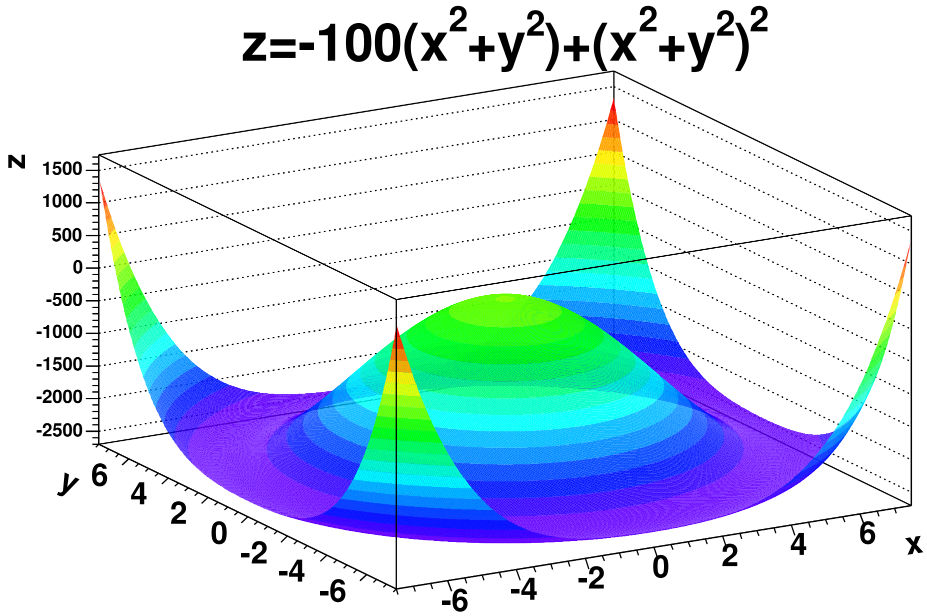

Figure 2.2: A Mexican hat potential. The minimum is not at $r = 0$ but instead at a set of points with $r = \sqrt{x^2 + x^2} = \sqrt{50}$.

with the photon remaining massless. A plethora of precise electroweak data provide measurements of g, g' and v. One useful set of such measurements [1] is as follows:

- The value of the fine structure constant, α, which is measured in $e^{\pm}$ anomalous magnetic moments, the quantum Hall effect, and other low-energy experiments; $\alpha = \frac{1}{4\pi}\frac{g^2 g'^2}{g^2+g'^2} = 1/137.03599911$.

- The Fermi coupling constant (G_F), which is determined in muon lifetime experiments; $G_F = \frac{1}{\sqrt{2}v^2} = 1.16637 \times 10^{-5}$ GeV^{-2}.

- The Z boson mass, measured at LEP; $M_Z = 91.1876$ GeV$/c^2$.

Using these measurements, v is found to be ≈ 246 GeV$/c^2$.

The Higgs field is a complex scalar doublet, so it has 4 degrees of freedom. Two degrees of freedom are used in giving mass to the two $W^{\pm}$ bosons, and one degree of freedom is used in giving mass to the Z boson. This leaves one degree of freedom unused, so that fluctuations in the field can produce a particle called the Higgs boson, with a mass (m_H) on the order of the electroweak scale. As another way to see this, consider perturbations h around the minimum of the field, so that v$\rightarrow$ v+h. An expansion of the potential terms (i.e. the last 2 terms) in Equation 2.3.2 gives

$$\mathcal{L}_{\text{V-higgs}} = \frac{\mu^4}{4\lambda} - h^2\mu^2 - \lambda v h^3 - \frac{\lambda h^4}{4}. \tag{2.3.6}$$

The second term in Equation 2.3.6 is identified as the mass term in a Lagrangian, indicating that $m_H = \sqrt{2}\mu = \sqrt{2\lambda}v$, so that, up to some self-coupling constant, the Higgs mass is determined by v.

At leading order, the mass of the W boson is predicted from precision electroweak observables:

$$M_W^2 = \frac{1}{4}g^2v^2 = \frac{1}{2}M_Z^2(1 + \sqrt{1 - \frac{4\pi\alpha}{\sqrt{2}G_F M_Z^2}}). \tag{2.3.7}$$

A simpler way to express M$_W$ is the definition of the Weinberg angle, θ_w, such that

$$M_W = M_Z \cos\theta_w. \tag{2.3.8}$$

Through contributions in loop diagrams, the Higgs boson can affect the properties of other particles, including the mass of the W boson. In addition, the top quark is so heavy that loop diagrams associated with it also contribute to the mass of the W. Figure 2.3 shows these two diagrams, which have the effect of modifying the W mass:

$$M_W \to \frac{M_W}{\sqrt{1-\Delta r}} \tag{2.3.9}$$

The top quark contribution to Δr,

$$(\Delta r)_{\text{top}} \approx -\frac{3G_F \mathrm{M_{top}}^2}{8\sqrt{2}\pi^2 \tan^2\theta_w}, \tag{2.3.10}$$

moves the mass of the W in the opposite direction as the Higgs boson contribution:

$$(\Delta r)_{\text{Higgs}} \approx \frac{11 G_F M_Z^2 \cos^2\theta_w}{24\sqrt{2}\pi^2} \ln\frac{m_H^2}{M_Z^2}. \tag{2.3.11}$$

Contributions to the observed W mass from diagrams involving the Higgs are proportional to $\ln(m_H)$, and contributions from the top are proportional to $\left(\mathrm{M_{top}}\right)^2$. Thus, a precision measurement of the W boson mass [8] can be turned around to make a statement about the expected mass of the Higgs, given a top quark mass. Unfortunately, the weak $\ln(m_H)$ Higgs mass dependence makes very precise measurements of $\mathrm{M_{top}}$ and M_W necessary in order to make a meaningful statement about m_H. Standard Model Higgs bosons with masses below 114 GeV/c^2 are ruled out by direct searches at LEP [9], but global fits to the electroweak data prefer light Higgs masses [10, 11]. Figure 2.4 shows the consistency of different Higgs masses in the SM, as indicated by shifts in the goodness of the global fit to electroweak observables and the top quark mass. Standard Model Higgs bosons above $\approx$ 160 GeV/c^2 are ruled out at the 95 % confidence level. Figure 2.5 shows the 68% confidence level contour

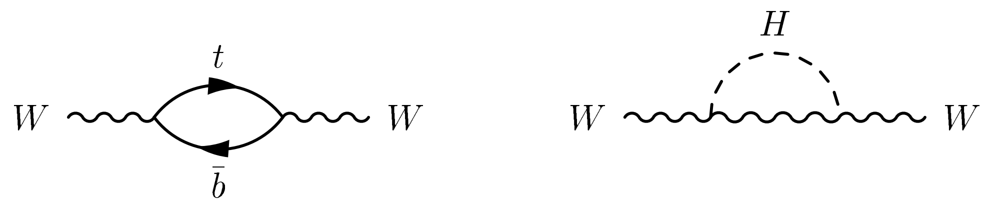

Figure 2.3: Loop diagrams involving virtual Higgs bosons and top quarks that contribute to the observed W mass.

for the Standard Model in the $\mathrm{M_{top}}$-M_W plane.

It seems, then, that the SM Higgs boson may be right around the corner. When and if something like the Higgs boson is found in the next few years, a very precise measurement of the top quark mass will be necessary to confirm that experiments are seeing the SM Higgs boson, and not something else. And if the Higgs is not found, the heavy mass of the top quark will make precision measurements throughout the top quark sector necessary to help theorists disentangle models of new physics [13]. As an example, Figure 2.6 shows the status of the SM in the $\mathrm{M_{top}}$-M_W plane, along with one region of applicable phase space (MSSM) for supersymmetry, an extension of the SM in which every fermion has a boson partner, and likewise, every boson in the SM has a fermion partner.

2.4 An interesting quark

The massive nature of the top quark makes it interesting to study in its own right. Due to the large phase space available to them, top quarks in the SM decay with an extremely short lifetime, $\sim 4 \times 10^{-25}$ s, well before the top has a chance to hadronize, as processes in QCD have a characteristic time scale $\tau_{\mathrm{QCD}} \sim 10^{-24}$ s. Thus, unlike lighter quarks that bind into hadrons before decaying, top quarks transmit spin in-

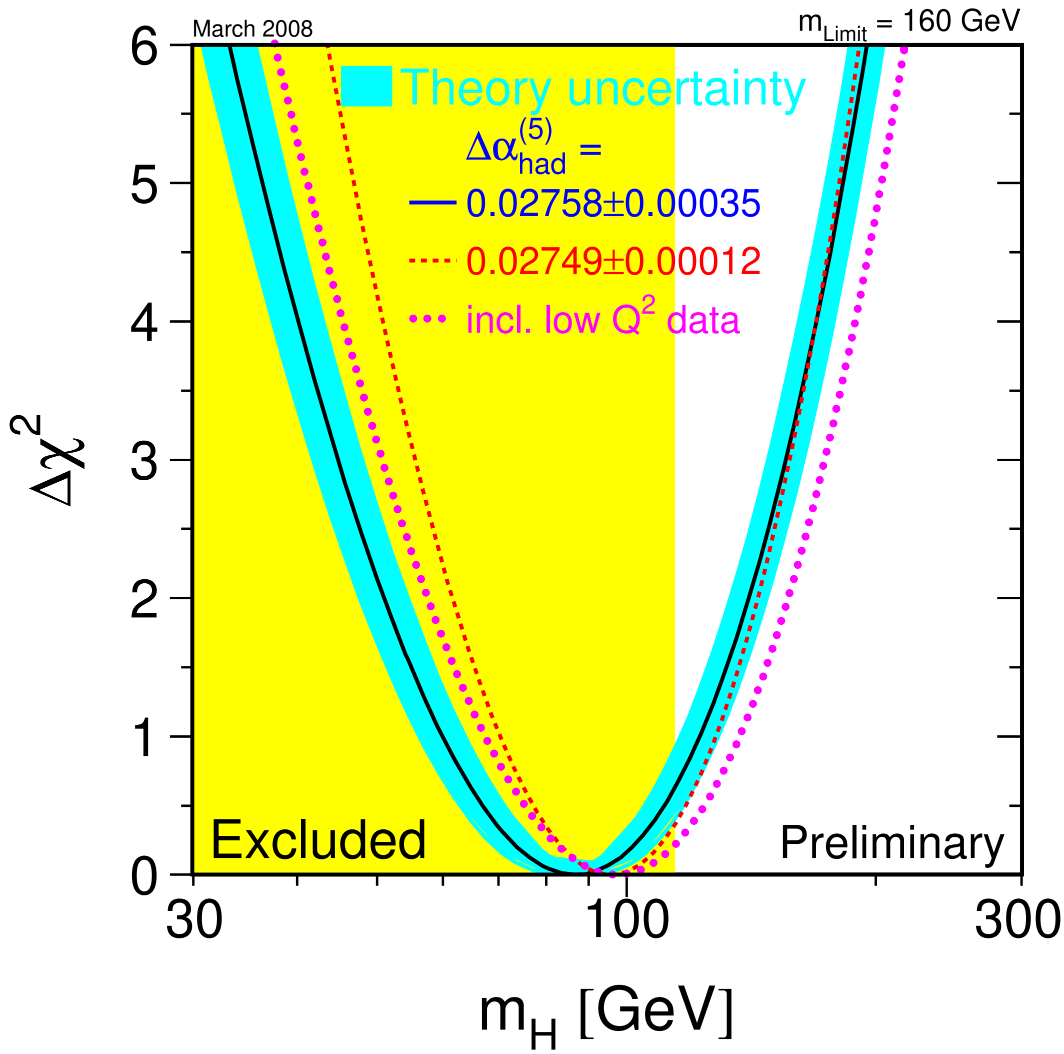

Figure 2.4: Shifts in the goodness-of-fit to electroweak and top quark observables as a function of the Higgs mass [12]. The global best fit to high-energy (high-Q^2) data in the SM is indicated by the dark blue line, with a theoretical uncertainty indicated by the turquoise band. The magenta curve includes additional measurements from low-energy (low-Q^2) experiments. The red dotted curve uses a theoretical prediction instead of an experimental measurement for the contribution to the photon vacuum polarization from light quarks. The 95% upper limit on the Higgs mass of 160 GeV/c^2 comes from the intersection of $\Delta\chi^2 = 2.7$ with the right-hand side of the turquoise band.

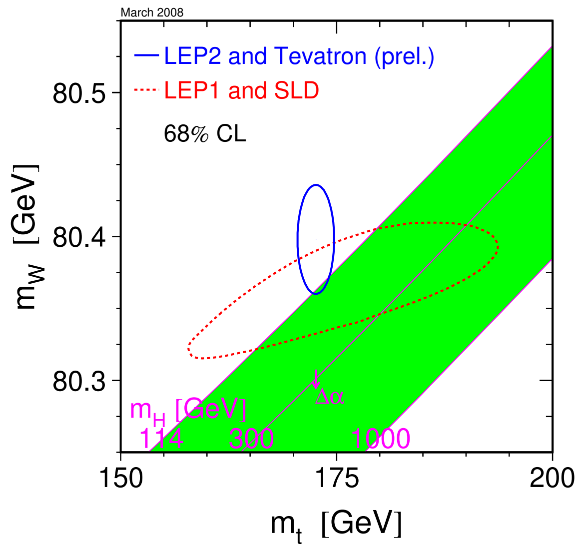

Figure 2.5: Consistency with the SM at a 68 confidence level in the M_{top}-M_W plane [12]. LEP1 and SLD data refer to measurements in e^+e^- machines at a center-of-mass energy at the Z pole mass, such that the mass of the top quark was inferred indirectly. LEP2 took place above the Z pole mass. Tevatron data include not only top quark mass measurements, but also a W mass measurement.

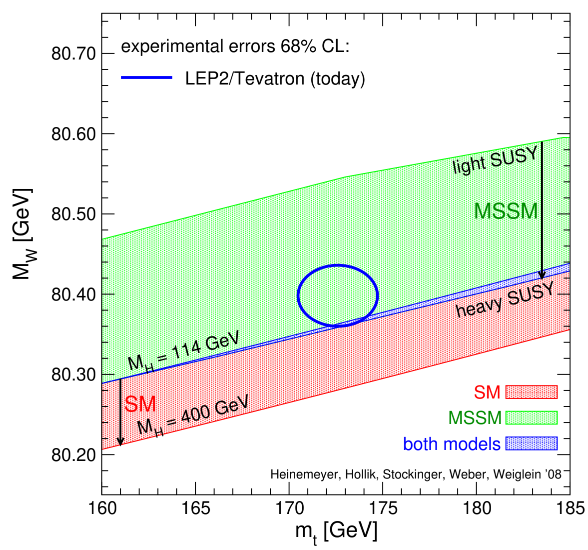

Figure 2.6: Consistency of both the SM and the MSSM version of supersymmetry at a 68% confidence level in the M_{top}-M_W plane. Taken from [14] as updated from [15, 16].

formation to their decay daughters, and are the only quarks for which it is possible to measure directly the bare mass, and not just the mass of hadrons.

The heavy nature of the top quark, with a mass near the electroweak scale, also suggests that beyond-SM physics may first appear in the top sector [13], making a precision measurement of M_{top} important for future physics searches. Finally, with the turn-on of the LHC, the Tevatron will cease to be the only place on the planet producing top quarks. Top quarks will be produced copiously at the LHC, and may become a standard candle used to help calibrate the ATLAS and CMS detectors [17, 18].

2.5 Top phenomenology

At the Tevatron, top quarks are produced predominantly in $t\bar{t}$ pairs via a gluon splitting to $t\bar{t}$. Roughly 85% of these events are produced via quark-antiquark anhiliation, with the remaining 15% coming from gluon-gluon fusion.[2] See Figure 2.7 for Feynman diagrams of the two main production mechanisms. Due to the large mass of the top quark, the pair production cross-section at the Tevatron is strongly dependent on M_{top}, as it becomes harder and harder with higher mass to pull partons with the necessary energy out of the colliding protons and antiprotons. The pair-production cross section is roughly 7 pb [19]. Production of single top quarks through the electroweak interaction occurs at a lower rate and has a much higher background than $t\bar{t}$ production [20].

The primary mechanism for flavor-changing processes in the SM is interaction with a W boson. The CKM matrix describing interactions between W bosons and quarks [21] surpresses the relative branching ratios of top quark decay to $W + [s, d]$ quarks

2. This is in contrast to the LHC, where the majority of $t\bar{t}$ events are initiated by gluons.

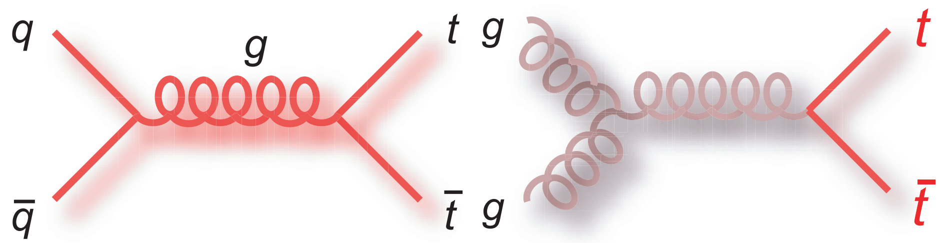

Figure 2.7: The two primary mechanisms for $t\bar{t}$ production. In the diagram on the left, a quark and an anti-quark annihilate into a gluon, which then splits into a $t\bar{t}$ pair. On the right, two gluons annihilate into a third gluon, which again produces the $t\bar{t}$ pair. The diagram on the left dominates at the Tevatron.

with respect to the $W + b$ decay by factors of $|V_{ts}/V_{tb}|^2 \sim 2 \times 10^{-3}$ and $|V_{td}/V_{tb}|^2 \sim 5 \times 10^{-5}$. That $|V_{tb}|$ is very close to 1 leads to predictions for branching ratios of $t \to Wb$ of nearly 100% [1], with direct searches showing no obvious discrepancies from the SM predictions [22]. Once $t\bar{t}$ pairs are produced and decay ($t\bar{t} \to W^+bW^-\bar{b}$), the phenomenology is given by the decay of the two W bosons. Each W boson can decay hadronically to a quark-antiquark pair, or it can decay leptonically, to a charged lepton and a neutrino. Due to the difficult nature of reconstructing τ leptons, for the purposes of this analysis we classify only electrons (e) and muons (μ) as leptons, and typically consider taus as a separate category of W decay products. The 4 classes of $t\bar{t}$ events are then:

- Dilepton Events, in which both W bosons decay leptonically. $t\bar{t}$ events have a small branching ratio of $\sim 5\%$ to the dilepton channel, and cannot be fully reconstructed due to the presence of two neutrinos that are not observed in the detector. The dilepton channel has the advantage of a low background rate and only 2 quarks per event.

- All-Hadronic Events, in which both W bosons decay hadronically. This channel has the largest branching ratio ($\sim 46\%$), but without any leptons, suffers from an overwhelming QCD background. All-Hadronic events each contain a minimum of 6 quarks, and produce a very crowded and busy detector.

- Tau Events, in which at least one of the W bosons decays to $\tau + \nu$. These events, constituting $\sim 20\%$ of all $t\bar{t}$ events, are difficult to accurately reconstruct because of the subsequent decay of the τ, and generally are avoided in top mass measurements.

- Lepton+Jets Events, in which one W decays hadronically, and the other W decays leptonically. These events have a branching ratio of $\sim 30\%$, with a minimum of 4 quarks per event. Backgrounds for Lepton+Jets events are not negligible, but can be kept under control. The neutrino escapes undetected, but can be reconstructed by using the overconstrained kinematics of the $t\bar{t}$ system.

This analysis considers only events consistent with $t\bar{t}$ production and decay via the Lepton+Jets channel.

CHAPTER 3

APPARATUS

Until the turn-on of the LHC, the Tevatron is the only place on the planet capable of producting top quarks, and Fermilab's two large, multi-purpose experiments, CDF and DZero, are the only collaborations capable of studying top quark properties. This chapter discusses the accelerator chain making it possible to produce $t\bar{t}$ pairs, and also the CDF detector used by this analysis to study these events.

3.1 Accelerator overview

In order to collide protons and antiprotons at a center-of-mass energy of 1.96 TeV, the Fermilab accelerator complex needs many distinct components to produce protons (p) and antiprotons ($\bar{p}$), boost particle energies many times and focus beams. As Run II progresses, Fermilab's accelerator physicists continue to push the performance of the lab's machines. By the middle of 2008, initial instantaneous luminosities at the Tevatron collision halls are routinely pushing close to $300 \times 10^{30}\mathrm{cm}^{-2}\mathrm{sec}^{-1}$. Figure 3.1 shows a schematic of all the components necessary to obtain high-energy collisions. The accelerator systems needed to initiate $p\bar{p}$ collisions can be broken down into 5 groups: the proton source (the pre-accelerator, linac and Booster), the Main Injector, the antiproton source (the target station, the Debuncher and the Accumulator), the Recycler and the Tevatron.

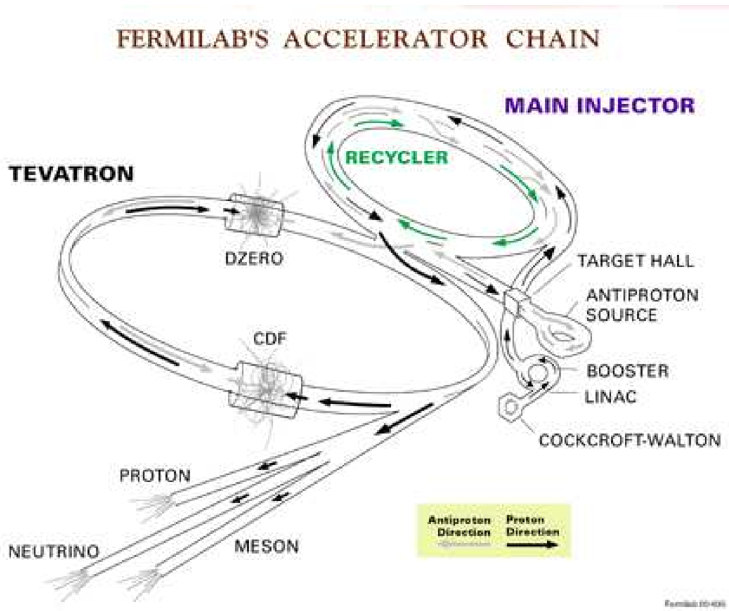

Figure 3.1: Schematic of Fermilab's accelerator complex.

3.1.1 The proton source

Creation of both p and $\bar{p}$ at Fermilab begins with a simple bottle of hydrogen gas. The pre-accelerator converts H_2 to an ionized hydrogen (H^-) gas via the addition of an extra electron to every hydrogen atom. A Cockroft-Walton dome accelerates the charged gas to energies of 750 keV with a repitition rate of 15 MHz. After exiting the Cockroft-Walton dome, the ionized gas travels toward the Linac, a 130 meter-long linear accelerator that uses radio frequency (RF) cavities to boost the energy of the ions to 400 MeV. After reaching 400 MeV, the ions travel to the Booster, where thin carbon foils strip the negatively charged hydrogen of both its electrons, giving positively charged protons. The Booster, a 75-meter radius circular accelerator, uses 18 RF cavities to accelerate the protons to 8 GeV, after which they are transferred to the Main Injector.

3.1.2 The Main Injector

The Main Injector is a synchrotron roughly 2 miles in circumference with an egg-like elliptical shape due to several straight sections in the system. The Main Injector can take p and $\bar{p}$ from several different sources and accelerate them to 120 or 150 GeV before injecting them into other accelerator subsystems. When producing and accumulating (stacking) $\bar{p}$, the Main Injector booster takes 2 batches of protons from the Booster, accelerates them to 120 GeV, and sends them toward the Antiproton source (see section 3.1.3). During what is called shot setup, protons from the Booster and antiprotons from the Recycler or Accumulator (see Sections 3.1.3, 3.1.4) are accelerated to 150 GeV before extraction to the Tevatron for studies or the initiation of collisions. The Main Injector also accelerates p from the Booster to 120 GeV for use in neutrino physics experiments.

3.1.3 The Antiproton source

Unfortunately, $\bar{p}$ are not obtained as easily as protons. Antiprotons are created, accumulated and cooled as quickly and often as possible at the Tevatron, but still remain a precious commodity that drives the ability of the Tevatron to deliver higher instantaneous and integrated luminosities. Creation of antiprotons begins with batches of roughly 5×10^{12} 120 GeV protons from the Main Injector that strike a rotating nickel target every 1.5 seconds. Before being sent to the target, the energy-time phase space of the protons is rotated to narrow the width of the bunches in time. This comes at the expense of widening their momentum distribution, but makes it easier to collect the antiprotons. The collisions with the nickel target produce sprays of particles with a variety of types, energies and angles. Occasionally, an antiproton is produced.

The showers of particles coming from the collisions enter a lithium lens, in which a very strong solenoidal magnetic field focuses the negatively charged secondary particles so they travel parallel to one another. Following the lens is a pulsed dipole magnet, which selects 8 GeV negatively charged particles with a 4% momentum acceptance. Within the angular and momentum acceptances, roughly 15 out of every million proton collisions produce a captured antiproton. At peak stacking rates, Fermilab accumulates approximately $6 \times 10^7 \bar{p}$ per second, with an average accumulation rate closer to $4 \times 10^7 \bar{p}$ per second. A comparison of these numbers with the 5×10^{12} protons striking the target every 1.5 seconds shows why $\bar{p}$ accumulation is the driving force behind Tevatron luminosities.

Antiprotons passing through the pulsed magnet selection enter the Debuncher, a triangular-shaped synchrotron with an average radius of 90 meters. The purpose of the Debuncher is to collect 8 GeV $\bar{p}$, perform a phase rotation to reduce their momentum spread, and cool them (in other words, to reduce their 6-dimensional phase space). This is performed with a stochastic cooling system, in which signals of small deviations of the beam around the ideal orbit are collected, amplified and sent across the Debuncher such that they arrive before that portion of the beam (this is possible because the signal can travel in a straight line whereas the beam must travel around the ring). Use of a single set of these small signals cannot cool the beam, but repeated application of the technique can significantly reduce its momentum spread.

Afer cooling, $\bar{p}$ from the debuncher are sent to the Accumulator, which was the original storage ring for antiprotons produced at the Tevatron. Housed in the same tunnel as the Debuncher, the Accumulator is also triangular-shaped and has an average radius of 75 meters. Additional cooling systems are available in the Accumulator, which is now typically used only as a temporary storage device for antiprotons.

3.1.4 The Recycler

When significantly large numbers of $\bar{p}$ collect in the Accumulator (usually 4×10^{11} – 1×10^{12}), the antiprotons are sent to the Recycler, a storage ring built along the ceiling of the Main Injector tunnel. The Recycler was originally built to recycle (reuse) antiprotons remaining in the ring after Tevatron collisions conclude, but is now used as a storage and cooling device instead.

The stacking rate in the $\bar{p}$ source decreases as antiprotons collect in the accumulator and occupy more and more phase space in the ring. Transferring $\bar{p}$ out of the accumulator and freeing up phase space thus increases the overall stacking rate, even with the loss of beam during transport. Additionally, stochastic cooling looses its effectiveness with higher $\bar{p}$ intensities, and a new way to cool the antiprotons becomes necessary. The Recycler uses a technique called electron cooling, which applies the thermodynamic principle that energy flows from energetic objects to cooler objects. Electrons are produced in a 5 MeV Pelletron and temporarily injected into the $\bar{p}$ beam with the same average velocity as the antiprotons but a much smaller spread in spatial coordinates and velocity (in other words, a smaller emittance). Since the electrons have less random motion, repeated $\bar{p}$-e collisions transfer momentrum from the antiprotons to the electrons, thus reducing the antiproton emittance.

3.1.5 The Tevatron

The lagest of the Fermilab accelerators and currently the highest-energy accelerator in the world, the Tevatron is a roughly circular synchrotron approximately 4 miles in circumference responsible for colliding protons and antiprotons at a center-of-mass energy of 1.96 TeV over durations (stores) typically lasting $\sim$ 24 hours. The Tevatron is the only accelerator at Fermilab using superconducting magnets, and needs to be

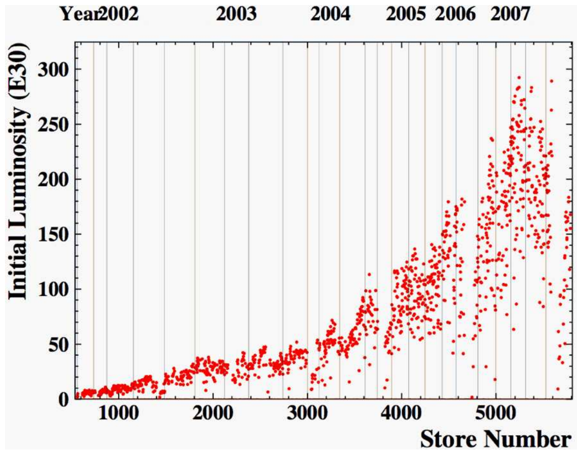

Figure 3.2: Initial instantaneous luminosity at the Tevatron as a function of store number (bottom axis) and year (top axis).

kept at extremely cold temperatures for the magnets to remain in their superconducting state. For each store, the Tevatron takes 150 GeV protons and antiprotons from the Main Injector, accelerates them to 980 GeV using RF cavities, and collides the beams, which travel in the same tunnels but are kept separated due their opposite charge by electrostatic separators.[1] The beams collide at two locations along the Tevatron ring (B0, where CDF is located, and D0, where DZero is located). Each beam consists of 36 separate bunches, with a crossing rate of 396 ns. At the beginning of a store, each proton bunch typically contains 3×10^{11} particles, and each antiproton bunch contains $3 - 10 \times 10^{10}$ particles, depending on the available number of antiprotons. Several years after the startup of Run II, initial instantaneous luminosities of $250 \times 10^{30}\mathrm{cm}^{-2}\mathrm{sec}^{-1}$ are commonplace, as seen in Figure 3.2.

1. The separators produce large electric fields perpendicular to the direction of the beams, putting the protons and antiprotons on different helical trajectories throughout the Tevatron.

3.2 CDF

The Collider Detector at Fermilab (CDF) is a forward-backward, azimuthally symmetric detector designed to study a wide range of physics at the Tevatron. Topics studied at CDF include botton and charm spectroscopy, precision electroweak measurements, Higgs searches, exotic searches and, of course, top quark physics. The detector has an onion-like structure typical of multi-purpose experiments in high-energy physics, with precision silicon tracking devices located near the collision point and additional detector subsystems layered on top of one another with increasing radial distance. CDF uses a right-handed coordinate system, with the z-axis pointing in a tangent to the Tevatron along the direction of the proton beam, the x-axis pointing outward in the plane of the Tevatron ring, and the y-axis pointing straight up. It is usually more convenient to use cylindrical coordinates $r = x^2 + y^2$, $\phi = \tan^{-1}(\frac{y}{x})$ and $\theta = \cos^{-1}(\frac{z}{r})$. The quantities of interest are often those defined in the plane transverse to the beam (ie perpendicular to the z axis), such as transverse energy ($E_T \equiv E \sin\theta$) and transverse momentum ($p_T = p \sin\theta$). Since the Tevatron collides constituents of the proton and antiproton (and not the p and $\bar{p}$ themselves), the center-of-mass energy and momentum along the z axis after collisions are unknown. It is commonplace to replace the polar angle (θ) with the rapidity ($y = \frac{1}{2}\ln(\frac{E+p_z}{E-p_z})$), a quantity that is better behavied under a Lorentz boost in the z direction - y would then shift by an amount independent of E or p_z. In the relativistic limit where mass is much smaller than a particle's momentum (usually a very good approximation in collisions involving top quarks), this reduces to the the pseudorapidity $\eta = -\ln(\tan\frac{\theta}{2})$.[2] Since

2. To give the reader some intuitive sense of the size of a unit of pseudorapapidity: $\eta = 0$ for objects moving only in the transverse plane ($\theta = 90^o$), $\eta = \pm 1$ at roughly 35^o from the z-axis. At 15^o from the z-axis, $\eta = \pm 2$. The pseudorapidity diverges for objects traveling along the beamline ($\theta = 0$ corresponds to infinite η).

the interaction region at B0 is 60 cm in length and collisions do not always occur at $z = 0$, the pseudorapidity can be defined in two ways: η refers to the pseudorapidity of an event, using the measured collision vertex as the origin of the coordinate system, and η_{det} refers to the pseudorapidity in the detector frame of reference, where the origin of the coordinate system is in the center of the detector. It is also common to reference a region in a cone of ΔR around an object. This refers to the nearby region in $\eta - \phi$ space: $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. For example, objects in a cone of $\Delta R = 0.4$ around an electron are any objects (such as tracks, jets, leptons and photons) with $\sqrt{(\eta_{electron} - \eta_{object})^2 + (\phi_{electron} - \phi_{object})^2} < 0.4$.

A cut-away view of the CDF detector is shown in Figure 3.3. Working outward from the interaction point, the detector can be divided into tracking systems (silicon and drift chamber systems to precisely measure the trajectories of charged particles), calorimeters (to measure the energy and locations of particles), and muon systems (to identify muons).

3.2.1 The tracking system

Tracking in CDF is done as close to the interaction region as possible, both to minimize scattering of charged particles off of additional material, and also to make precise position measurements of decays near the interaction region. This is particularly important for identifying the decays of b hadrons, which are metastable and on average travel several millimeters in the detector before decaying. The tracking system can be divided into two parts: the silicon systems and the drift chamber. Both are located inside a 1.4 Tesla uniform magnetic field pointing in the z direction, created by a superconducting solenoid 5 m in length at a radius of 1.5 m. Combining knowledge of the magnetic field with the curvature of tracks enables precise measurements of

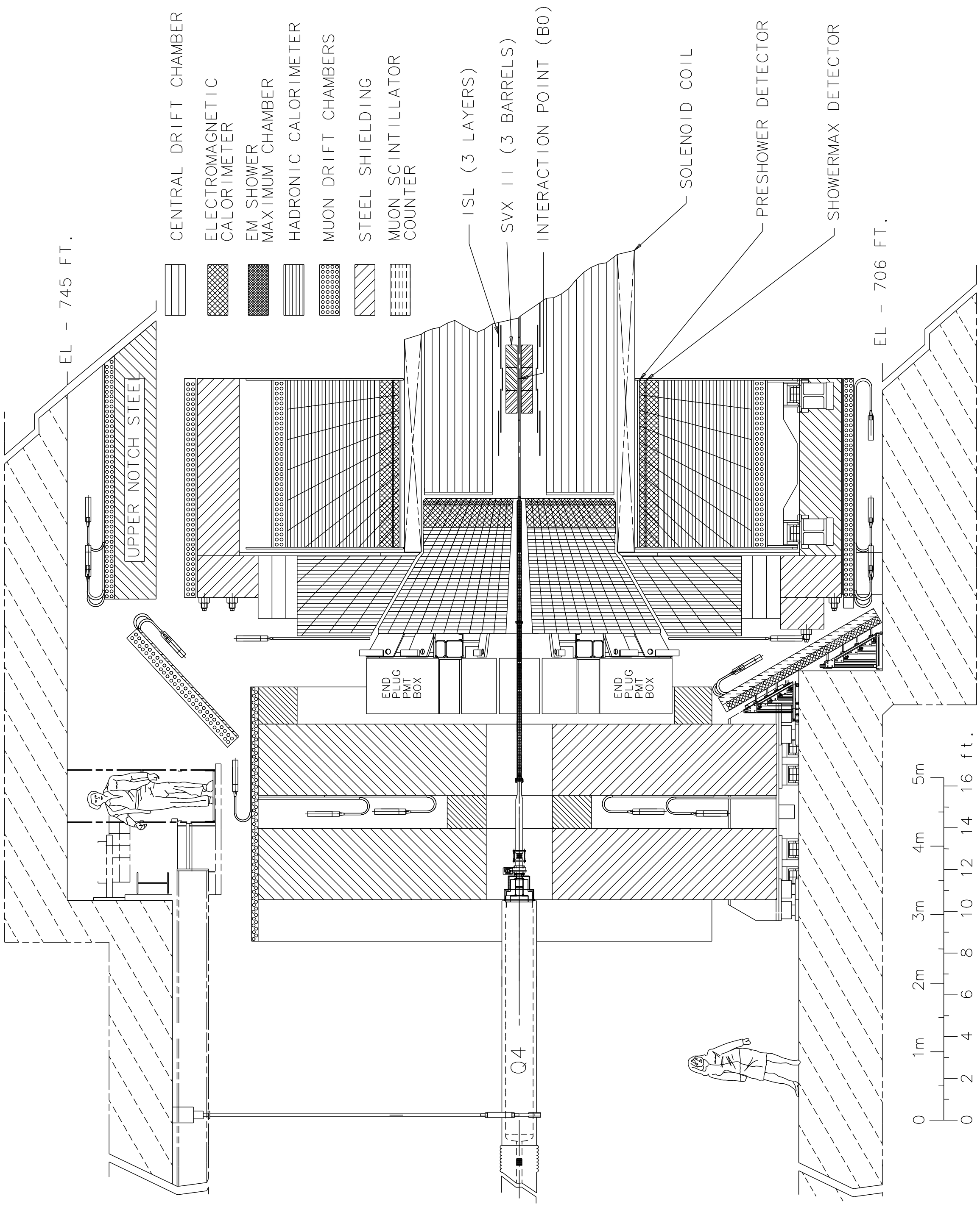

Figure 3.3: One slice of the CDF detector. The other half of the detector is identical.

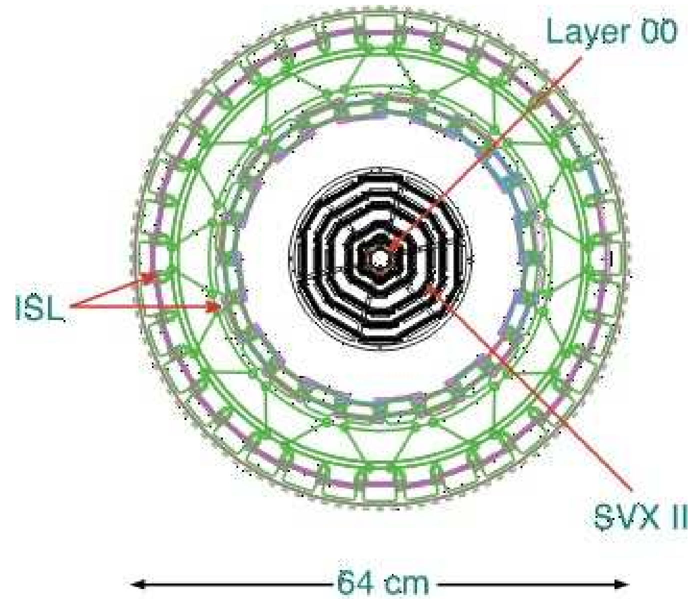

Figure 3.4: Schematic view of the 3 silicon subsystems in the $r - \phi$ plane.

charged particle momenta and charge determination.

The silicon system consists of 3 subcomponents, shown in Figure 3.4. From smallest to largest radius, the first, Layer 00 (L00), is a single layer of radiation-hard silicon mounted directly on the beampipe ($r = 1.6$ cm), providing a measurement of charged particle trajectories as close to the interaction region as allowed. L00 provides only axial information, with a readout pitch of 50 μm. The second system, the Silicon VerteX detector (SVX-II), contains 5 layers of double-sided silicon located at radii from 2.5 cm to 10.6 cm. Each wafer of silicon has an axial-only side, making measurements only in the $r - \phi$ plane. On the other side, the 1st, 3rd and 5th layers make stereo measurements completely in the $z - \phi$ plane, and the other 2 layers contain small-stereo angle strips pitched 1.2° from the strips making axial measurements. The SVX-II is 90 cm long, and is divided into 12 identical wedges in ϕ and 3 mechanical barrels in z. The strip pitch is 55-65 μm for axial strips, 60-75 μm for small-stereo angle strips, and 125-145 μm for 90° stereo angle strips. The advantage to having

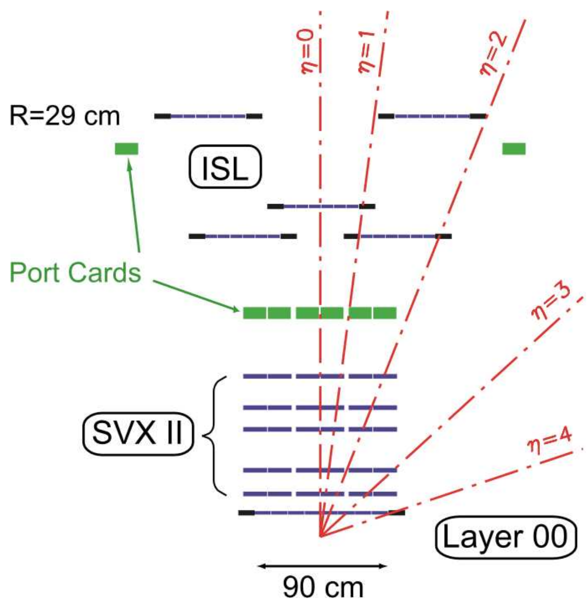

Figure 3.5: Schematic view of the 3 silicon subsystems in the $r - z$ plane.

stereo information is the ability to fully reconstruct 3-dimensional track helices. The two layers of the Intermediate Silicon Layers (ISL), located at radii of 20-29 cm, make up the third silicon subsystem. One side of the ISL contains strips with a pitch of 55 μm making axial measurements, while the strips on the opposite side make small-angle stereo measurements with a pitch of 73 μm. Figure 3.5 shows the silicon in the $r - z$ plane. Coverage is good out to $|\eta_{det}| = 2$, and doesn't completely fall off until $|\eta_{det}| \approx 3 - 4$. The resolution on the impact parameter (distance of closest approach to the origin in the $r - \phi$ plane) of tracks is roughly 40 μm, including a $\approx 30\mu m$ contribution from the width of the beam itself. The resolution on the z_o of tracks is 70 μm. In total, the CDF silicon system contains 722,432 channels.

The Central Outer Tracker (COT) makes up the second half of CDF's tracking system. An open-cell drift chamber located at much larger radii than the silicon, the COT is particularly useful for measuring particle curvature (and thus momenta). The COT is 310 cm in length, and covers radii from 43 to 132 cm and $|\eta_{det}| < 1.0$. There are 96 position measurements of a particle's trajectory, grouped into 8 "superlayers" that alternate between axial and 2^o stereo measurements. The gas used in the COT is a 1:1 mixture of argone and ethane, giving a relatively quick drift velocity and a maximum drift time of 100-200 ns, which prevents pileup of events from bunch crossings occuring every 396 ns. The transverse momentum resolution is $\sigma_{p_T}/p_T \approx 0.15\% \times p_T$ [GeV/c].

3.2.2 Calorimetry

Sitting outside the solenoid, the calorimeter system measures the energy and position of particles passing through and interacting with dense material. The calorimeters sample a fraction of the total energy deposited by both charged and neutral particles, making it possible to deduce original particle energies. Electromagnetic calorimeters sample the deposition of energy via electromagnetic showers caused by pair productions of electrons and bremsstralung of photons. Hadronic calorimeters sample the deposition of hadronic energy via interactions with nucleii in the detector. The CDF calorimeters are segmented into unit cells called towers that are read out independently, with each tower pointing toward the origin of the coordinate system. The calorimeters subtend the full 2π in ϕ and are roughly divided into two regions: the central region ($|\eta_{det}| \lesssim 1.1$) and the plug region ($1.1 \lesssim |\eta_{det}| \lesssim 3.6$).

The calorimetry in the central region consists of the central electromagnetic (CEM), hadronic (CHA) and wall hadronic (WHA) calorimeters, with each tower subtending 15^o in ϕ and 0.1 in η. Consisting of alternating layers of lead and scintillator amounting to 18 radiation lengths of material in depth, the CEM is the first calorimeter (smallest r) in the central region. Embedded in the CEM is the shower maximum detector (CES), which provides position measurements of electromagnetic showers at a depth of 5.9 radiation lengths, and is used in electron identification. Wire chambers with a 1.45 cm cell size are used to provide the axial position measurements, and cathode strips 1.67-2.01 cm wide provide stereo information. Behind the CEM lies the CHA, which consists of 4.7 interaction lengths of alternating steel and scintillator and provides energy measurements of hadronic particles. The CHA covers the region $|\eta_{det}| < 0.9$, and is augmented by the similarly designed WHA, which covers the region $0.7 < |\eta_{det}| < 1.2$. Steel, which is significantly cheaper than lead, is used for the hadronic calorimeters, which need to be much larger to measure the energy of hadrons, as hadrons have a much larger mean free path in material than electrons.

Inclusion of the calorimetry in the plug region increases the acceptance for $t\bar{t}$ events, allows for better background vetoing, and improves on the measurement of unbalanced momenta in an event, indicative of an escaping neutrino from W decay. The plug electromagnetic calorimeter (PEM) consists of 23.2 radiation lengths of alternating lead absorder and scintillating tile, and is read out with wavelength shifting fibers. A shower maximum detector (PES) imbedded in the PEM measures the position of electromagnetic showers. The plug hadronic calorimeter (PHA) has 6.8 interaction lengths of alternating iron and scintillating tile to measure the energy of hadronic particles.

3.2.3 Muon detectors

With a mass over 200 times larger than an electron, a muon traversing the CDF detector does not emit much radiation. The cross section for a muon to undergo electromagnetic or nuclear interactions is small, so muons typical travel through the CDF detector intact and leave behind only ionization energy, making the calorimeter systems described in section 3.2.2 inadequate to detect muons. To find such minimum ionizing particles, the muon detectors are located behind both the calorimeters and large amounts of steel shielding that remove remnant hadronic particles. Muons pass through the shielding and leave behind hits (stubs) in the muon detectors, which are single-wire drift cells four layers deep.

Directly behind the CHA and the steel shielding is the central muon detector (CMU), which covers $|\eta_{det}| < 0.6$. The central muon upgrade (CMP) adds coverage in the central region and provides background rejection via an additional 60 cm of steel shielding separating it from the CMU. The central muon extension (CMX) covers the more forward region of $0.6 < |\eta_{det}| < 1.0$.

3.2.4 Luminosity counters

Any measurement of a $t\bar{t}$ cross section needs an integrated luminosity for normalization, and as we shall see, some of the estimates for backgrounds to $t\bar{t}$ production are normalized as well to integrated luminosity. The Cherenkov Luminosity Counter (CLC) measures the instantaneous (and thus integrated) luminosity of $p\bar{p}$ collisions at B0. The CLC consists of gaseous Cherenkov counters located at each end of the detector, covering $3.6 < |\eta_{det}| < 4.6$. Photomultiplier tubes detect the Cherenkov light emitted by charged particles passing through the counters. As the luminosity rises, so does the rate of inelastic collisions. This produces more charged particles traversing the forward region and more light measured by the CLC. Given the knowledge of the inelastic cross section for $p\bar{p}$ collisions and the CLC acceptance, the instantaneous and integrated luminosities can be measured with a relative uncertainty of 6%.

3.2.5 Trigger system

The 396 ns $p\bar{p}$ bunch crossing rate at the Tevatron implies a 2.5 MHz collision rate inside the detector.[3] In most of these bunch crossings, nothing takes place of much interest to a modern-day high-energy physicist, and recording what the detector sees at these rates would take a prohibitively large amount of disk space. CDF employs a three-tiered trigger system to select potentially interesting events for a wide variety of physics analyses, ultimately reducing the 2.5 MHz rate down to roughly 100 Hz, a reduction of 2×10^4 in data size. [4] Roughly 10 billion inelastic collisions occur for every $t\bar{t}$ event produced at the Tevatron.

At the first trigger level, custom-designed hardware makes decisions about events based on crude physics quantities. The level-1 (L1) trigger looks for simple signs of interesting physics objects. Examples include energetic calorimeter towers, missing energy in the transverse plane of the detector, stubs in the muon detectors and stiff tracks found in the drift chamber. This rough reconstruction of physics objects takes place in a pipeline capable of buffering 42 beam crossings, with a decision made in 5.5 μs so that the system is deadtimeless. The accept rate for the L1 trigger varies

3. Every 12 of the 36 bunches of p and $\bar{p}$ are grouped into what are called bunch trains, which are separated by (relatively) large areas in which there should be no circulating beams. These spots, called the abort gaps, give the Tevatron enough time to ramp up septum magnets and safely aim the beam toward beam dumps in case of an accelerator failure or at the end a store. Thus, although the instantaneous crossing rate can be as high as 2.5 MHz, the average beam crossing rate is more like 1.7 MHz. At an instantatenous lumonisity of $100 \times 10^{30}\mathrm{cm}^{-2}\mathrm{sec}^{-1}$, which is the typical order of magnitude for Tevatron luminosity in Run II, roughly 3 inelastic collisions occur per bunch crossing. The remaining $\sim 10^{10}$ particles do not interact with one another and continue on, remaining in the Tevatron ring.

4. To give the reader an idea for why a trigger system is needed - raw event data sizes on the order of 100 kB at a data-logging rate of 100 Hz produce on the order of 1 terabyte of data a day. Without any reduction in the trigger rate, we would need a way to write tens of petabytes of data to tape every single day. One petabyte $\equiv$ 1000 terabytes $\equiv$ 1 million gigabytes.

depending on luminosity, beam conditions and the triggers being used, but is on the order of 25 KHz, giving a rejection factor of roughly 100.

Events passing the L1 trigger are sent to the level-2 (L2) trigger, which also uses custom-designed hardware. L2 performs a rudimentary event reconstruction, and looks for interesting, more high-level objects such as energetic clusters of towers in the calorimeter or tracks in the silicon that have displaced vertices consistent with secondary decays inside the detector. This event information is passed to a single high-speed commodity PC, which makes the L2 trigger decision. The accept rate for L2 also varies, but is usually on the order of 500 Hz, giving an additional rejection factor of roughly 50.

Events passing the L2 trigger are sent to the level-3 trigger, which consists of a massive CPU farm performing offline-quality event reconstruction. In order to perform full event reconstruction, the data from all detector subsystems must be read out and assembled together. The L3 trigger confirms the L2 trigger decision, rejecting another factor of $\sim$ 5x and reducing the data-logging rate to tape to the more manageable 75-100 Hz.

CHAPTER 4

EVENT SELECTION

As discussed in Section 2.5, this thesis applies to the Lepton+Jets decay channel of $t\bar{t}$ events, with an event signature of four quarks (two of them b quarks), a single, energetic electron or muon, and an escaping neutrino. A typical Feynman diagram for a Lepton+Jets event is shown in Figure 4.1. These events are produced close to threshold at the Tevatron. The average p_T of the $t\bar{t}$ system is ≈ 20 GeV/c and the average p_T of a single top quark is ≈ 100 GeV/c. [1] Due to the threshold effects, $t\bar{t}$ events are rather spherical in nature, and most of the partons are found in the central region of the detector. Unfortunately, there is no easy way to pair what we see in the detector with the partons from the hard-scatter process. There are two top quarks in each event, so simply taking the invariant mass of all objects in the detector does not give a good measurement of the individual top quark mass; any kinematic reconstruction needs some way to identify and separate the different types of hadronic objects (jets) in the event.

4.1 Trigger

There are several striking things about Lepton+Jets $t\bar{t}$ events - a minimum of 4 jets, momentum imbalance in the transverse plane due to the escaping neutrino and 2 b quarks. Yet from the point of view of the trigger system, the easiest thing

1. Top quarks are not produced ultrarelativistically at the Tevatron. The average $\beta \approx 0.5$.

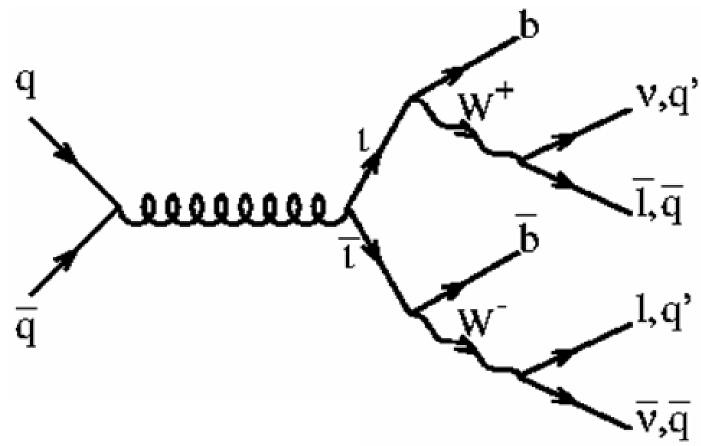

Figure 4.1: A typical diagram for $t\bar{t}$ production (through quark annihilation) and decay (through the Lepton+Jets channel) at the Tevatron. The charged lepton refers to either an electron or muon.

to identify is the single high-p_T, isolated electron or muon. The energetic charged lepton found in these events distinguishes them from the otherwise overwhelming QCD background. For central electron events, the L1 trigger requires an EM cluster with $E_T > 8$ GeV, a matching track in the COT having $p_T > 8$ GeV/c, and a ratio of hadronic-to-electromagnetic energy less than 1:8. At L2, the requirement is tightened to $E_T > 16$ GeV, and L3 makes basic electron identification cuts and further tightens the energy requuirement to $E_T > 18$ GeV. These events are referred to as CEM events, with a trigger path $\sim 96\%$ efficient. The PEM is used only to veto dilepton events; events with the high-p_T trigger lepton in the plug are not used in this analysis.

For muons in the central region, L1 requires stubs in both the CMU and CMP detectors consistent with muons with $p_T > 6$ GeV/c, and a matching track in the drift chamber with $p_T > 4$ GeV/c. At L2, the p_T requirement is increased to 15 GeV/c, and at L3 to 18 GeV/c. These events are referred to as CMUP events. For muons in the plug region, the trigger is similar. These events are referred to as CMX events. The efficiency for muon triggers is roughly 93%.

4.2 Offline Reconstruction

Events passing the trigger requirement are logged temporarily in B0 (the building housing the CDF detector) before being sent to the Feynman Computing Center at Fermilab, where they are stored on tapes. The events are grouped into datasets corresponding to different data-taking periods and the trigger that the event fired. Calibrations are applied for shifting beamlines and calorimeter response, after which high-level event reconstruction is undertaken. High-level event quantities are stored in large relational ntuples that contain all the necessary information for kinematic reconstruction of $t\bar{t}$ events.

Before event reconstruction, a good run list is applied to every event to ensure that all detector components were active and working correctly during the period when the event was recorded. The total integrated luminosity of good data corresponding to this analysis is 1.9 fb^{-1}.

4.3 Lepton Selection

Electron events are required offline to have an EM cluster with $E_T > 20$ GeV. The electron is required to be isolated from other information in the event: 90% or more of the energy in a cone of $\Delta R = 0.4$ around the electron is required to be associated with the electron. Additional track-quality cuts are also made to ensure a clean electron sample.

Muon events are required to have a track with $p_T > 20$ GeV$/c$ and a matching stub in the muon chambers. Energy deposition in the calorimeters is required to be consistent with a minimium ionizing particle. Similar to the isolation cut for electrons, 90% or more of the energy in a cone of $\Delta R = 0.4$ around the muon is required to be associated with the lepton.

Dilepton channel $t\bar{t}$ is vetoed by rejecting events with more than one lepton. Additional categories of leptons such as PEM electrons and muons that strike only the CMU or CMP chambers are included in this dilepton veto.

4.4 Jets

The property of quark confinement in QCD [23, 24] ensures that bare quarks are not directly observable. Quarks and gluons instead manifest themselves in the detector as jets of particles in the direction of the original parton. After production, quarks immediately undergo parton showering, in which gluon radiation is emitted, leading to $q\bar{q}$ pairs. Additional $q\bar{q}$ pairs are pulled out of the vacuum, and hadronic bound states are formed. It is energetically favorable for this process to continue until a collimated, macroscopic region of the detector, called a jet, is filled with lower energy particles - mostly protons, neutrons, pions and kaons. Some more massive hadrons may decay, however, into charged leptons and neutrinos.

Though the jet as a whole is in roughly the same direction as the original parton, some of the low-energy particles may exist at very wide angles with respect to the original particle direction, either due to the nature of fragmentation kinematics or because low-momentum charged particles are swept away by the magnetic field in the detector. Additionally, some other particles the $t\bar{t}$ decay or from beam remants may enter into the region of the jet, making it difficult to make a precise measurement of the original parton energy. It is also possible for objects within another jet or the Tevatron beam itself to exchange color information with the jet in question and alter reconstructed distributions. Nevertheless, the sum of the total energy within a cone around the direction of the original particle should be strongly correlated with its energy.

Jet reconstruction, then, begins with the defintion of a jet in the detector. If the size of the jet is too small, information about the original parton energy is lost. If the jet is too big, it becomes possible to wash out kinematic or directional information on the original parton. Jets from different partons are also more likely to merge together when a larger jet cone is used. Jets for this analysis are reconstructed with a cone-based clustering algorithm called JETCLU [25], using a cone size of $\Delta R = 0.4$. A detailed explanation of CDF jet calibration and its associated systematics and uncertainties can be found in [26].

Jets are first identified offline by looking for energetic seed towers with $E_T >$ 1 GeV. Towers within the initial cone of $\Delta R = 0.4$ are then added to the jet, and a new jet center is calculated by an E_T-weighted average of every tower in the jet. This process is repeated until the jet no longer changes. The sum of the energies in the final jet gives the raw jet energy in the detector. Overlapping jets are merged if the overlap is more than 50%, and if a tower belongs to two jets it is assigned to the closest jet in ΔR space.

4.4.1 Relative corrections

From the raw jet energy many stages of corrections are applied that attempt to remove effects that mask the initial parton energy. In the first stage, the relative corrections normalize the detector response as a function of η_{det} so that all jets have the E_T response of the well understood central region of the detector, $0.2 < |\eta_{det}| < 0.6$. The response varies across η due to cracks in the detector and different amounts of material in front of the calorimeters. The relative corrections are derived using large samples of dijet events in which the E_T of one jet in the central region is balanced against a probe jet in another region of the detector. Modulo the small intrinstic

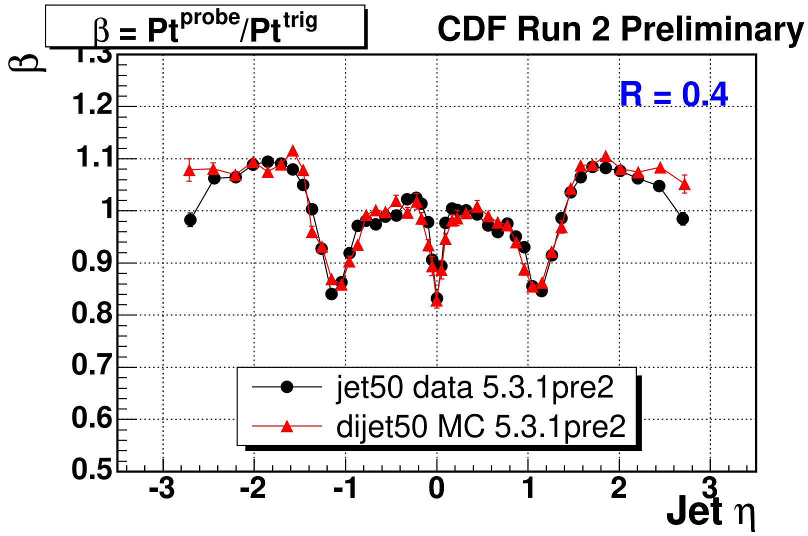

Figure 4.2: Dijet balance ratio for both data and Monte Carlo simulation as a function of η_{det}. β is the average p_T of the probe jet (the jet that needs correcting) vs the trigger jet (the central jet). The multiplicative correction to the trigger jet is then $1/\beta$. Red points are PYTHIA Monte Carlo, black points are data. The Monte Carlo is used to generate the correction.

energy of the initial partons in the transverse plane, vetoing events with additional jet activity ensures that the energy of the two jets should balance. The relative corrections are typically 5-20 % of the raw jet E_T, and depend on the E_T and η_{det} of the jet (they can both increase and decrease the jet energy). Figure 4.2 shows the average p_T of the probe jet compared to the trigger jet as a function of η_{det}, for both MC and data. Varying the selection requirements and the fitting procedure provides systematic uncertainties, shown in Figure 4.3.

4.4.2 Multiple interactions

The next level of jet corrections accounts for the average additional energy in a jet cone due to additional $p\bar{p}$ inelastic interactions occuring during the same bunch crossing. The multiple interaction (MI) correction is parameterized as a linear dependence on

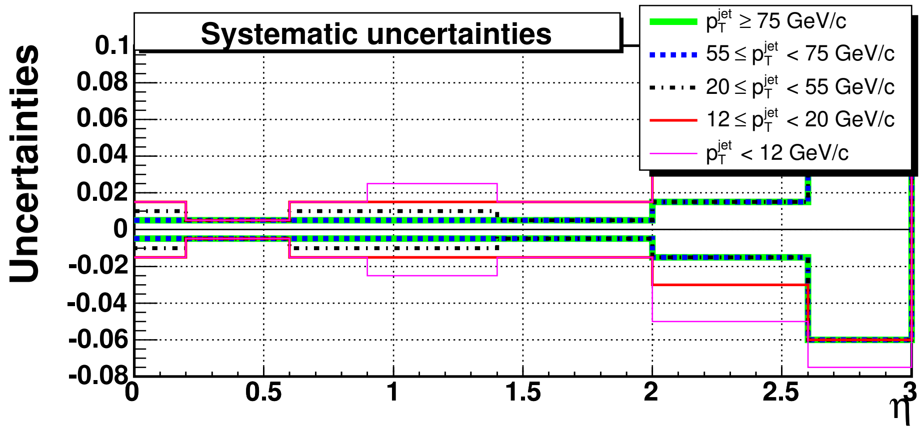

Figure 4.3: Systematic uncertainties on the relative corrections as a function of η_{det} for jets with different p_T values. The uncertainty shown is an absolute uncertainty on the multiplicative correction needed.

the number of additional vertices found in the event - every extra interaction should produce an additional vertex. Using minimum bias data, each additional interaction is found to add 356 MeV of energy, on average, to a jet of cone size $\Delta R = 0.4$. This energy, $((n_{vertex} - 1) \cdot 356$ MeV), is small compared to typical energies in $t\bar{t}$ events, and is subtracted from the energy of each jet. The correction is shown in Figure 4.4, and the systematic uncertainty, given both by the different corrections estimated in non-minbias data and by their luminosity dependence, is shown in Figure 4.5.

4.4.3 Absolute corrections

After accounting for energy due to extra $p\bar{p}$ interactions, the absolute corrections are applied. These corrections account for the non-linearity of the calorimeters and any overall detector response. After the absolute corrections, jet energies should be independent of the CDF detector, and correspond to the total energy of all the hadrons inside the jet cone (these are also called "particle jets"). The absolute corrections are derived from simulations, test-beam data and single-track energy measurements

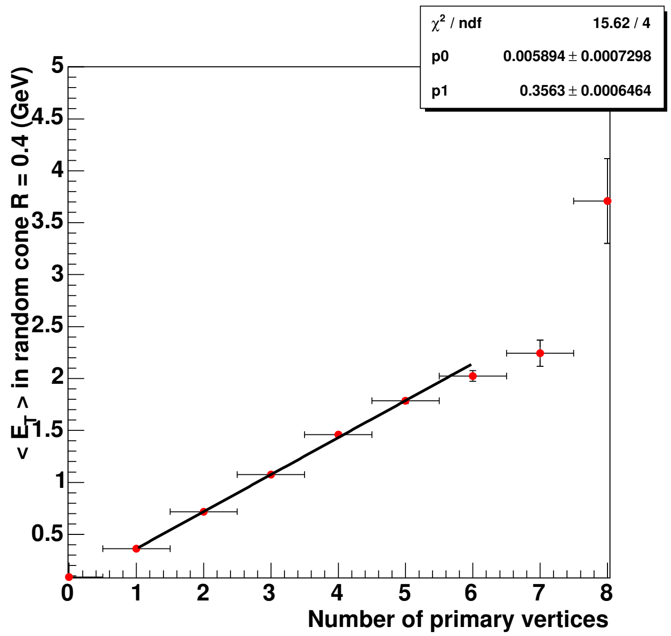

Figure 4.4: Additional average energy in each jet as a function of extra primary vertices found in the event.

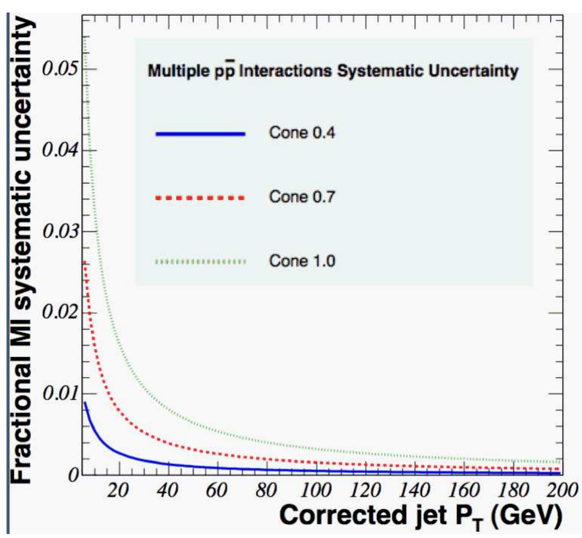

Figure 4.5: Systematic uncertainty on the multiple interaction correction. This analysis uses jets of cone size $\Delta R = 0.4$, shown in the solid blue (the line with the smallest fractional uncertainty).

in the calorimeter, and depend both on the multiplicity and p_T spectrum of particles in jets, as well as on the calorimeter response to these individual particles. The corrections return the most probable value of the particle jet energy as a function of the jet p_T, with corrections as high as over 40% for low-energy jets and as low as 10% for high-energy jets, as shown in Figure 4.6. The total uncertainty on the absolute corrections consists of many pieces, and is dominated by fragmentation uncertainties and the uncertain E/p response in the calorimeter. The total fractional uncertainty, shown in Figure 4.7, depends on the p_T of the jet, and is typically 2% of the jet p_T.

4.4.4 Jet categories

Jets with $|\eta_{det}| < 2.0$ and $E_T > 20$ GeV after applying the relative, MI and absolute corrections are referred to as "tight jets." Jets not passing the tight cuts but having $E_T > 12$ GeV and $|\eta_{det}| < 2.4$ are referred to as "loose jets." As will be described in Section 4.7, the jet requirements depend on the number of b-tags found in the event.

4.4.5 Additional jet corrections

Applying absolute corrections gives the most probable value of the energy within the jet cone, but does not give back the original parton energy. Two effects make the numbers different. Energy from spectator partons in the hard collision process that breaks up the proton and anti-proton to form the $t\bar{t}$ system must be accounted for. This correction is called the underlying energy (UE) correction. There is also a correction that accounts for out-of-cone (OOC) effects, in which some of the original parton energy lies outside the cone of $\Delta R = 0.4$, either because of fragmentation effects or because low-energy particles are swept out of the cone by the magnetic field. These corrections are derived from balancing dijet events, and can reach almost

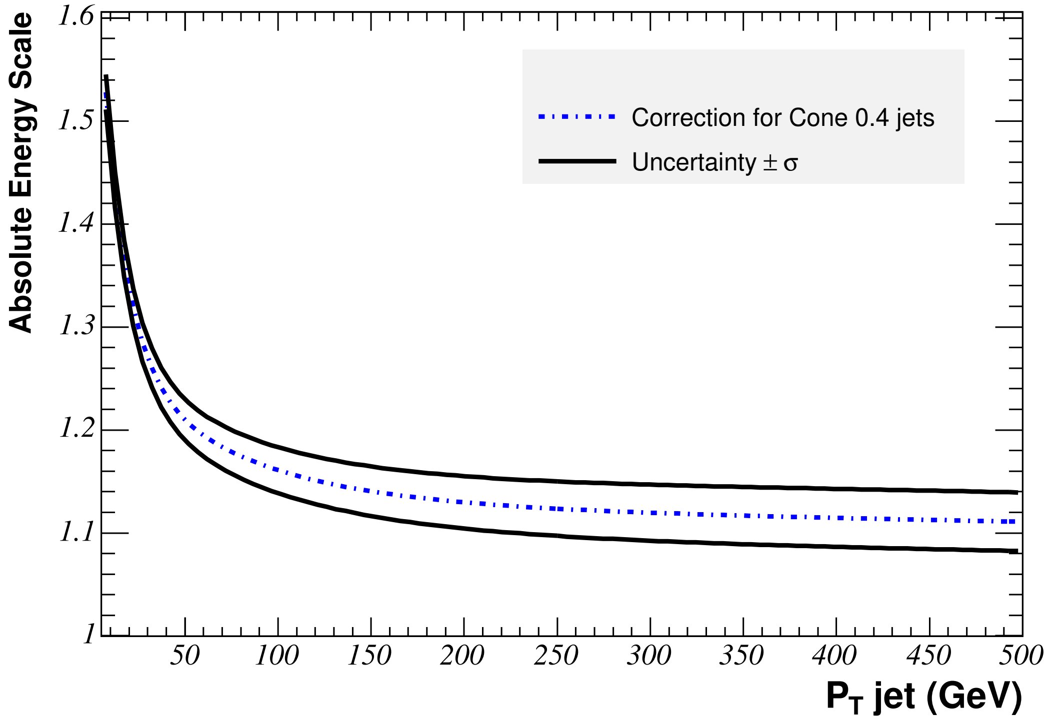

Figure 4.6: Absolute energy correction as a function of jet p_T. The correction is a multiplicative factor on the jet energy that returns the most probable energy of all hadrons inside the jet cone.

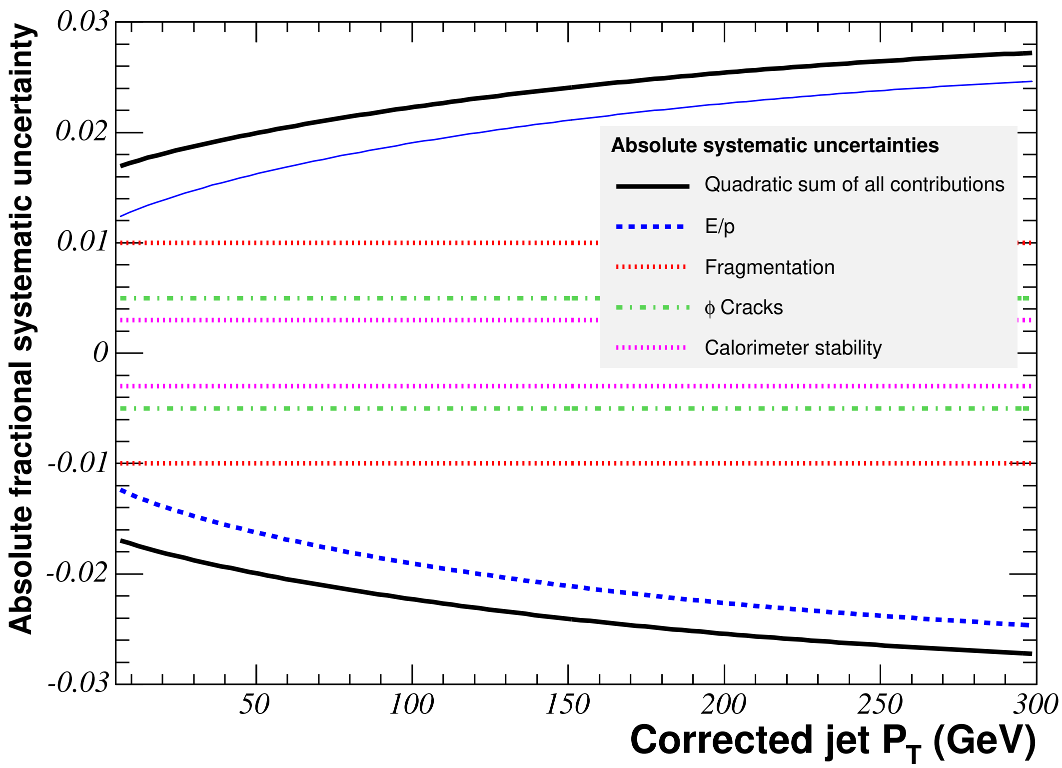

Figure 4.7: Fractional systematic uncertainty on the absolute jet energy correction, as a function of jet p_T.

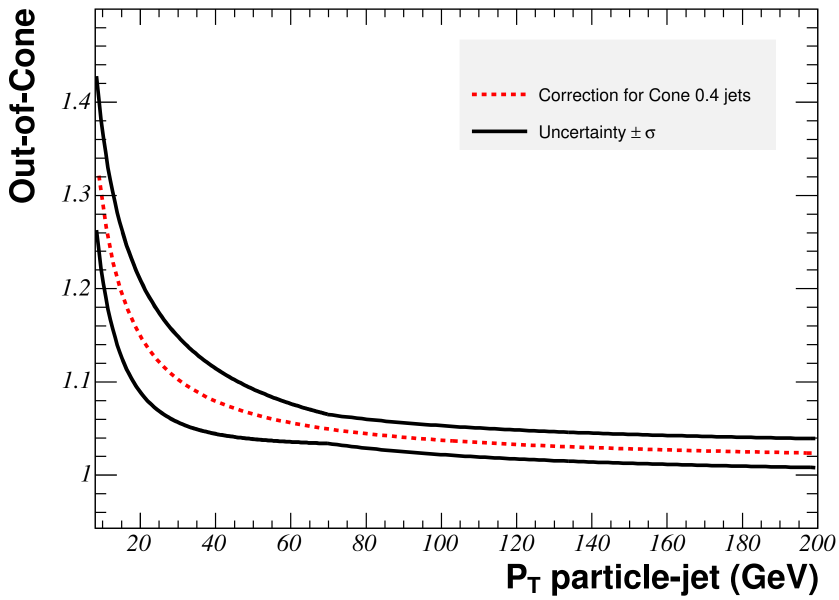

Figure 4.8: Out-of-cone energy correction and $\pm 1\sigma$ systematic uncertainty as a function of jet p_T.

20% for low-energy jets. As will be described in Section 5.1.1, these last corrections are not used, though the systematic effects arising from the MC mismodeling the data still remain. The OOC correction and systematic uncertainty dominate over the UE effects. Out-of-cone effects provide the dominant systematic uncertainty for low p_T jets, for which many of the fragmentation hadrons are swept outside the jet cone. The correction and its systematic uncertainty are shown in Figure 4.8. The correction is derived from MC, and the systematic uncertainty is found from MC-data differences.

4.4.6 Overall jet energy scale uncertainty

The combined fractional uncertainty on the jet energy calibration is shown in Figure 4.9. The overall jet energy calibration is typically referred to as the jet energy scale

(JES). The unit of JES at CDF[2] is defined as σ_C, and shifts in the JES from the expected value are referred to as Δ_{JES}. A JES calibration of 0 corresponds to the best estimate of a jet's energy after all CDF calibrations are applied. As discussed above, the calibrations have many sources of uncertainty. If every calibration is shifted up by one σ uncertainty (up in the sense that the calibration returns a larger jet energy), we say that the energy was obtained with $\Delta_{\text{JES}} = +1\sigma_C$. Similarly, if the calibrations are shifted down by one unit of uncertainty, we say that the jet was calibrated with $\Delta_{\text{JES}} = -1\sigma_C$. The expected JES at CDF is thus $0 \pm 1\sigma_C$. Note that the calibration shift for $\Delta_{\text{JES}} = \pm 1\sigma_C$ is different for a low-energy jet and a high-energy jet. The calibration might shift a jet energy up by 5% for a low-energy jet, but only 3% for a high-energy jet. This is because the uncertainty on the calibrations is different for the two jets. Similarly for jets in different regions of the detector.

4.5 Missing E_T

The neutrino from the leptonic W decay escapes undetected, leading to an imbalance of momentum in the transverse plane of the detector. Requiring large missing energy in the transverse plane ($\not{E}_T$)[3] significantly reduces the background from QCD events, in which there is no W decaying to a neutrino. The $\not{E}_T$ in each event is defined with respect to the event's interaction point along the z axis; the interaction region at B0

2. The JES has no physical meaning, and no relationship to quantities outside of the CDF detector.

3. It is conventional in hadron collider experiments to speak of an object's energy as if it were a vector like the momentum. This is to designate the source of the measurement - the magnetic spectrometer for momentum and the calorimeter for energy. Since the detected particles are traveling close to the speed of light, the magnitudes of energy and momentum are the same. Thus, a particle's E_T vector has magnitude E $\sin\theta$ and is directed along its transverse momentum vector. If the sum of the transverse momenta of all particles is zero, then the sum of the E_T vectors will also be zero and there will be no "missing E_T".

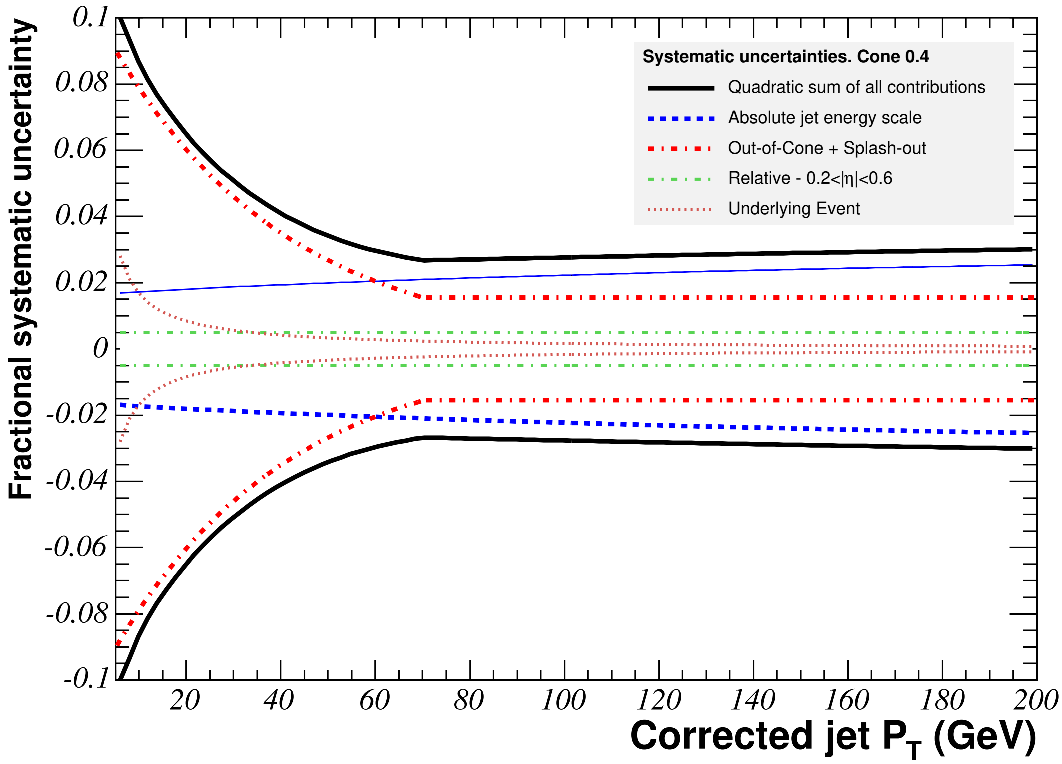

Figure 4.9: Fractional uncertainty on jet energy scale calibration as a function of corrected jet p_T.

is 60 cm in length. Good-quality tracks from the drift chamber are extrapolated to the interaction region. Vertices are defined where tracks overlap in this region, and the primary interaction vertex is selected as the one closest to the charged lepton. The $\not\!\!E_T$ is then defined as:

$$\not\!\!E_T = \left| \sum_i -\vec{n^i} E_T^i \right|, \tag{4.5.1}$$

where the sum on i runs over all calorimeter towers, $\vec{n^i}$ is a vector pointing from the vertex to the face of the ith calorimeter tower, and E_T^i is the E_T in the ith tower. In events with muons, the measured p_T of the muon track is used in the $\not\!\!E_T$ calculation, and not the energy deposited by the muon in the muon chambers or calorimeters. All loose and tight jets have relative, MI and absolute corrections applied. Events are required offline to have a corrected $\not\!\!E_T > 20$ GeV to pass the event selection.

4.6 B-tagging

The b hadrons produced in the fragmentation of b quarks have an average lifetime on the order of 1.5 ps. Given typical boosts in $t\bar{t}$ events at the Tevatron, this implies that b hadrons on average travel several millimeters in the detector before decaying, leading to the striking signature of a multitrack vertex displaced from the $p\bar{p}$ collision point. The lifetime for b hadrons is so long because $|V_{tb}|$ is very nearly unity. The hadron would prefer to decay through a flavor-changing process with a top quark in the final state, but cannot do so because the top is so much more massive than the b. The hadron must decay, then, through $b \to c$ or the even more surpressed $b \to u$ transitions. If the displaced vertex from a b hadron decay can be observed with the silicon detector, a jet can be tagged as coming from a b quark, reducing both the combinatorics of jet-parton assignments as well as light-flavor backgrounds, since light-flavor quark jets do not produce secondary vertices.

This analysis uses the SECVTX algorithm [27] to find jets containing secondary vertices consistent with the decay of a b hadron. The algorithm searches inside a jet for good-quality tracks with both COT and silicon hits that form secondary vertices. The distance in the $r-\phi$ plane between the possible secondary vertex and the primary vertex (L_{xy}) is calculated, along with the uncertainty on L_{xy}, $\sigma(L_{xy})$. If $L_{xy}/\sigma(L_{xy}) > 7.5$, the jet is tagged as coming from the decay of a b quark. Only tight jets are allowed to have b-tags, and tags are considered only on the four leading (highest E_T) jets. Figure 4.10 shows a toy picture of a jet being tagged, and Figure 4.11 shows the $\sim 40\%$ per-jet efficiency to tag a b quark in $t\bar{t}$ events. Jets arising from charm quarks have finite lifetimes as well, and have a per-jet tagging efficiency of roughly 8%. The efficiency to tag a b quark falls off at higher E_T because extremely energetic jets have a very high track density, making it hard to find good quality tracks. The efficiency rapidly falls off past $|\eta_{det}| > 1.0$, where coverage from the COT disappears.

4.7 Sample division

The Lepton+Jets decay channel of $t\bar{t}$ events has 4 quarks in the final state, so we ask offline for at least 4 jets in the detector. Tightening the requirement to events having exactly 4 jets increases the fraction of good, clean signal events in the sample, but removes some real $t\bar{t}$ events with initial state radiation gluon (ISR) or final state gluon radiation (FSR). Only events with at least one b-tag are considered; the 0-tag sample has a rather substantial background from production of W bosons with 4 or more jets, and also suffers from large systematic uncertainties. For events with single tags,

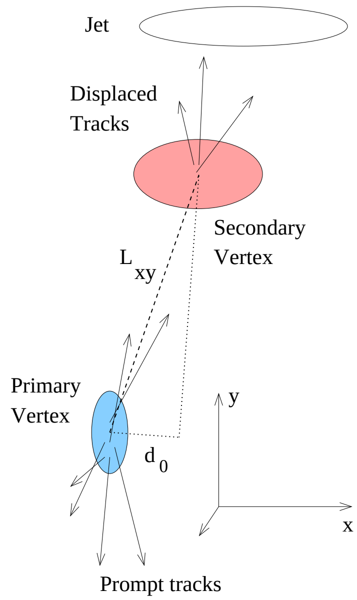

Figure 4.10: Toy picture of a jet tagged as arising from the hadronization and fragmentation of a b quark.

Table 4.1: Event selection summary. Jets are corrected to the particle jet level.

	1-tag	2-tag
b-tags (Leading 4 jets)	==1	> 1
Lepton p_T (GeV/c), E_T (GeV)	>20	> 20
$\not{E}_T$ (GeV)	>20	> 20
Leading 3 jets E_T (GeV)	>20	>20
4th jet E_T (GeV)	>20	>12
Extra jets E_T (GeV)	<20	Any

exactly 4 tight jets are required; events with additional tight jets and events with only 3 tight jets but 1 or more loose jets were found to have very little statistical power. Events with two or more tags have extremely small background contamination, and also have fewer ways to assign jets-to-partons, so we allow events with 3 tight jets and one or more loose jets, as well as events with 4 or more tight jets. Event selection is summarized in Table 4.1.

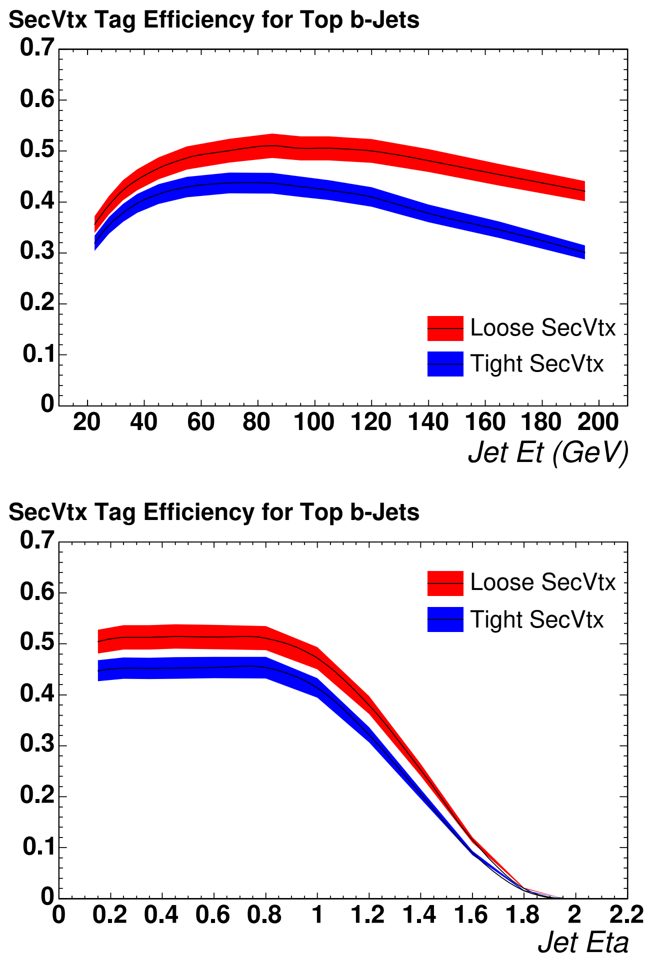

Figure 4.11: Per-jet efficiency to tag a b-jet in $t\bar{t}$ events, as a function of jet E_T (top) and jet η_{det} (bottom). This analysis uses the tight tagger, corresponding to the curves of lower-efficiency.

CHAPTER 5

TEMPLATES

After selection and reconstruction, the data give 4-vectors for the jets, leptons and $\not\!\!E_T$ in each event. This information must somehow be processed to form an estimator for the top quark mass. Unfortunately, little at the Tevatron is quite as clean as implied by simple lists of 4-vectors. Jets from the quarks in the event can be lost down the beampipe or into cracks in the detector, or they can merge with one another. Additional jets can arise from ISR off of one of the incoming partons, or from FSR off of one of the final-state quarks in the event. Additional charged leptons can appear from the semi-leptonic decay of a B hadron, and non-Lepton+Jets events can enter our event sample when charged leptons and jets are not found in the detector or fake one another, or when taus decay to electrons or muons. There is also the daunting task of trying to assign correctly the jets in the detector to the partons at the hard scatter level, the additional problem of mistagged light quarks, as well as the need to make an unbiased measurement in the presence of non-negligible background.

We solve these problems by constructing a quantity strongly correlated to the top quark mass and comparing the data to MC predictions that include all of the above effects. This quantity will be called the reconstructed top quark mass ($\mathrm{M}_t^{\mathrm{reco}}$). A set of MC predictions is referred to as a template, which is simply a probability density estimate derived from a given MC sample. Thus, a measurement of the top quark mass is a determination of the most likely parent template for the data. Further complicating the analysis, however, is the strong correlation between the JES in the detector and quantities sensitive to the top quark mass, including M_t^{reco}. When a top quark decays to a W boson and a b quark, the narrow width and large mass of the W boson ensures that most of the mass information is carried by the b quark. As explained in Section 4.4, scaling measured jet energies back to original parton energies is a difficult task, and any uncertainty on the JES directly translates to a systematic uncertainty on the measured top quark mass. To reduce this effect, we introduce a second template that uses the hadronic decay of the W boson to make an *in situ* measurement of the JES in the detector that can be applied to all jets in the event, including those from b decay. The narrow resonance of the W will make its dijet mass (W_{jj}) a good estimator for Δ_{JES}. Using PYTHIA version 6.216 as the MC generator [28], we produce both types of templates (M_t^{reco} and W_{jj}) in a grid of possible values of M_{top} and Δ_{JES}, and the final measurement is then a comparison of the sets of two numbers in data to the templates.

5.1 χ^2 Kinematic Fitter

The Lepton+Jets decay channel gives overconstrained kinematics for the $t\bar{t}$ system. There are 24 quantities that define the 4-vectors of the 6 decay objects. The 4-vectors of the 4 jets and the single charged lepton in the event give $4 \cdot 5 = 20$ measurements. We also measure the $\not{E}_T$, which has 2 components, $\not{E}_x$ and $\not{E}_y$, bringing the number of measurements up to 22. We know that the mass of the neutrino is essentially 0, bringing the total up to 23. Finally, there are several constraints that we can apply to the system: the invariant mass of the two jets from the hadronic W decay must give back the W mass, within resolution effects and the W width. Similarly, the invariant mass of the charged lepton and the neutrino must also give the W mass. Finally, in reconstructing the kinematics of the system, we can demand that the mass of the top quark equal the mass of the anti-top quark. These 3 constraints, combined with our 23 measurements, overconstrain the system of 24 quantities, and allow for a 2C-fit.

MINUIT is used to minimize a χ^2-like function[1] for the overconstrained kinematic system:

$$\begin{aligned}
\chi^2 &= \Sigma_{i=\ell,4jets}\frac{(p_T^{i,fit}-p_T^{i,meas})^2}{\sigma_i^2}+\Sigma_{j=x,y}\frac{(U_j^{fit}-U_j^{meas})^2}{\sigma_j^2} \\
&+ \frac{(M_{jj}-M_W)^2}{\Gamma_W^2}+\frac{(M_{\ell\nu}-M_W)^2}{\Gamma_W^2} \\
&+ \frac{(M_{bjj}-M_t)^2}{\Gamma_t^2}+\frac{(M_{b\ell\nu}-M_t)^2}{\Gamma_t^2}.
\end{aligned} \tag{5.1.1}$$

The first term constrains the p_T of the lepton and the 4 jets in the event to their measured values, within their uncertainties. The second term constrains the x and y components of the unclustered energy in the detector close to their measured values, also within uncertainties. The unclustered energy is the energy in the calorimeter not associated with the primary lepton, one of the jets, or the $\not\!\!E_T$. The neutrino momentum is not a direct parameter in the χ^2 minimization, but is instead related to the unclustered energy, and is calculated at every stage of the minizmiation process:

$$p_{x,y}(\nu) = -\left(p_{x,y}(\text{jets}) + p_{x,y}(\text{lepton}) + U_{x,y}\right). \tag{5.1.2}$$

The third term in the χ^2 constrains the dijet mass of the 2 jets assigned as W decay daugters to be near the well measured W mass. The fourth term similarly constrains

1. Note that the parameters of the functions are correlated. The χ^2 after minimization does not strictly follow a χ^2 distribution, but the overall shape is not used in the analysis. One way to think of the χ^2 is as a test for agreement with the $t\bar{t}$ hypothesis for a particular jet-quark assignment.

the invariant mass of the leptonic W decay daugters. The last two terms constrain the invariant masses of the three-body top decay daugters to be consistent within the top quark decay width. The value of $\mathrm{M}_{\mathrm{top}}$ is a free parameter in the fit, and is taken as the reconstructed mass used in the templates. Since the detector resolutions are much wider than the natural lines widths of the top and W, the line width can be approximated by a Gaussian.

Assuming that the leading (most energetic) 4 jets in the detector come from the 4 final partons at the hard scatter level[2], there are 12 ways to assign the jets in the detector to the partons at the hard-scatter level.[3] In addition, we need some way to initialize the value of the longitudinal momenutum of the neutrino - all other parameters in the χ^2 minimization have initial values set to their measured values (the initial value of the top quark is set to 175 GeV/c^2, though this does not strongly affect the minimization process), but the neutrino p_z is an unconstrained parameter in the fit. To start off the minimization in a reasonable corner of phase space, the parameter is initially set such that the invariant mass of the charged lepton and the neutrino gives the W mass. In doing so, there is a sign ambiguity due to solving a quadratic equation. We examine both solutions as unique possibilities (though they often converge on the same value of $\mathrm{M}_t^{\mathrm{reco}}$), giving a total of $12 \cdot 2 = 24$ possible jet-parton assignments for each event. The fit minimization is performed on all combinations, with M_t^{reco} taken from the combination giving the best (lowest) χ^2. A χ^2 cut of 9 is applied to reject poorly reconstructed events not fitting the $t\bar{t}$ hypothesis: any event with

2. We could try applying any extra, energetic non-leading jets as coming from one of the original partons, but this significantly increases the combinatorics, worsening the power of our templates.

3. There are $4 \cdot 3 \cdot 2$ ways to assign the 4 jets to the 4 partons, but the W decay daughters are identical from the point of view of jet-parton assignment, and thus we only have $\frac{4 \cdot 3 \cdot 2}{2}$ $= 12$ possible assignments.

$\chi^2 > 9.0$ is removed from the sample. B-tagging reduces the combinatorics - any leading jet that is tagged is allowed assignment only to a b parton. In rare events with more than two tags among the leading 4 jets, only two tags are considered in any jet-parton assignment, so that the event is not rejected. Not counting the factor of two in combinatorics from the neutrino p_z sign ambiguity, there are 6 possible jet-parton assignments in one-tag events, and 2 possible jet-parton assignments in two-tag events.

5.1.1 Top-specific Corrections

The jet corrections described in Section 4.4 are generic alorithms derived for application in all high-energy CDF analyses. As such, they miss out on several key features of $t\bar{t}$ events in the Lepton+Jets channel. The generic jet corrections assume flat p_T spectra for all jets, meaning that an 700 GeV/c^2 parton is assumed to be just as likely as a 75 GeV/c^2 particle, which is assumed to be just as likely as a 1.5 GeV/c^2 parton. This is not true for $t\bar{t}$ events, which have well defined and different p_T spectra for both the W decay jets as well as for the b jets.[4]

The generic jet corrections also do not account for differences between jets coming from b quarks and jets coming from more generic light-flavor quarks. Jets coming from heavy flavor quarks have different types of fragmentation and hadronization,

4. Imagine a jet measured as 700 GeV/c^2 after all jet corrections are applied. Let's naively assume that our resolution on this jet is 100 GeV/c^2. If we assume a flat p_T spectrum, we are essentially saying that the 700 GeV/c^2 jet is just as likely to be a mismeasured 800 GeV/c^2 parton as it is to be a mismeasured 600 GeV/c^2 parton. This is wrong, however, as we know that jets with such high energy are extremely rare at the Tevatron. If, however, we fold in an assumed p_T spectrum, then the 700 GeV/c^2 jet is seen much more likely to be the result of a lower energy jet fluctuating high, and much less likely to be the result of a jet fluctuating low in our measurement. Thus, the assumed p_T spectrum influences the most probable value of the original parton energy, and a flat spectrum is not a good choice.

and may have semi-leptonic decays that include escaping neutrinos. There is also a different flow of QCD color in b jets and W decay jets - the b jets can be color connected via strings of gluons to other particles, whereas the W decay daughters are color singlets. The Lepton+Jets channel is also a very busy environment - the 4 jets in an event all crowd into the detector, and there is a chance for overlap of the jets and the possibility that such a hard collision alters the underlying event structure. The χ^2 minimization should perform better if this knowledge is included in our jet corrections, as well as in the uncertainties on each jet, represented by the σ terms in the χ^2. To account for all these effects, jet corrections specific to the Lepton+Jets $t\bar{t}$ environment are derived and applied. These top-specific (TS) corrections are derived from millions of $t\bar{t}$ MC events created by the PYTHIA event generator and run through the full CDF detector simulation. Input masses from 30 different top mass samples ranging from $160 - 180$ GeV$/c^2$ are used to increase statistics in the derivation of the corrections. Only Lepton+Jets events where all 4 partons cleanly and uniquely match to jets within a cone of $\Delta R = 0.2$ are used. The nominal event selection procedure is applied, giving over 380,000 $t\bar{t}$ events, and over 760,000 each of light quark and b-quark jet-parton matches.

After event selection, relative, multiple interaction and absolute corrections are applied to jets. The corrected jets are divided by flavor (whether they came from a b quark or a light quark), and then into 6 bins of η. The η bins have different expected p_T distributions, and though the relative corrections should make the average response uniform in η, the average resolution and jet p_T are not uniform. The bins in $|\eta|$ are 0-0.2, 0.2-0.5, 0.5-0.8, 1.2-1.6 and 1.6-2.0. Additionally, the jets in each η bin are divided into 10 p_T bins. Response functions are plotted in each η-p_T-flavor bin, where the response is given by:

$$\text{Response} = \frac{p_T(\text{parton}) - p_T(\text{jet})}{p_T(\text{jet})} = \frac{p_T(\text{parton})}{p_T(\text{jet})} - 1. \qquad (5.1.3)$$

The key to the top-specific corrections, then, is to correct the response of a jet such that it returns zero. Of course, this can only be done in some probabilistic manner, and there is a choice as to whether the final corrections should return an average response of 0, a median response of 0, or some other statistic associated with 0. The corrections are derived such that the most probable value (MPV) of the jet is the parton level quantity. The MPV is a somewhat ill-defined quantity, but we try to correct such that the peak of the response function is zero. The mean and median are not good choices, as the response function can have large asymmetric tails, and the most important events for top mass resolution are those in the core of the distribution. To find the MPV, we fit a double Gaussian to each response function, and select the MPV as the peak of the larger Gaussian. Response functions for central light quark jets are shown in Figure 5.1, and for forward light quark jets in Figure 5.2. Similar plots for b quark jets are shown in Figures 5.3 and 5.4.

The MPV for the different η regions are shown in Figures 5.5 and 5.6, along with four-parameter fits to $R = e^{a+bp_T+c/p_T} + d$.

To get the value of the response function for any arbitrary η and p_T, the values of the four parameters are fit as a function of η. Given the p_T, η and flavor hypothesis of any jet, the four-vector can be corrected back to the MPV of the parton level by rearranging the response function equation:

$$p_T(\text{parton}) = (\text{Response} + 1) \cdot p_T(\text{jet}). \qquad (5.1.4)$$

The result of the above parameterizations are used as inputs to a single 2d fit to the response as a function of p_T and η. This should take into account the correlations

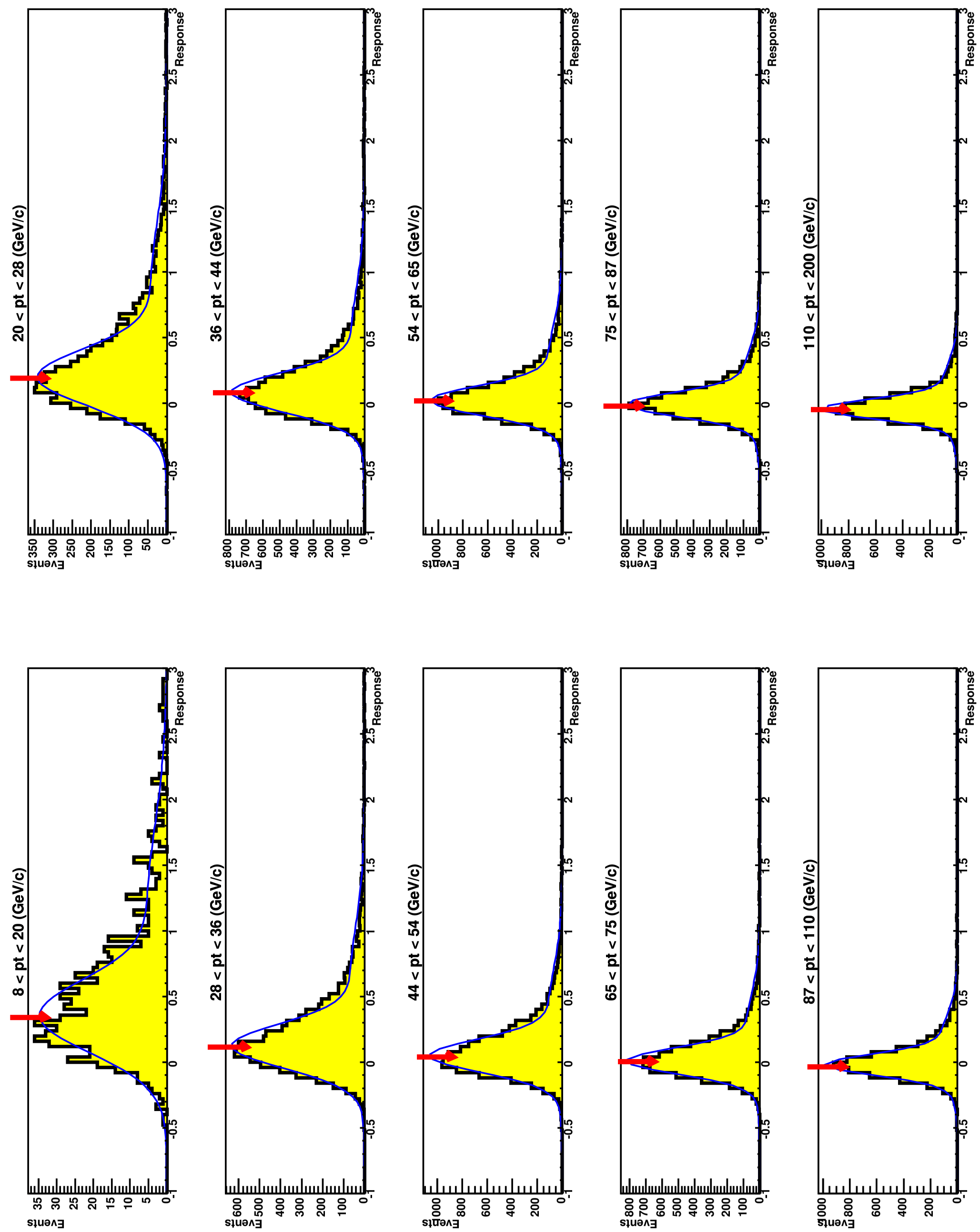

Figure 5.1: Response functions (response vs # events) for light quarks jets with $|\eta| < 0.2$ in different p_T regions. Overlaid are the fits, along with the MPV determined in each distribution.

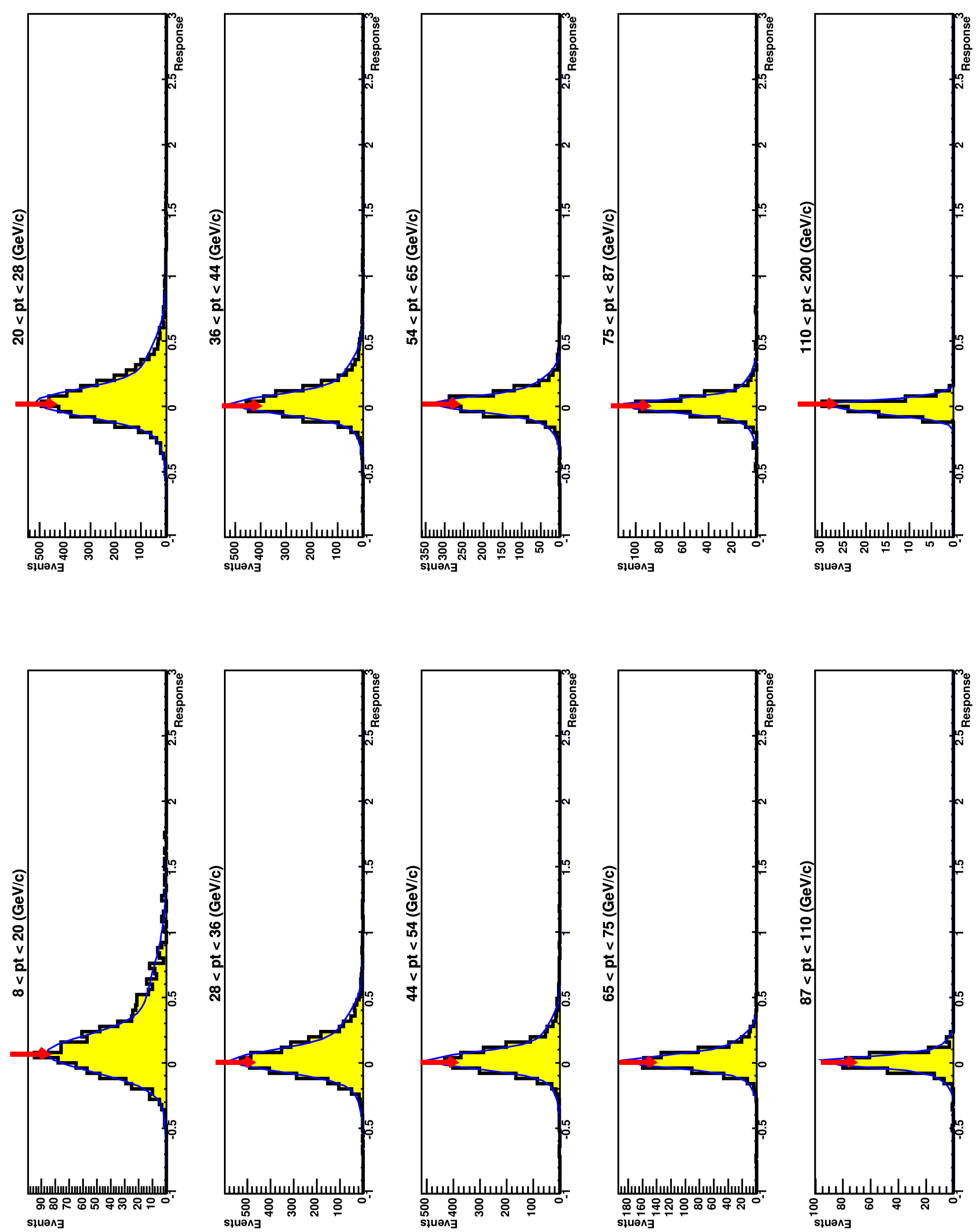

Figure 5.2: Response functions (response vs # events) for light quarks jets with $1.6 < |\eta| < 2.0$ in different p_T regions. Overlaid are the fits, along with the MPV determined in each distribution.

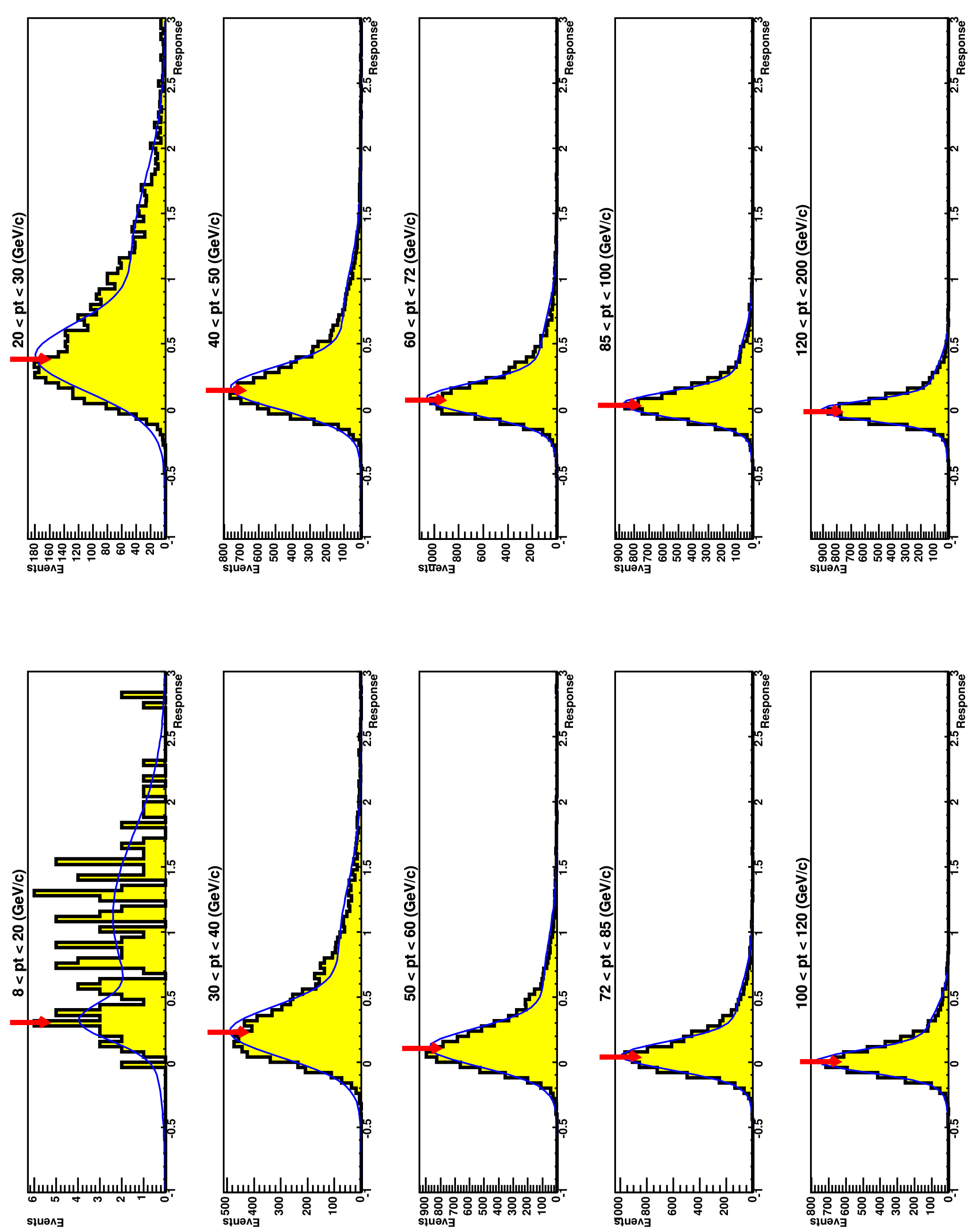

Figure 5.3: Response functions (response vs # events) for b jets with $|\eta| < 0.2$ in different p_T regions. Overlaid are the fits, along with the MPV determined in each distribution.

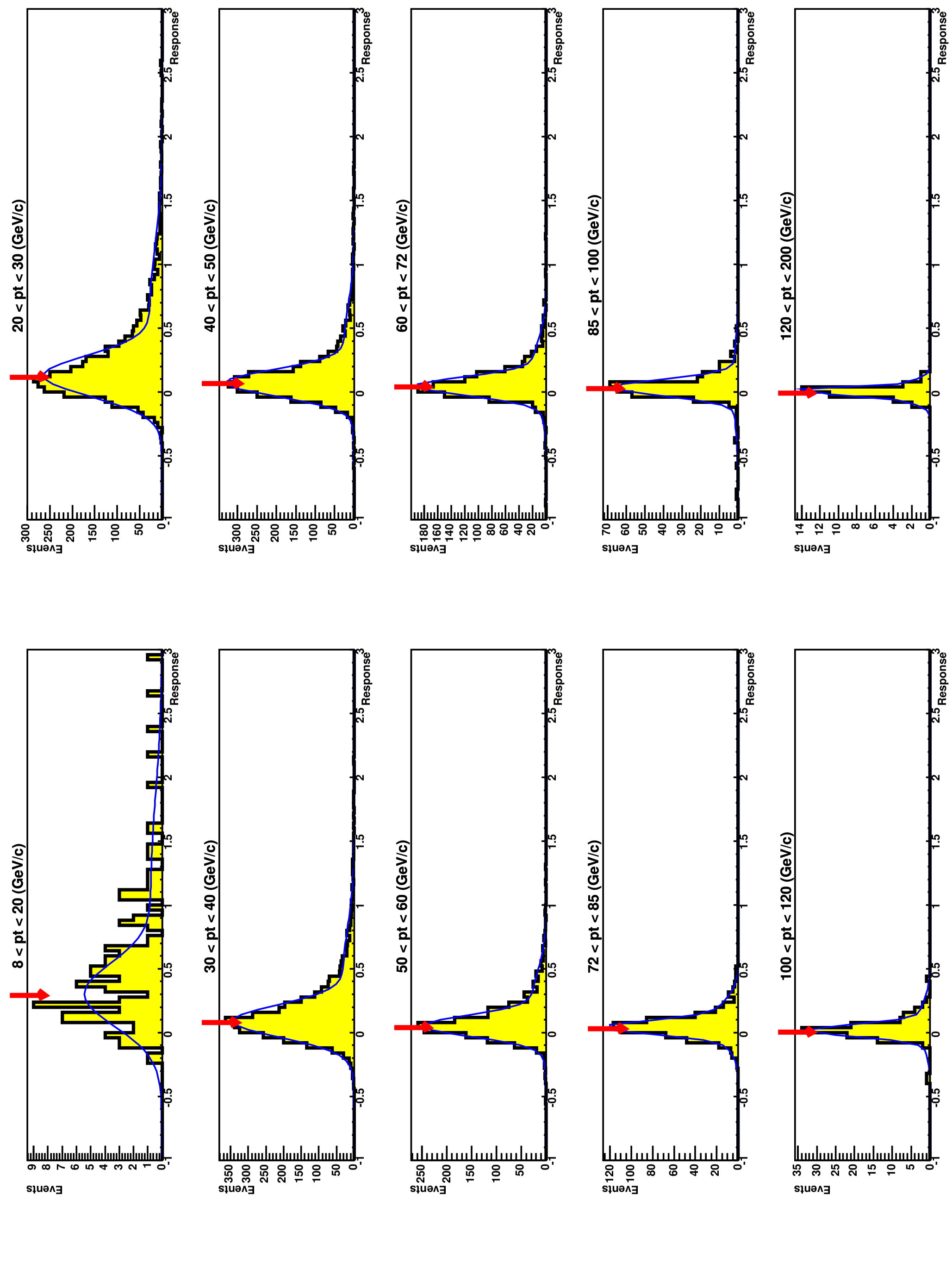

Figure 5.4: Response functions (response vs # events) for b jets with $1.6 < |\eta| < 2.0$ in different p_T regions. Overlaid are the fits, along with the MPV determined in each distribution.

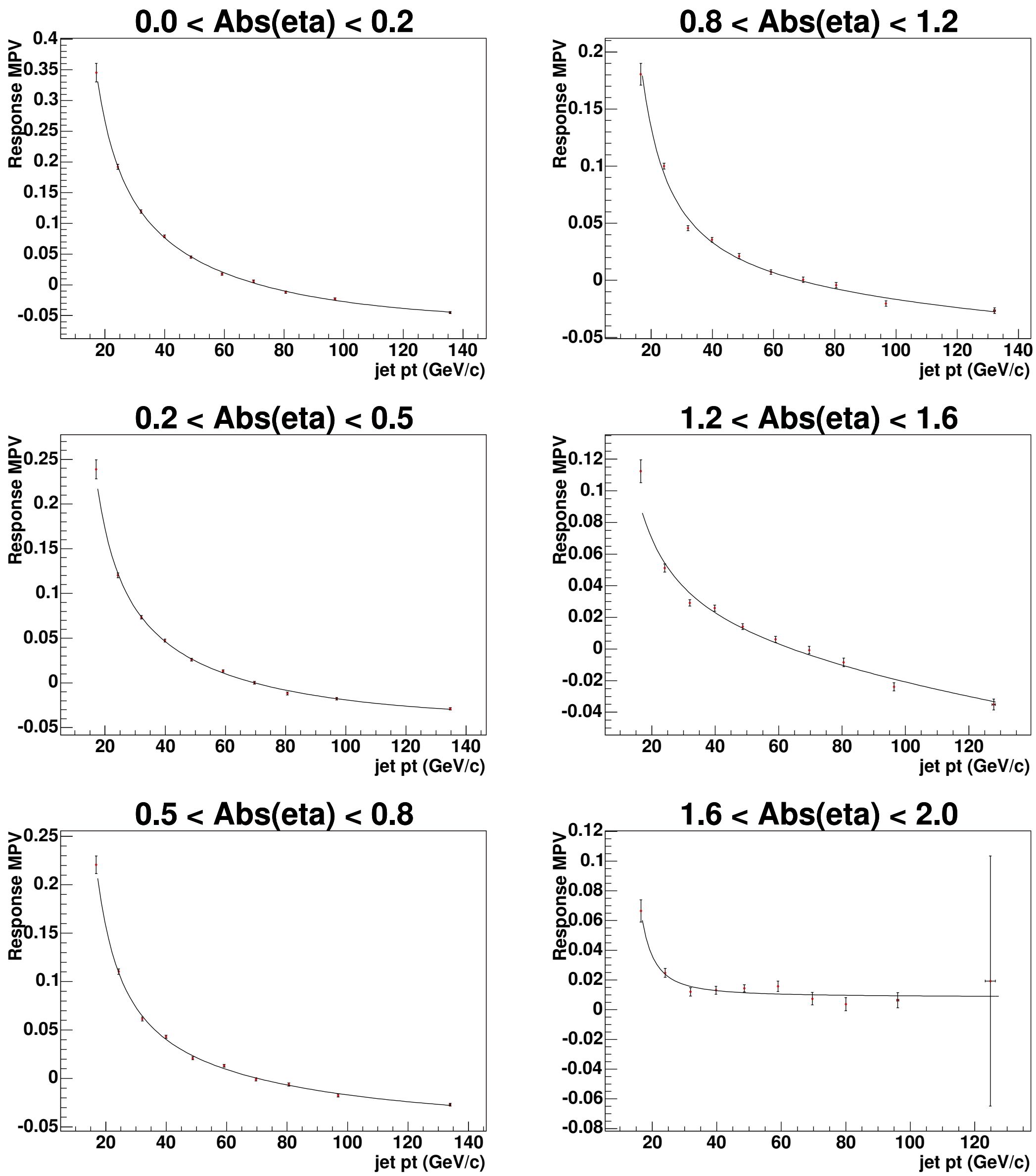

Figure 5.5: MPV of the response for light quark jets as a function of jet p_T.

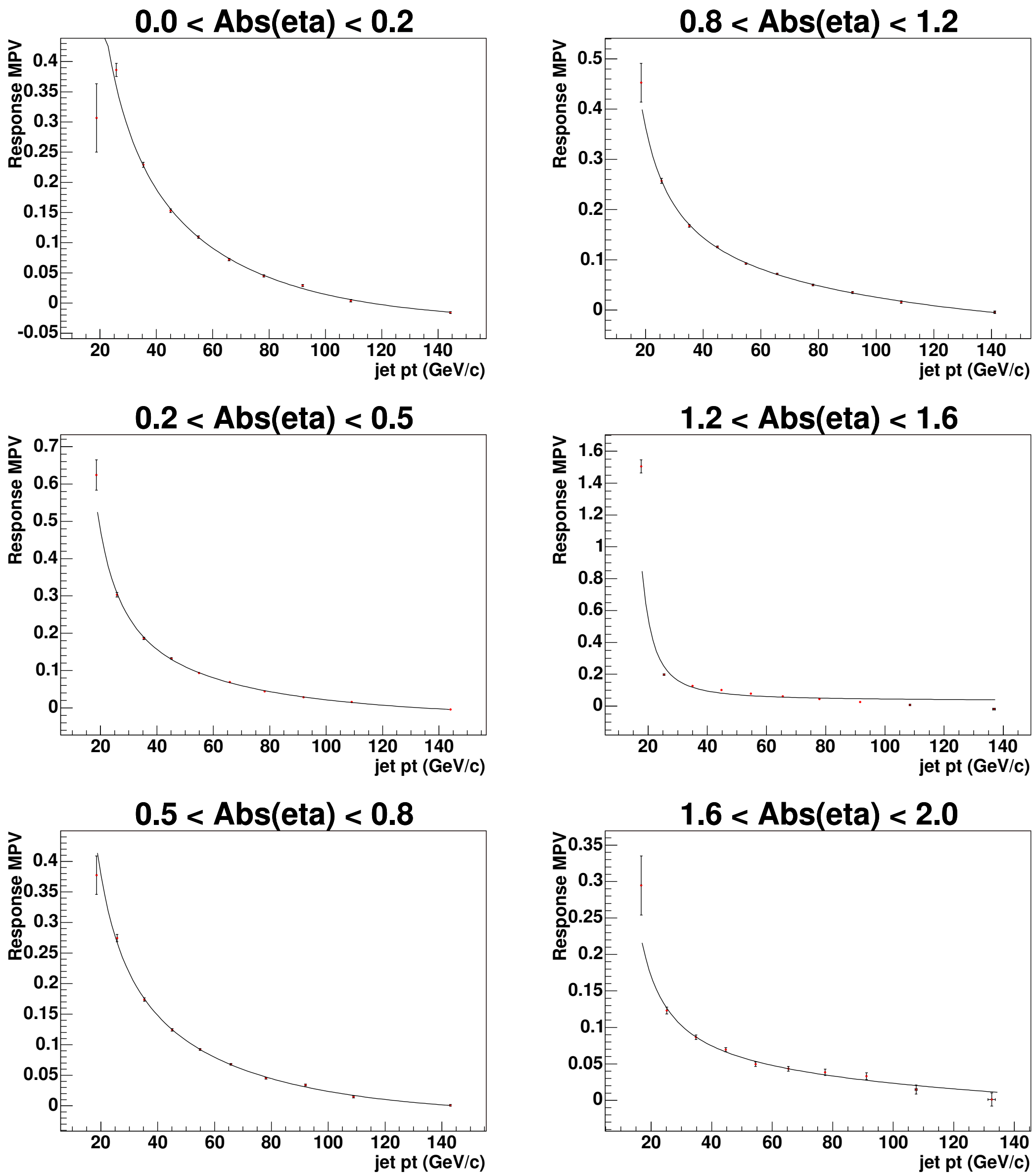

Figure 5.6: MPV of the response for b quark jets as a function of jet p_T.

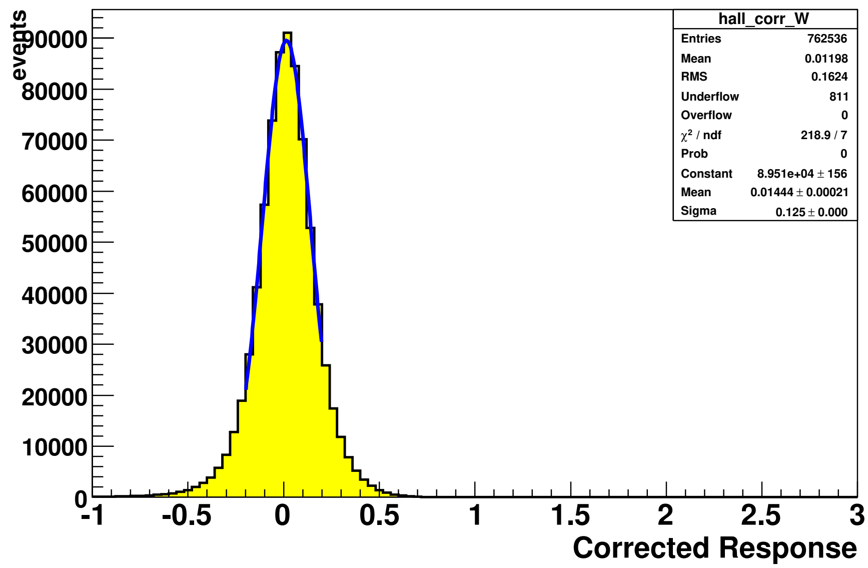

Figure 5.7: Response functions for light quarks jets after applying the final top-specific corrections.

between the variables, and improve the fit, even if visualizing the results becomes more difficult. The same functional form is used. The response functions using these final TS-corrections are shown in Figures 5.7 and 5.8.

The above corrections are applied to the p_T of the jet. The direction of the jets (η and ϕ) given to the kinematic fitter is taken from the measured quantity, with no uncertainty on the values. The mass of jets assumed to come from b quarks is fixed to 5.0 GeV/c^2 and the mass of jets assumed to come from light quarks is fixed to 0.5 GeV/c^2, though the mass effects are small compared to typical jet energies in $t\bar{t}$ events.

The TS corrections also provide the resolution terms in the χ^2 fitter. The response is once again plotted in bins of p_T and η using the fully corrected jets. Single Gaussians are fit to the distributions, with the σ of the fit taken as the resolution on the jet p_T. The resulting two-dimensional distribution is fit to the same functional

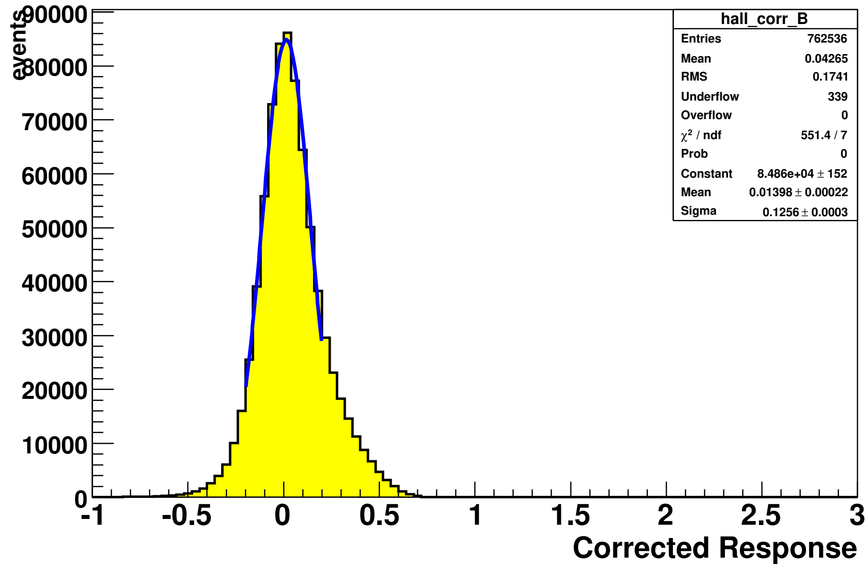

Figure 5.8: Response functions for b quarks jets after applying the final top-specific corrections.

form as was used for the original response curve, giving a resolution for a jet of any p_T and η, assuming a certain flavor hypothesis.

5.1.2 Other input and parameters to kinematic fitter

Leptons in the fitter are treated similarly to jets - the direction of the leptons is assumed to have zero uncertainty, and the mass is fixed to 0, a very good assumption given the small masses of the electron and muon compared to the typical lepton energies. Only the p_T of the leptons is allowed to fluctuate around the measured value within uncertainties.

The uncertainty on the muon p_T is given by the tracking resolution:

$$\frac{\sigma_{p_T^\mu}}{p_T^\mu} = 0.0011 \cdot p_T^\mu[\mathrm{GeV/c}]. \tag{5.1.5}$$

The resolution on the electron p_T is given by the resolution of the EM calorimeter:

$$\frac{\sigma_{p_T^e}}{p_T^e} = \sqrt{\left(\frac{0.135}{\sqrt{p_T^e[\mathrm{GeV/c}]}}\right)^2 + (0.02)^2}. \tag{5.1.6}$$

The uncertainty on the unclustered energy in the event is taken from studies of minimum bias events, and is given by:

$$\sigma^{\mathrm{UE}} = 0.4 \cdot \sqrt{\sum E_T^{\mathrm{cal}}[\mathrm{GeV}]}, \tag{5.1.7}$$

where the sum is over all calorimeter towers, and is a measured of the total transverse energy in an event. The resolution out of the kinematic fitter is insensitive to the defintion of σ^{UE}.

The mass of the W is fixed to 80.42 GeV/c^2, and the width of the W is fixed to 2.12 GeV. The top width is taken as a constant value, 1.5 GeV.

5.1.3 Kinematic Fitter Results

Monte Carlo events with different input values of $\mathrm{M_{top}}$ are sent through full detector simulation and have all corrections applied. After running through the kinematic fitter, templates for the 1-tag and 2-tag samples are shown in Figures 5.9 and 5.10. Though there is a strong correlation between the peak of the templates and the input value of $\mathrm{M_{top}}$, they are not the same thing, and $\mathrm{M}_t^{\mathrm{reco}}$ is only an estimator for $\mathrm{M_{top}}$. The 2-tag subsample, with fewer jet-parton assingments, has narrower templates, making its events more useful than 1-tag events.

5.1.4 Efficiency of χ^2 cut

Figure 5.11 and 5.12 show the efficiency for passing the $\chi^2 < 9.0$ cut for signal events. The efficiency does have some dependence on $\mathrm{M_{top}}$ and Δ_{JES}, but the effect is small,

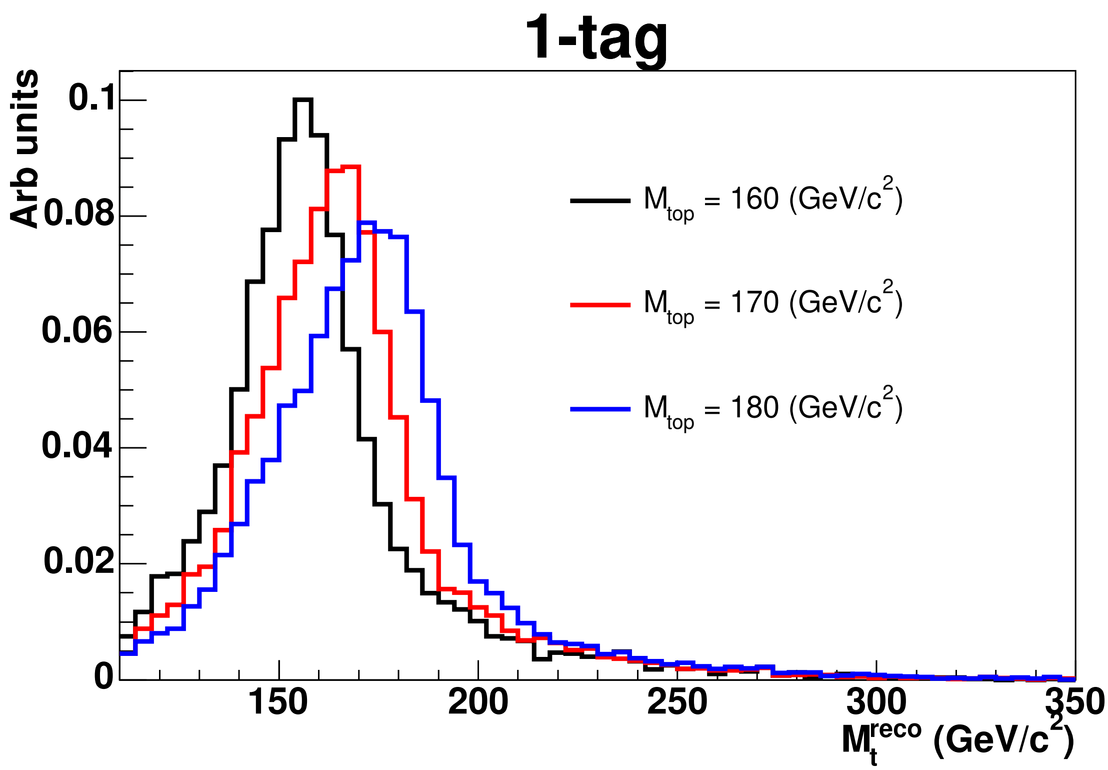

Figure 5.9: Output of the kinematic fitter for MC events passing the 1-tag event selection, after the χ^2 cut is applied.

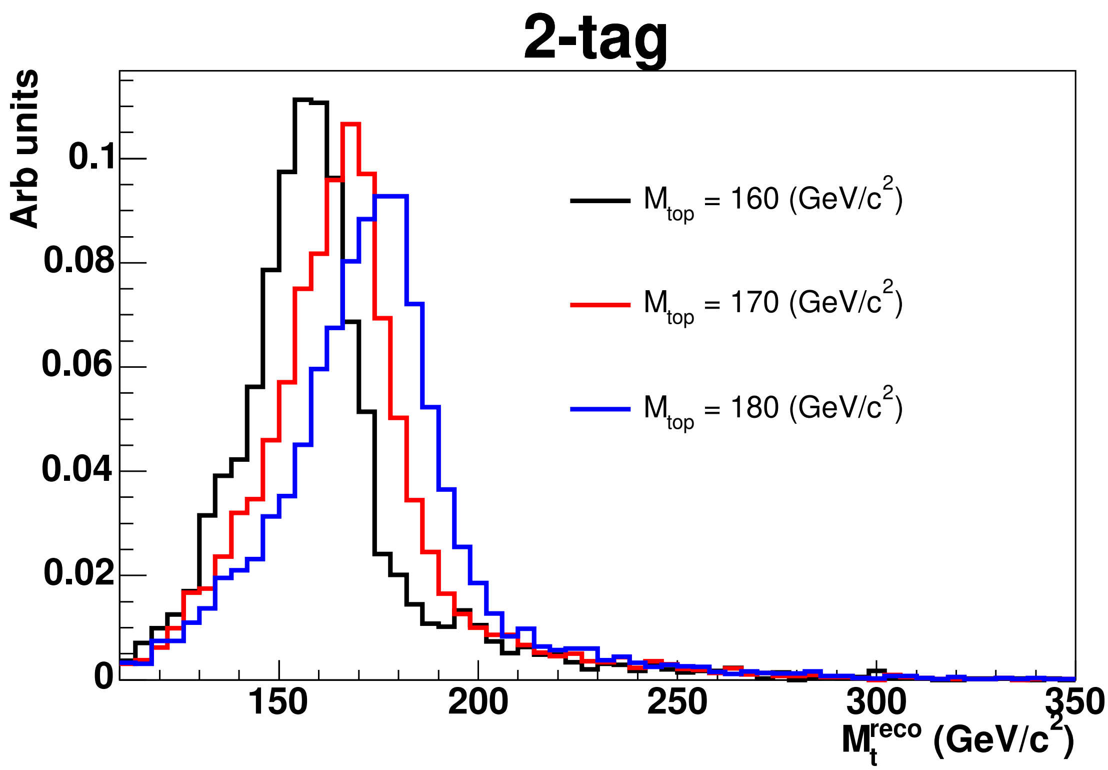

Figure 5.10: Output of the kinematic fitter for MC events passing the 2-tag event selection, after the χ^2 cut is applied.

and the number of signal events passing all cuts does not enter into the likelihood as an input. The efficiency is lower for 2-tag events, which have fewer possible jet-parton assignments, and therefore fewer chances to pass thie χ^2 cut. An attempt was made to optimize the value at which the χ^2 cut is made, but the expected statistical uncertainties from the two subsamples are very insensitive to the cut value over a wide range of χ^2, so 9.0 was chosen for consistency with previous measurements.

5.2 Dijet template

There are many possible ways of forming a dijet mass from two jets among the four or more jets in $t\bar{t}$ events. To simplify the list of choices, jets tagged as arising from b quarks are assumed to come from real b quarks, and not from mistags. Additionally, the two jets from the W decay quarks are assumed to be among the leading four jets. For two-tag events, there are only 2 non-tagged jets from among the leading 4 jets, giving only one choice for the dijet mass. The jets are corrected using the light quark TS corrections, and the single dijet mass per event is used. For 1-tag events, there are three possible dijet masses to be made from the 3 non-tagged leading jets. Several variables based on jet p_T and angular distribution were studied for sensitivity to measuring Δ_{JES}, including linear combinations of multiple dijet masses. In the end, the variable with the best performance was the dijet mass closest to the well known W mass. This sculpts the distribution, but is the most likely to select the two jets from the W decay daughters. The biggest drawback to this variable is that the sculpting causes the background, most of which has no real hadronic W decay, to look very much like the signal. Templates for W_{jj} masses for three different values of Δ_{JES} are shown in Figures 5.13 and 5.14. The 1-tag W_{jj} templates are narrower than the 2-tag templates due to the sculpting. The narrowness is a good thing, but

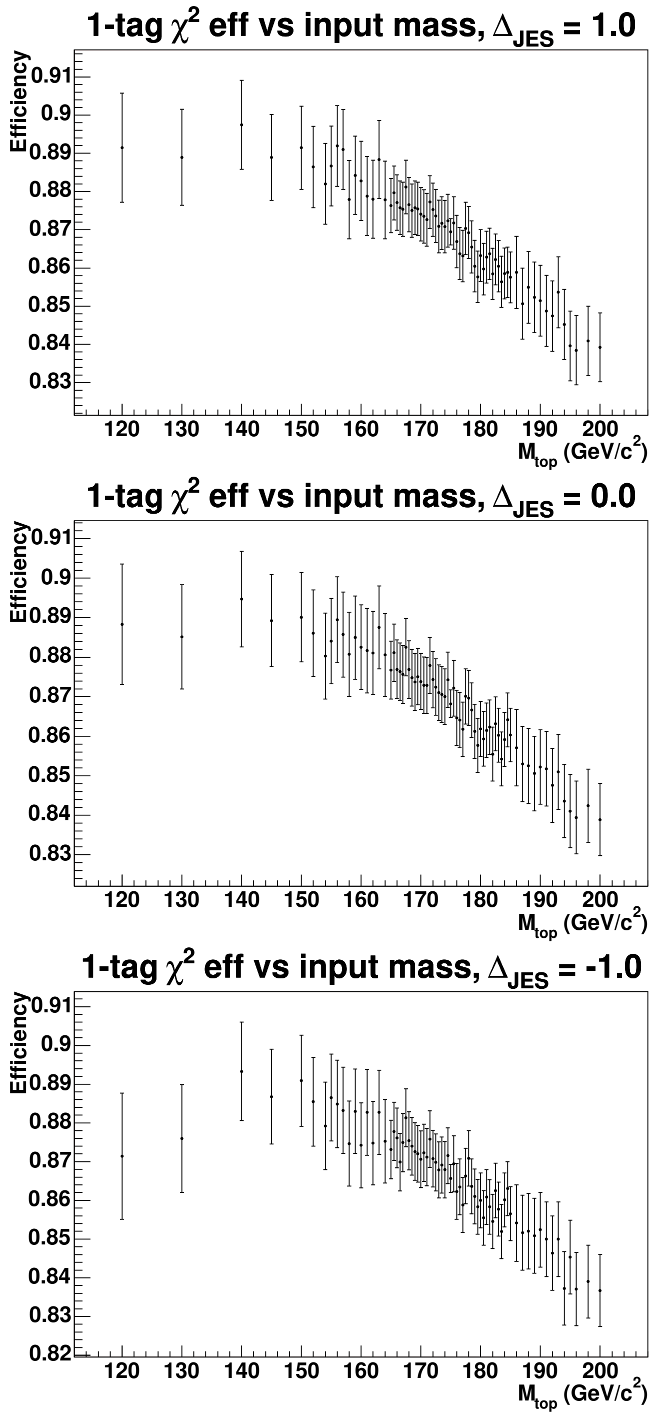

Figure 5.11: Efficiency to pass the $\chi^2 < 9.0$ cut for 1-tag signal events for Δ_{JES} = + 1.0 (top), 0.0 (middle) and -1.0 (bottom) σ_{C}. Note the surpressed zeros.

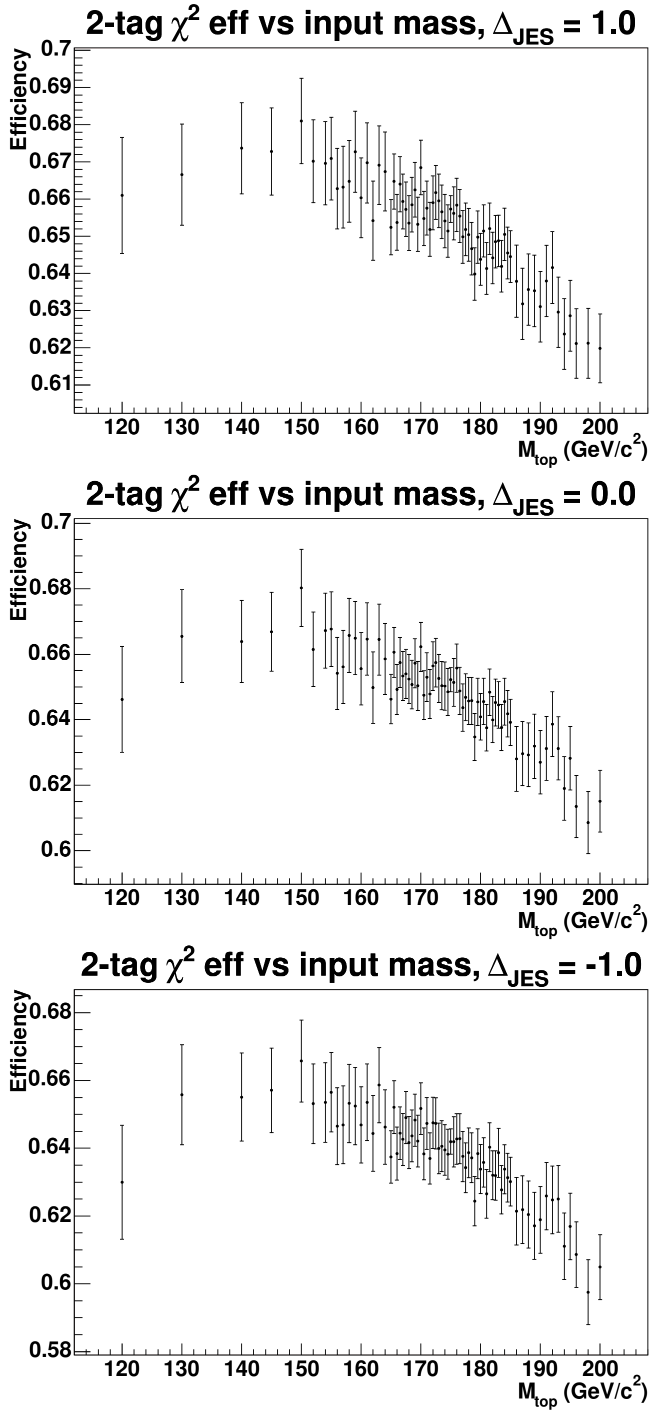

Figure 5.12: Efficiency to pass the $\chi^2 < 9.0$ cut for 2-tag signal events for $\Delta_{\rm JES}$ = + 1.0 (top), 0.0 (middle) and -1.0 (bottom) $\sigma_{\rm C}$. Note the surpressed zeros.

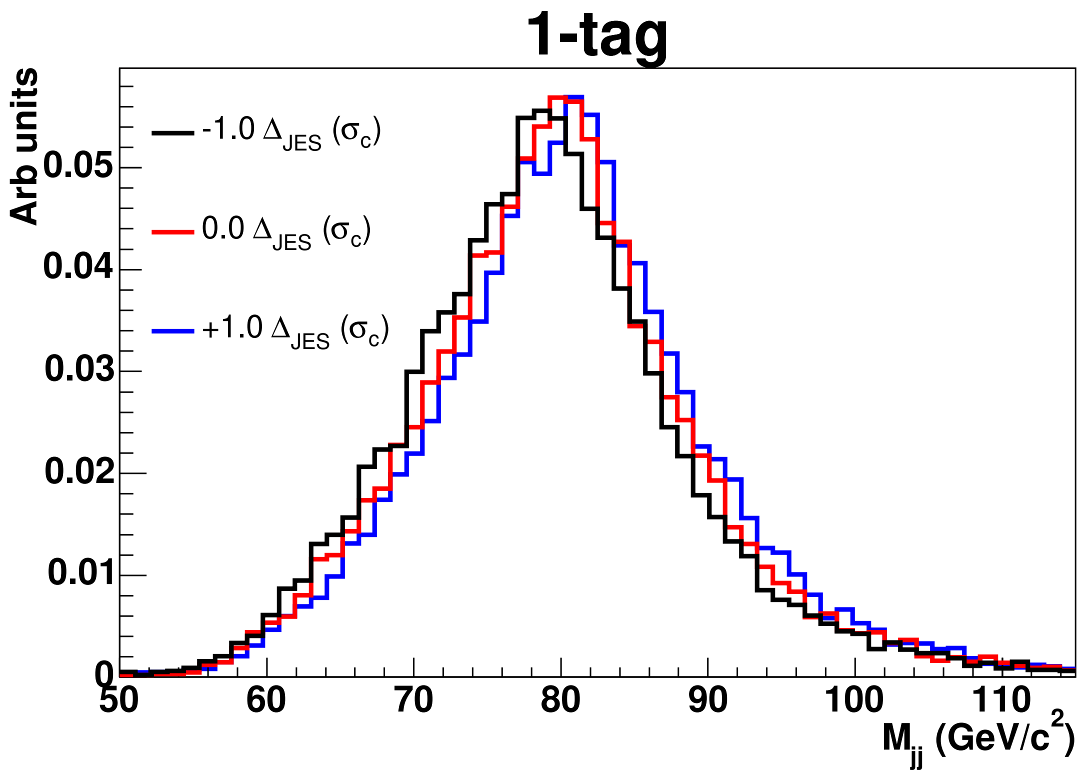

Figure 5.13: Dijet masses for MC events passing the 1-tag event selection, after the χ^2 cut is applied. The input top quark mass is 170 GeV/c^2.

the 1-tag W_{jj} templates have smaller shifts than the 2-tag W_{jj} templates as Δ_{JES} varies.

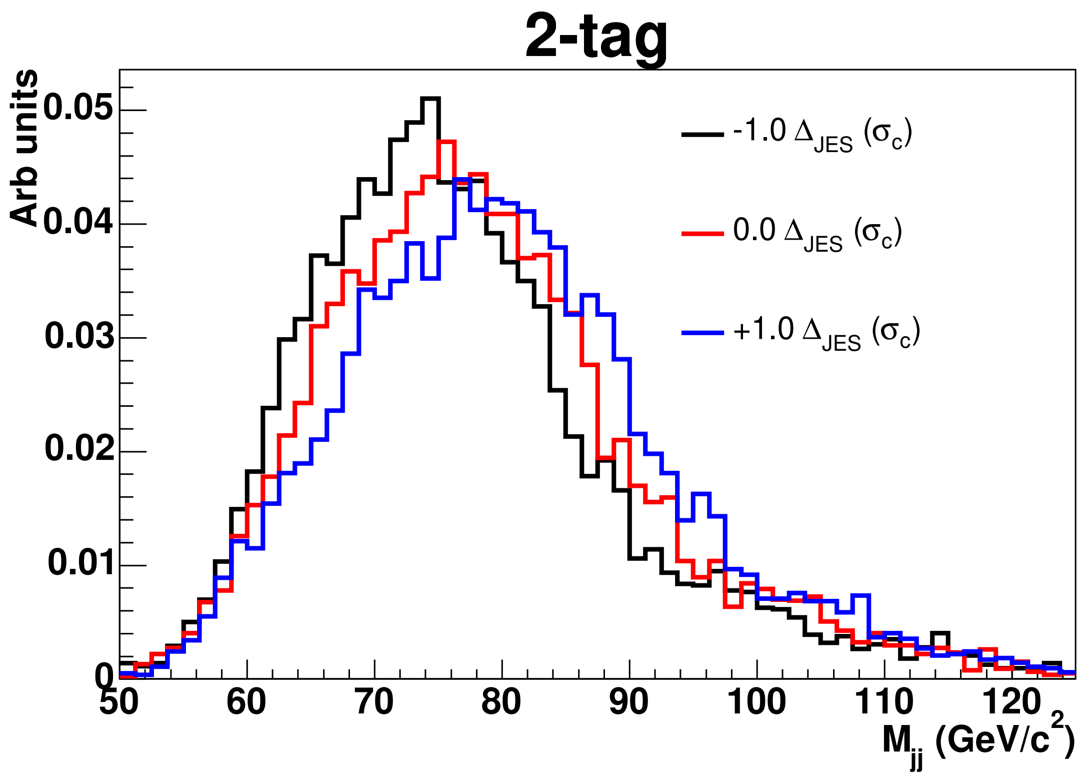

Figure 5.14: Dijet masses for MC events passing the 2-tag event selection, after the χ^2 cut is applied. The input top quark mass is 170 GeV/c^2.

CHAPTER 6

KERNEL DENSITY ESTIMATES

Without unlimited computing resources, the finite MC statistics for M_t^{reco} and W_{jj} must somehow be smoothed out to produce templates needed for the likelihood fit described in Chapter 8. Previous CDF template-based measurements of $\mathrm{M_{top}}$ parameterize the templates using an ad-hoc, arbitary function with no physical meaning [29, 30]. The parameterizations are adequate to make an unbiased measurement, but are difficult to produce with each iteration of MC, can lead to slightly different minima, and are mathematically poor fits. Perhaps more importantly, it is impossible to account for the small but non-zero correlation between $\mathrm{M}_t^{\mathrm{reco}}$ and $\mathrm{W_{jj}}$ using a parametric approach. Instead of using multiple 27-parameter functional fits, this measurement uses a non-parameteric kernel density estimate (KDE) approach.

6.1 Histograms

Much of the literature on Kernel Density Estimation (KDE) begins with a discussion of histograms as density estimators. Physicists are accustomed to thinking of histograms as tools to graphically represent data, but they are more than that: Histograms give an estimate for the density in a given bin. The value returned by the histogram does not change across the bin and the derivative of the estimate is not defined at the bin boundary, but the histogram is still an estimate of the average density in the bin. Therefore, it may help to think about histograms before moving to KDE. Some useful introductions to KDE can be found in [31, 32, 33, 34].

To be a bit more formal, consider a histogram $\hat{f}(x)$ (the template), determined by sampling an underlying distribution $f(x)$ n times ($f(x)$ is the true distribution one would get with infinite MC statistics). The histogram has constant binning width h, and n entries. For simplicity, start the bin origin at $x = 0$ so that the kth bin (B_k) of the histogram contains the points with $x \in [kh, (k+1)h]$. Bin B_k has a bin count ν_k, which is simply the number of entries falling into the kth bin (ν_k is a non-negative integer). Each entry in a histogram contributes a norm of the area of a rectangle of width h and height equal to the weight of the entry. There are n entries, so the total normalization of the histogram is $nh \cdot$ single entry weight. If the histogram is to be normalized to unity (as is necessary for probability density functions), then the weight of each event is $\frac{1}{nh}$, and

$$\hat{f}(x) = \frac{\nu_k}{nh}. \tag{6.1.1}$$

Let p_k be the probability for an entry from the underlying (ie true) distribution to fall into the kth bin. Since the distribution is sampled n times, ν_k will be distributed as a binomial random variable with n trials and probability p_k:

$$< \nu_k >= np_k, \tag{6.1.2}$$

$$\mathrm{Var}(\nu_k) = np_k(1 - p_k). \tag{6.1.3}$$

Combining Equation 6.1.1 with Equations 6.1.2 and 6.1.3 gives

$$< \hat{f}(x) >= \frac{p_k}{h}, \tag{6.1.4}$$

$$\mathrm{Var}\hat{f}(x) = \frac{np_k(1-p_k)}{(nh)^2} = \frac{p_k(1-p_k)}{nh^2}. \tag{6.1.5}$$

At this point, the formalism gives the expectation value and variance of the histogram at any point $x \in B_k$. When deciding how to choose the optimal bin width h, it will be easier to focus on global quantities integrated over the entire histogram. In the end, this may not be the best thing for top mass analyses, but global quantities are simpler, and this tends to be the route taken by most of the literature. Therefore, instead of focusing on the variance at a point, attention is paid to the integrated variance (IV), which is the variance of $\hat{f}(x)$ integrated across all possible values of x:

$$\mathrm{IV} = \int_{-\infty}^{\infty} \mathrm{Var}\hat{f}(x)dx. \tag{6.1.6}$$

This large integral can be broken up into integrals over each of the infinite number of bins (if the histogram is defined for only a finite number of bins, it can be extended to an infinite number of bins and set it equal to 0 in the extra regions):

$$\mathrm{IV} = \sum_{k=-\infty}^{\infty} \int_{B_k} \mathrm{Var}\hat{f}(x)dx = \sum_{k=-\infty}^{\infty} \int_{B_k} \frac{p_k(1-p_k)}{nh^2}dx. \tag{6.1.7}$$

The IV has two pieces (the two separate terms in the numerator of equation 6.1.7): one term is $\propto p_k$ and the other term is $\propto -p_k^2$. As we scan across a bin, p_k remains a constant (it's the probability to find an entry in that bin, regardless of where in the bin the histogram is evaluated). The integral of a constant over each bin is h times that constant (since the width of the bin is h), and $\sum p_k = 1$, so that the first term in the IV is $\frac{1}{nh}$. The second term is $\frac{-1}{nh}\sum_{k=-\infty}^{\infty} p_k^2$, so an expression for $\sum p_k^2$ is needed.

The approximation $p_k \sim h \cdot f(x_k)$ for some x_k in the kth bin is a good one, as the probability to be in a bin is equal to the value of the function somewhere in that bin, normalized by the bin width. This is a reasonable approximation if $f(x)$ is a smoothly varying function, which should be the case for the M_t^{reco} and W_{jj} distributions. Plugging this in, $\sum p_k^2 = \sum [f(x_k)h]^2$ for some x_k in each of the B_k. If the bin widths are small, another approxmation is that the sum of many small things approaches the integral of that thing: $\sum [f(x_k)h]^2 \sim h \int f(x)^2 dx$ (as $h \to 0$, $h^2 \to hdx$). Thus,

$$\text{IV} = \frac{1}{nh} - \frac{1}{n} \int_{-\infty}^{\infty} f(x)^2 dx. \tag{6.1.8}$$

Introducing new notation:

$$R(\phi) = \int_{-\infty}^{\infty} \phi(x)^2 dx, \tag{6.1.9}$$

$$\text{IV} = \frac{1}{nh} - \frac{R(f)}{n}. \tag{6.1.10}$$

Equation 6.1.10 says that as the statistics of the sample are increased, the variance decreases as 1/n, or, $\sqrt{\text{Variance}} \propto \frac{1}{\sqrt{n}}$, as might be expected. The first term also scales as 1/h - if the bin width is doubled, on average, the number of entries in each bin is doubled. As for the second term, it depends on the number of entries in the histogram and a quantity related to the true, underlying distribution, but not on h or $\hat{f}$. Since it does not depend on bin width, which in the end is the quantity of interest, we will ignore this term.

The estimator $\hat{f}(x)$ can also be biased at the point x. Let's Taylor expand the probability p_k in the kth bin. Using the first bin as an example, so that $k = 0$ and

$x \in [0, h)$, the Taylor expansion of p_k gives:

$$p_0 = \int_0^h f(t)dt = \int_0^h [f(x) + (t - x)f'(x) + ...]dt, \tag{6.1.11}$$

$$p_0 = hf(x) + h(\frac{h}{2} - x)f'(x) + ... \tag{6.1.12}$$

The bias of $\hat{f}(x) = < \hat{f}(x) > - f(x) = \frac{p_0}{h} - f(x) = (\frac{h}{2} - x)f'(x) + ...$ This is the smallest at the center of the bin, where the Order(h) term disappears, and an expansion to additional powers of h is needed to find a bias.

Also important is the square bias integrated across this bin: $\int_{B_0} (\frac{h}{2} - x)^2 (f'(x))^2 dx$, as before distilling down to one number how well the histogram performs. The mean value theorem says that this is $(f'(\eta_0))^2 \int_0^h (\frac{h}{2} - x)^2 dx = (f'(\eta_0))^2 \frac{h^3}{12}$ for some η_0 in the bin (ie $\eta_0 \in [0, h)$). This is the total square bias in the first bin. Summing up the total integrated square bias (ISB) in all the bins:

$$\text{ISB} = \frac{h^2}{12} \sum_{k=-\infty}^{\infty} hf'(\eta_k)^2 \sim \frac{h^2}{12} \int_{-\infty}^{\infty} f'(x)^2 dx + ... \tag{6.1.13}$$

Adopting the previous notation as before,

$$\text{ISB} = \frac{h^2}{12} R(f'). \tag{6.1.14}$$

How can the ISB be interpreted? As the bin width increases, the histogram becomes more susceptible to biases - in each bin, it is asked to give a single estimate for the density across an even wider range of points. The number $R(f')$ can be thought of as a way to measure the average amount that the underlying function is or is not constant in each of the bins bins. If the average slope of the underlying distribution is small, estimating the density with a single number per bin should give only small biases. If the average slope of the distribution is large, then the histogram has large biases, especially at bin edges.

A useful measure of the performance of a histogram is the sum of the h-dependent part of the IV plus the ISB. This gives a total error that is a function of h:

$$\text{AMISE} = \frac{1}{nh} + \frac{1}{12}h^2 R(f'). \tag{6.1.15}$$

AMISE stands for Asymptotic Mean Integrated Square Error (the asymptotic refers to the fact that it includes only the leading order terms in our expansions). This is used because the MISE (Mean Integrated Square Error) is a global quantity that describes us how well the histogram can be expected to perform, integrated over the entire distribution:

$$\text{MISE} = \int < [\hat{f}(x) - f(x)]^2 > dx = \int [< \hat{f}^2(x) > + f^2(x) - 2f(x) < \hat{f}(x) >] dx. \tag{6.1.16}$$

Adding and subtracting $< \hat{f}(x) >^2$ to the integral gives

$$\text{MISE} = \int [< \hat{f}^2(x) > + f^2(x) - 2f(x) < \hat{f}(x) > + < \hat{f}(x) >^2 - < \hat{f}(x) >^2] dx, \tag{6.1.17}$$

and rearranging leads to

$$\text{MISE} = \int [< \hat{f}(x) >^2 + f^2(x) - 2f(x) < \hat{f}(x) > + < \hat{f}^2(x) > - < \hat{f}(x) >^2] dx. \tag{6.1.18}$$

The first three terms are an expansion of the ISB $\int[<\hat{f}> - f]^2$ and the last two terms are the IV, giving some justification for adding the ISB and IV together with unit coefficients. Setting $\frac{d}{dh}\text{AMISE} = 0$ gives the bin width of a histogram that minimizes the total error:

$$h_{min} = [\frac{6}{nR(f')}]^{1/3}. \tag{6.1.19}$$

Plugging this back in to the AMISE gives an error that decreases at a rate of $n^{-2/3}$. Of course, choosing the optimal bin width depends on knowing a function (R) of the true distribution, and if the true distribution were known, the histogram would not be needed as a density estimator.

6.2 KDE

Histograms are great tools for visualizing data, but as density estimators they have many shortcomings - they are not continuous, suffer from effects due to the choice of origin, and give the same density estimate across an entire bin. Many of these shortcomings can be overcome, but we will move away from histograms and instead focus on kernel density estimation. Essentially, the underlying density will be estimated by smoothing out the sampling data using a smoothing function called the kernel, K(x). One way to think about doing this is placing δ-functions on the axis at every point returned by our MC (for example, the reconstructed top masses). Now smooth out the points by replacing the delta functions with normalized Gaussians centered at each point. Summing over all the Gaussians gives the density estimate. All that remains is to select the width of these Gaussians, and therein lies the rub: similar to the bin width choice for histograms, there is a parameter h (in this example, h would be the σ of the Gaussians, more generally called the smoothing parameter or bandwidth for other functions) that determines the tradeoff between variance and bias, or how much the sampling points are smoothed out. A large h will give a small variance and a large bias, whereas a small h will give a large variance and a small bias.

Formally, the density estimate becomes:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h}) \equiv \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) \equiv K_h(x - X) \equiv K_h(t) \equiv \frac{1}{h}K(t). \tag{6.2.1}$$

The kernel, K(x), is a function that determines how to smooth out the sampling data to estimate the density at the point x. The values of $x_i \in X$ are the n sampling values used to estimate the true distribution. Similar to before, we define expectation values and variances for the estimate:

$$< \hat{f}(x) > = < K_h(x - X) >, \tag{6.2.2}$$

$$\text{Var } \hat{f}(x) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var } K_h(x - x_i) = \text{Var } K_h(x - X). \tag{6.2.3}$$

where the extra factor of $\frac{1}{n}$ comes from the general rule that for constant A, Var(Ax) = A^2Var(x), and not AVar(x). Expanding $\hat{f}(x)$:

$$< \hat{f}(x) > = \frac{1}{nh} < \sum_{i=1}^{n} K(\frac{x - x_i}{h}) > = \frac{1}{nh}\int f(x) \sum_{i} K(\frac{x - x_i}{h}), \tag{6.2.4}$$

$$< \hat{f}(x) > = \int f(x) K_h(x - X) dx = \frac{1}{h}\int f(x) K(t) dx. \tag{6.2.5}$$

From Equation 6.2.1, $t = \frac{x-X}{h}$, so $x = ht + X$ and $dx = hdt$:

$$< \hat{f}(x) >= \int f(X + ht)K(t)dt. \tag{6.2.6}$$

A Taylor expansion of $f(X + ht)$ around $x = X$ gives

$$f(X + ht) = f(X) + htf'(X) + \frac{1}{2}h^2t^2f''(X) + ..., \tag{6.2.7}$$

and plugging back in to Equation 6.2.6 gives

$$< \hat{f}(X) >= [f(X) \int K(t) + hf'(X) \int tK(t) + \frac{1}{2}h^2f''(X) \int t^2K(t) + ...]dt. \tag{6.2.8}$$

There is some freedom in choosing the Kernel - it should be normalized, and it makes sense to choose a Kernel that is symmetric about 0 (ie with a vanishing first moment):

$$\int K_h(w)dw = 1, \tag{6.2.9}$$

$$\int wK_h(w)dw = 0. \tag{6.2.10}$$

It is possible to choose additional, higher-order moments to vanish, but this brings a risk of the density estimate becoming negative. Avoiding this,

$$\int w^2K_h(w)dw \equiv \sigma_k^2, \tag{6.2.11}$$

where σ_k^2 depends only on the kernel function, and not on the smoothing parameter. Combining equations 6.2.8, 6.2.9, 6.2.10 and 6.2.11 gives the bias to leading order:

$$\text{Bias(x)} = < \hat{f}(x) > - f(x) \sim \frac{1}{2}\sigma_k^2 h^2 f''(x), \tag{6.2.12}$$

and the ISB:

$$\text{ISB} = \frac{1}{4}\sigma_k^4 h^4 R(f''). \tag{6.2.13}$$

Unlike for histograms, where the lowest-order term in ISB $\propto R(f')$, the ISB for KDE $\propto R(f'')$. This is because histograms have an intrinsic bias across the bin as the function varies, whereas kernel estimate tracks linear changes in underlying distributions via the sampling distribution. In other words, if the underlying distribution has a single slope, KDE will track this via larger number of points sampled at higher densities. The density estimate develops a bias only when the second derivative of the true distribution does not vanish.

Once again following the approach with histograms,

$$\text{Var}\hat{f}(x) = < K_h(x - X)^2 > - < K_h(x - X) >^2 . \tag{6.2.14}$$

Using the same tricks as before to calculate the first term in the variance:

$$< K_h(x - X)^2 > = \frac{1}{n^2 h^2} \int f(x) \sum_i K^2(\frac{x - x_i}{h}) dx, \tag{6.2.15}$$

$$< K_h(x - X)^2 > = \frac{1}{nh^2} \int f(x) K^2(\frac{x - X}{h}) dx. \tag{6.2.16}$$

Making the change of variables once more gives:

$$< K_h(x - X)^2 > = \frac{1}{nh} \int f(X + ht)) K^2(t) dt. \tag{6.2.17}$$

Keeping only the leading-order term in the $f(X+ht) = f(x) + ...$ expansion gives:

$$< K_h(x - X)^2 > \sim \frac{1}{nh} \int f(x) K^2(t) dt, \tag{6.2.18}$$

$$< K_h(x - X)^2 > \sim \frac{f(x)R(K)}{nh}. \tag{6.2.19}$$

Using the first order term in the Taylor expansion of $< K_h(x - X) >^2$ gives:

$$\mathrm{Var}(\hat{f}(x)) = \frac{f(x)R(K)}{nh} - f(x)^2, \tag{6.2.20}$$

$$\mathrm{IV} = \frac{R(K)}{nh} - R(f), \tag{6.2.21}$$

where, similar to histograms, the second term in the IV does not depend on h, only on the underlying distribution and the number of sample points. Since as before the quantity of interest is the choice of smoothing parameter, we ignore this term and combine the remaining IV and the ISB (Equations 6.2.21 and 6.2.13):

$$\mathrm{AMISE} = \frac{R(K)}{nh} + \frac{1}{4}\sigma_k^4 h^4 R(f''), \tag{6.2.22}$$

yielding an optimal smoothing parameter of

$$h_{min} = [\frac{R(K)}{n\sigma_k^4 R(f'')}]^{1/5}. \tag{6.2.23}$$

Plugging this in gives an error that decreases at a rate of $n^{-4/5}$, which is faster than the $n^{-2/3}$ rate from histograms.

6.3 On the choice of kernel

It turns out that the choice of kernel is less important than the choice of smoothing parameter [35]. We use the Epanechnikov Kernel:

$$K_h(t) = \frac{3}{4}(1-t^2) \text{ for } |t| < 1 \text{ and } K(t) = 0 \text{ otherwise.} \tag{6.3.1}$$

In other words, $K_h(t) = \frac{3}{4}(1-t^2)$, but only points with $|x - x_i| < h$ contribute at all to the density. The Epanechnikov Kernel is normalized and has a vanishing first moment. The value of R(K) = 3/5, and $\sigma_k^2 = 1/5$.

6.4 The oversmoothing rule

A great deal of effort goes into various rules for estimating the optimal smoothing parameter h. The techniques go by names such as unbiased cross-validation, biased crossed-validation, plug-in cross-validation and bootstrapping. We adopt a more simple approach - we use an upper limit on the smoothing parameter ($h = h_{OS}$) given by the oversmoothing rule, derived below [31, 32, 36] . The value of h_{OS} will depend on the number of entries and RMS of the sampled distribution, so it will vary by MC sample and the sub-category being examined, as well as for both signal and background.

The optimal value of $h \propto \frac{1}{R(f'')^{1/5}}$ (see Equation 6.2.23), so that a lower limit on $R(f'')$ gives an upper limit on the optimal value of h. Thus, we want to find the function $g(x)$ that gives the minimum possible value of $\int g''(x)^2 dx$ given some prior knowledge that $\int g = 1$ and $\int x^2 g = \sigma_f^2$. The true value of σ_f^2 is unknown, but a good, unbiased estimate of it can be obtained from the sample distribution. In other words, the variance of the templates sets the scale of the smoothing parameter. To do this, set up a Lagrangian with 2 multipliers:

$$L(g) = \int_{-\infty}^{\infty} [g''(x)^2 + \lambda_1(g(x) - 1) + \lambda_2(x^2 g - \sigma_f^2)]dx. \qquad (6.4.1)$$

The functional derivative of L in some vanishing direction η (η is some arbitrarily small variation in g) will be needed:

$$D[\eta][L(g)] = \frac{1}{\epsilon}[L(g + \epsilon\eta) - L(g)] \text{ for } \epsilon \to 0. \qquad (6.4.2)$$

For example, $D[\eta](g'')^2 = \frac{1}{\epsilon}[((g + \epsilon\eta)'')^2 - (g'')^2] = 2g''\eta''$. Thus,

$$D[\eta][L] = \int (2g''\eta'' + \lambda_1\eta + \lambda_2 x^2\eta)dx. \qquad (6.4.3)$$

If g is the function that minimzes the Lagrangian, this functional derivative must be zero for any η. Thus we will want to pull out a factor of η in front of every term. To do this, integrate the first term by parts twice:

$$D[\eta][L] = 2(g''\eta' - g'''\eta)|_{-c}^{c} + \int_{-c}^{c} (2g^{iv} + \lambda_1 + \lambda_2 x^2)\eta dx. \qquad (6.4.4)$$

Here, the symmetric boundaries $\pm c$ are introduced as the arbitary (and possibly infinite) limits of integration for the support of the function g, and g^{iv} is the 4th derivative of f. Before continuing, it is useful to step back for a moment to reiterate the plan of action. The function $g(x)$ is going to be the function that will give the minimum value of R for σ_f (which will be estimated from the MC template). The function $\eta(x)$ is some small variation in this function. The functional derivative $D[\eta][L]$ should vanish for all η, and g smoothly goes to 0 at the boundaries, so the first two boundary terms must be 0. Thus, all that is remaining is the requirement that:

$$\int_{-c}^{c} (2g^{iv}(x) + \lambda_1 + \lambda_2 x^2)\eta dx = 0. \tag{6.4.5}$$

This must hold for any η, so the terms in parentheses must vanish. Since as usual things should be symmetric, we can require that g be symmetric. It must have terms $\propto x^6$ and $\propto x^4$ so that g^{iv} can cancel the x^2 terms and x^0 terms. The symmetry rules out x^1, x^3 and x^5 terms, and there can be no terms with powers higher than x^6, since they will not cancel.

Thus, $g = Px^6 + Qx^4 + Rx^2 + S$, leaving 4 unknowns. However, $g(\pm c) = 0$ and $g'(\pm c) = 0$ (since the function should be smooth). These two constraints leave only two unknowns, and $g(x) = A(x-c)^3(x+c)^3$ for $x \in [-c, c]$ and $g(x) = 0$ otherwise. Think of c as a parameter of g; it corresponds to the point along the x-axis where the function goes to zero permanently, or, in other words, the range of support of g. The value of c can be increased, but it is constrained by the normalization and the choice of σ_f, and hence it is a parameter of g.

Given that $g(x) = A(x-c)^3(x+c)^3$ for $|x| < c$, things are almost finished, since there are two additional constraints ($\int_{-c}^{c} g(x)dx = 1$ and $\int_{-c}^{c} x^2 g(x)dx = \sigma_f^2$) and two unknowns ($A$ and c). The first constraint gives $Ac^7 = \frac{-35}{32}$ and the second constraint gives $Ac^9 = \frac{-315}{32}\sigma_f^2$. Combining, $c = -3\sigma_f$ and $A = \frac{35}{69984\sigma_f^7}$, or

$$g = \frac{35}{69984\sigma_f^7}(x - 3\sigma_f)^3(x + 3\sigma_f)^3. \tag{6.4.6}$$

Plugging this into $R(g) = \int_{-3\sigma_f}^{3\sigma_f} g''(x)^2$ gives:

$$h_{OS} = 3\sigma_f[\frac{R(K)}{35n\sigma_k^4}]^{1/5}, \tag{6.4.7}$$

so that

$$h < 3\sigma_f[\frac{R(K)}{35n\sigma_k^4}]^{1/5}. \tag{6.4.8}$$

For the Epanechnikov Kernel, this becomes

$$h < 3\sigma_f[\frac{3}{7n}]^{1/5}. \tag{6.4.9}$$

Typical values of h_{OS} for density estimates in the signal are 11 GeV/c^2 for M_t^{reco} and $3-6$ GeV/c^2 for W_{jj}. For background density estimates, these numbers are larger, as the number of events passing all the cuts is smaller and the templates are wider. Typical values are $15-20$ GeV/c^2 for M_t^{reco} and $5-10$ GeV/c^2 for W_{jj} for background KDE.

6.5 Adaptive kernels

There is no intrinsic reason to choose a constant smoothing factor over the entire template range. Near the core of the distrbution, where poor statistics are not a problem and the leading source of error is the bias, it would be nice to use a smaller value of h. And in the tails of the distribution, where the statistics are small and the variance is large, a larger value of h is better. There are many approaches to such improvements to KDE, one of which is adaptive smoothing, where $h = h(x_i)$, but $h \neq h(x)$. In other words, the smoothing parameter varies for every individual kernel function, but not for the value at which the density is being estimated. It helps to think about this from the point of view of summing many small, normalized bumps centered around the MC points. For example, if n Gaussians centered around the reconstructed top quark masses in MC are added, there is no reason for the widths of all the Gaussians to be the same.

Formally, then, the adaptive Kernel becomes

$$\hat{f}_{adaptive}(x) = \sum_{i=1}^{n} \frac{1}{nh_i} K(\frac{x - x_i}{h_i}), \tag{6.5.1}$$

leaving a choice for how to choose h_i. We follow [37], and seek to choose $h(x_i) \propto f(x_i)^{-0.5}$. As usual, $f(x_i)$ is unknown, but can be estimated from $\hat{f}(x_i)$ by using an initial pilot kernel density estimation. Then $h(x_i) = h_0(\hat{f}(x_i))^{-0.5}$. Here, h_0 cannot be the same size as the normal smoothing parameter unless it is normalized to some average density estimate. Following the suggestion of [32], we choose

$$h(x_i) = h_0 \sqrt{G/\hat{f}(x_i)}, \tag{6.5.2}$$

where G is the geometric mean of all the $\hat{f}(x_i)$. For the adaptive kernel method there are two choices to make for the smoothing parameter - a choice for the first pilot density estimation, and a choice for h_0. It's not clear what type of rule should be used for such kernels, but we take a simple approach and use $h = h_{OS}$ for all pilot density estimates. The claim from [37] is that, under certain restrictions, the bias in the density estmation is now $\propto h^4$, and not h^2.[1] Figures 6.1 and 6.2 show the improvements gained by adaptive KDE. The core of the distributions are typically better matched with a sharper peak, and fluctuations in the tails of the distribution are smoothed out. At this stage in our machinery, one final source of bias is removed - when forming the density estimates, an event is not compared to itself. In other

1. It would be nice to derive this, but [32] claims that showing this "involves extremely tedious algebra" and we have been unable to do so.

words, when finding the density estimate for the ith event, the very same ith event is removed in the sum over the N entries in the MC. This is not a problem when running over the data, but can be a problem when running bias checks in the MC. In particular, some MC samples have more events than other samples such that comparing an event to itself can skew the likelihood values.

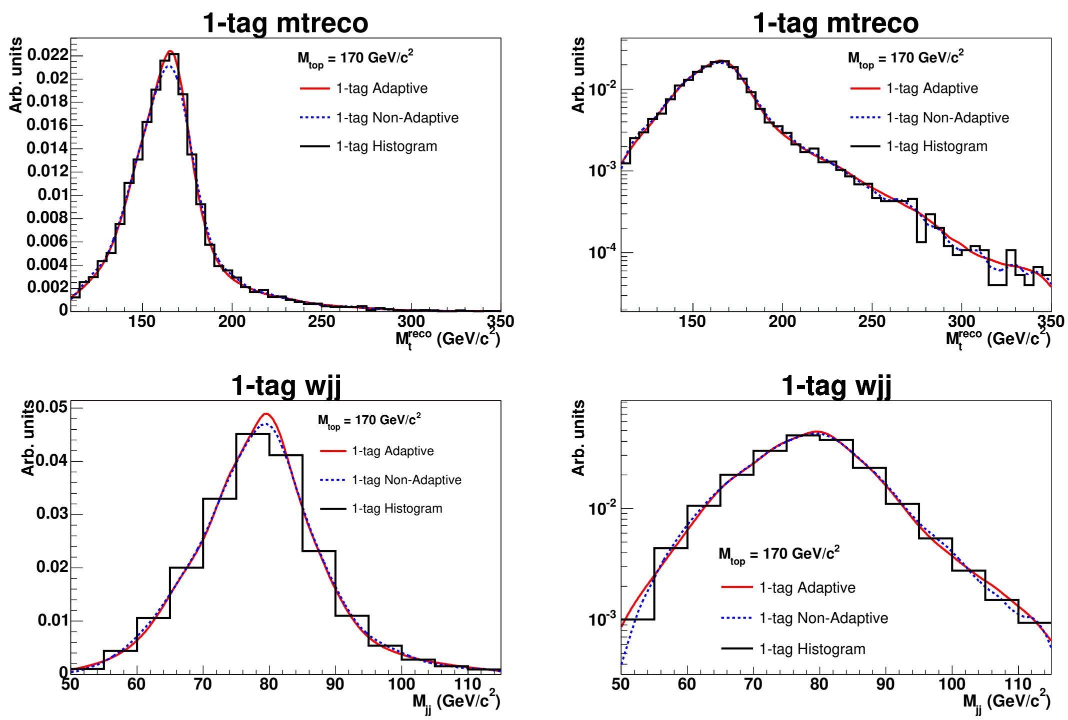

Figure 6.1: One-dimensional templates from KDE for 1-tag events. On the top are M_t^{reco} templates, and on the bottom are W_{jj} templates. The plots on the left are on a linear scale, and the plots on the right are on a log scale. On top of the histograms in solid (dashed) lines are the distributions from adaptive (non-adaptive) KDE.

6.6 KDE in multiple dimensions

The most basic multivariate kernels are product kernels. If $K_1(x)$ and $K_2(y)$ are independent kernels, define a new, single multivariate (2d) kernel by $K_{product}(x, y) \equiv K_1(x) \otimes K_2(y)$, so that

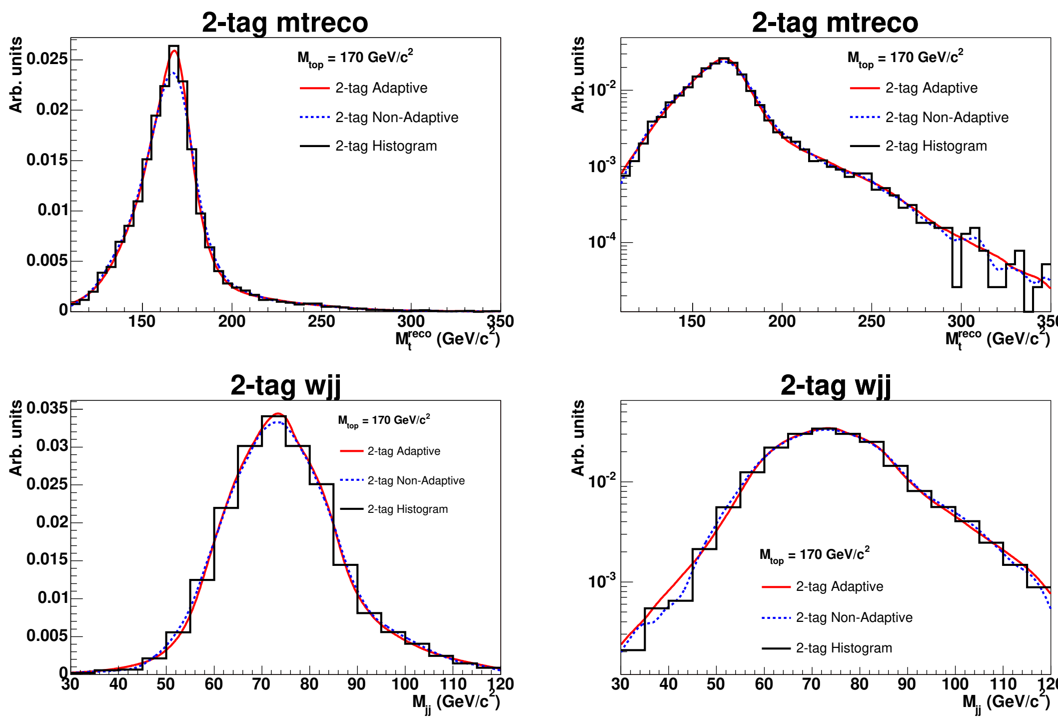

Figure 6.2: One-dimensional templates from KDE for 2-tag events. On the top are M_t^{reco} templates, and on the bottom are W_{jj} templates. The plots on the left are on a linear scale, and the plots on the right are on a log scale. On top of the histograms in solid (dashed) lines are the distributions from adaptive (non-adaptive) KDE.

$$K_{product}(x,y) = \frac{3^2}{4^2 h_1 h_2 n}\sum_{i=1}^{n}[1-(\frac{x-x_i}{h_1})^2][1-(\frac{y-y_i}{h_2})^2]. \tag{6.6.1}$$

Note that the ith event in the MC contributes to $K(x,y)$ if and only if $|x-x_i| < h_1$ and $|y-y_i| < h_2$. In other words, even if only one of the components of the product kernel fails the distance test, the event contributes 0 to the density estimation. Also note that $h_1 \neq h_2$. Requiring the two smoothing parameters to be equal is an unnecessary restriction that makes no sense unless the two variables have the same scale and RMS.

Reference [38] gives an overview of the more general forms of 2-dimensional KDE, and [39] describes the advantages of adaptive methods in multivariate KDE. In the brief math that follows for bivariate KDE, we will assume that the reconstructed top mass and dijet mass have 0 correlation. Of course, this is not true - if they had no correlation, then there would be no reason to move to multivariate density estimates. But since the correlation is small (correlation here refers to a local correlation in the neighborhood under evaluation, and not a global correlation), this will be a good introduction to the mathematical expressions that arise when adding an extra dimension.

6.7 Bivariate math

Start off with the basics, where H now refers to the collective (h_x, h_y):

$$\begin{aligned}<\hat{f}(x,y)> \; &= \; <K_H(x-X, y-Y)> \; = \int f(X,Y)K_H(x-X,y-Y)dXdY = \\ & \frac{1}{h_x h_y}\int f(X,Y)K(t_X,t_Y)dXdY. \end{aligned} \tag{6.7.1}$$

Changing variables gives

$$< \hat{f}(x,y) >= \int f(h_x t_X + X, h_y t_Y + Y) K(t_X, t_Y) dt_X dt_Y. \tag{6.7.2}$$

Recalling the formula for Taylor expansions of functions of 2 variables:

$$\begin{aligned} < \hat{f}(x,y) > &= \int K(t_X, t_Y)[f(X,Y) + f_x(X,Y) h_x t_X + f_y(X,Y) h_y t_Y + \\ & \frac{1}{2} f_{xx}(h_x t_X)^2 + \frac{1}{2} f_{yy}(h_y t_Y)^2 + f_{xy} h_x t_X h_y t_Y] dt_X dt_Y. \end{aligned} \tag{6.7.3}$$

As before, the kernel is normalized and has vanishing first moments in both dimensions, so that:

$$< \hat{f}(x,y) >= f(x,y) + \frac{1}{2} \int K(t_X, t_Y)[f_{xx}(h_x t_X)^2 + f_{yy}(h_y t_Y)^2] dt_X dt_Y. \tag{6.7.4}$$

Thus, the ISB is:

$$\text{ISB} = \frac{1}{4} \sigma_k^4 [h_x^4 R(f_{xx}) + h_y^4 R(f_{yy}) + 2 h_x^2 h_y^2 \int f_{xx} f_{yy}]. \tag{6.7.5}$$

To find the variance of our KDE, write:

$$\text{Var} \hat{f}(x,y) =< K_H(x,y)^2 > - < K_H(x,y >^2, \tag{6.7.6}$$

and then reach into the usual bag of tricks to get

$$
\begin{aligned}
< K_H(x,y)^2 > &= \frac{1}{nh_x h_y} \int f(X + h_x t_X, Y + h_y t_Y) K^2(t_X, t_Y) dt_X dt_Y \\
&\sim \frac{f(x,y) R^2(K)}{nh_x h_y}.
\end{aligned}
\tag{6.7.7}
$$

The extra power of R(K) come from the fact that for the 2d product kernel, $K(x,y) = K(x) \otimes K(y)$ and $\int K^2(x)dx = \int K^2(y)dy = R(K)$. Thus,

$$
\mathrm{IV} = \frac{R(K)^2}{nh_x h_y} - \frac{R(f)}{n}, \tag{6.7.8}
$$

and (dropping the usual term with no h dependence),

$$
\mathrm{AMISE} = \frac{1}{4}\sigma_k^4 [h_x^4 R(f_{xx}) + h_y^4 R(f_{yy}) + 2h_x^2 h_y^2 \int f_{xx} f_{yy}] + \frac{R(K)^2}{nh_x h_y}. \tag{6.7.9}
$$

In general, there is no closed-form expression that minimizes Equation 6.7.9, let alone for the general case of d dimensions. However, note that for $h_x = h_y \equiv h$, even though the ISB $\propto h^4$, the IV $\propto 1/nh^2$ (where it formerly went as only $1/nh$). Because the MC must cover an area in parameter space in the 2d method (as opposed to only a line in parameter space in the 1d method), the number of MC points required to keep the variance at bay increases. Unlike in 1d KDE, where $h_{min} \propto n^{-1/5}$, in 2d KDE $h_{min} \propto n^{-1/6}$. In general, with each added dimension (in other words, with each additional observable), the size of the optimal smoothing parameter increases (some authors refer to this exponential increase in the necessary statistics as the curse of dimensionality). In d dimensions, $h_{min} \propto n^{-1/(4+d)}$ and the AMISE converges at a rate of $n^{-4/(4+d)}$.

When moving to full two-dimensional KDE, it is necessary once again to choose smoothing parameters. Reference [36] provides an oversmoothing rule for multiple dimensions that is also mentioned in [31]. The variance still sets the scale in each dimension, as it must from dimensional analysis. In any dimension x,[2]

$$h_x^{OS} = \left[\frac{R(K)^d d}{nC_f \sigma_K^4}\right]^{1/(d+4)} \sigma_x, \tag{6.7.10}$$

$$C_f = \frac{16\Gamma(\frac{d+8}{2}) d(d+2)}{(d+8)^{(d+6)/2} \pi^{d/2}}, \tag{6.7.11}$$

where, for us, $d = 2$.

The values of h_{OS} for 2d density estimates are typically $\sim 20\%$ wider than the 1d versions.

6.8 Boundary kernels

The KDE described above knows nothing about boundaries or kinematic cut-offs in the observable quantities. Mathematically, the density functions for $\mathrm{W_{jj}}$ and $\mathrm{M}_t^{\mathrm{reco}}$ can take on any real values, given large enough smoothing. In reality, there are soft limits to the values that can be possibly be observed in data. The value of $\mathrm{M}_t^{\mathrm{reco}}$ cannot shift much below the sum of the W mass and an additional hard jet assigned to a b-jet. On the high side, $\mathrm{M}_t^{\mathrm{reco}}$ is bounded by the 1.96 TeV center-of-mass energy, and in practice, the parton distribution functions fall off so rapidly that we

2. Be careful - both [36] and [31] have only a single power of R(K) in their OS rule, not $\mathrm{R(K)}^d$. This all depends on how R(K) is defined - we define it here to be based on the 1d Kernel that we will multiply together d times to get to the product kernel. But we could also define R(K) (as the references do) to be an integral over all d dimensions, which for the product kernel automatically gives us d powers of the number.

don't expect to see events with $\mathrm{M}_t^{\mathrm{reco}}$ approaching 1 TeV. Additionally, the W mass has a low-side boundary at roughly the invariant mass of two jets passing kinematic cuts separated by $\Delta\mathrm{R} = 0.4$, and a high-side boundary that is set by the W mass constraint and the requirement that $\chi^2 < 9.0$.

Adaptive kernels make the above problem even worse - in the tails of the distribution, events have very large amount of smoothing that can extend far past any reasonable soft boundary. To solve the above problems, we force the KDE to fall within some pre-determined regions and ensure that the density estimates are normalized to unity within the boundaries. From Equation 6.2.8, note that the kernel normally integrates to unity and the first term simply becomes $f(x)$, ensuring that the density estimates are unbiased to 0th order. The second term is also usually zero because the kernel is chosen to be a symmetric function, with a vanishing first moment. Thus, the normal bias in KDE is given by the third term, and is $\propto h^2$.

When the probability density extends beyond a soft boundary where the data are not found, the above statements about the normalization and first moment of the kernel do not hold. Ideally, a fix would enforce both of these conditions everywhere, ensuring that the bias remains of order h^2. In practice, this is somewhat difficult to do, and near the boundary we may need to be satisfied with enforcing only the normalization condition. When density estimates are calculated, we check that the individual kernels are normalized to 1.0 within the boundary. If they are not, their normalization N (which will be less than 1.0) is analytically calculated, and all density estimate contributions are divided by this factor to ensure unit normalization.

More formally,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) \rightarrow \frac{1}{n}\sum_{i=1}^{n} \frac{K_h(x - x_i)}{N_i}, \tag{6.8.1}$$

where, for lower boundary a and upper boundary b,

$$N_i = \int_a^b K_h(x - x_i)dx. \tag{6.8.2}$$

Using the Epanechnikov kernel and adaptive density estimates, the normalization on the ith kernel becomes:

$$N_i = \frac{3}{4}\left[\left(\frac{b_i - x_i}{h_i}\right) - \left(\frac{a_i - x_i}{h_i}\right) - \frac{1}{3}\left(\frac{b_i - x_i}{h_i}\right)^3 + \frac{1}{3}\left(\frac{a_i - x_i}{h_i}\right)^3\right], \tag{6.8.3}$$

where, for lower boundary a_0 and upper boundary b_0,

$$\frac{a_i - x_i}{h_i} = \max\left[\left(\frac{a_0 - x_i}{h_i}\right), -1\right], \tag{6.8.4}$$

$$\frac{b_i - x_i}{h_i} = \min\left[\left(\frac{b_0 - x_i}{h_i}\right), 1\right]. \tag{6.8.5}$$

The values of a_i and b_i have this min/max form because the Epanechnikov kernel has support only when $|t| \leq 1$. Renormalization of the 2d density estimates follows from the 1d renormalization, as the multi-dimensional kernel used in this analysis is a product kernel with two integrals in the normalization that factorize cleanly: $K_{2d}(x, y) \sim K_{h1}(x) \cdot K_{h2}(y)$ so that $N_{2d}(x, y) = N_{h1}(x) \cdot N_{h2}(y)$. The choice of boundaries is arbitrary; we choose to make very loose cuts to maximize the number of $t\bar{t}$ events in the sample. For both 1-tag and 2-tag events we require:

$$110 < M_t^{\mathrm{reco}} < 350 \text{ (GeV/c}^2\text{)}. \tag{6.8.6}$$

Table 6.1: Efficiencies to pass the boundary cuts.

	1-tag	2-tag
Pre-χ^2 ($\mathrm{M_{top}}$ = 170 GeV/c^2, Δ_{JES} = 0.0 σ_{c})	0.865	0.650
Post-χ^2 ($\mathrm{M_{top}}$ = 170 GeV/c^2, Δ_{JES} = 0.0 σ_{c})	0.989	0.981
Pre-χ^2 ($\mathrm{M_{top}}$ = 170 GeV/c^2, Δ_{JES} = 3.0 σ_{c})	0.861	0.651
Post-χ^2 ($\mathrm{M_{top}}$ = 170 GeV/c^2, Δ_{JES} = 3.0 σ_{c})	0.989	0.981
Pre-χ^2 ($\mathrm{M_{top}}$ = 170 GeV/c^2, Δ_{JES} = -3.0 σ_{c})	0.846	0.601
Post-χ^2 ($\mathrm{M_{top}}$ = 170 GeV/c^2, Δ_{JES} = -3.0 σ_{c})	0.987	0.977
Pre-χ^2 ($\mathrm{M_{top}}$ = 150 GeV/c^2, Δ_{JES} = 3.0 σ_{c})	0.880	0.671
Post-χ^2 ($\mathrm{M_{top}}$ = 150 GeV/c^2, Δ_{JES} = 3.0 σ_{c})	0.988	0.980
Pre-χ^2 ($\mathrm{M_{top}}$ = 150 GeV/c^2, Δ_{JES} = -3.0 σ_{c})	0.860	0.618
Post-χ^2 ($\mathrm{M_{top}}$ = 150 GeV/c^2, Δ_{JES} = -3.0 σ_{c})	0.980	0.975
Pre-χ^2 ($\mathrm{M_{top}}$ = 200 GeV/c^2, Δ_{JES} = 3.0 σ_{c})	0.818	0.602
Post-χ^2 ($\mathrm{M_{top}}$ = 200 GeV/c^2, Δ_{JES} = 3.0 σ_{c})	0.987	0.975
Pre-χ^2 ($\mathrm{M_{top}}$ = 200 GeV/c^2, Δ_{JES} = -3.0 σ_{c})	0.814	0.562
Post-χ^2 ($\mathrm{M_{top}}$ = 200 GeV/c^2, Δ_{JES} = -3.0 σ_{c})	0.986	0.979

For W_{jj}, we make different cuts for the two samples. The 2-tag sample has a looser cut, as it does not have the same sculpting as the 1-tag sample:

$$
\begin{aligned}
&50 < W_{\mathrm{jj}} < 115 \ (\mathrm{GeV/c^2}) \text{ for 1-tag,} \\
&50 < W_{\mathrm{jj}} < 125 \ (\mathrm{GeV/c^2}) \text{ for 2-tag.}
\end{aligned}
\tag{6.8.7}
$$

The boundary cuts are not very efficient as a first cut, but for those events that pass the χ^2 requirement, the efficiency is high, even across Δ_{JES} and $\mathrm{M_{top}}$. Table 6.1 summarizes the boundary cut efficiencies for signal events.

6.9 Clipped adaptive kernels

In the adaptive method, the smoothing parameter for the ith MC point, $h_i \propto \hat{f}(x_i)^{-0.5}$. In other words, the width of the smoothing goes as the inverse square root of the pilot density estimate. This improves the density estimates, but can lead to strange results. If the shape of the pilot density estimate changes faster than $\sqrt{x}$, the density estimates become nonlocal. Points on the tail of the distribution can contribute probability weight such that a density estimate at point x is influenced by two types of entries - MC points in the neighborhood of x (these are the points that we want to contribute), as well as MC points only very far from x (which should not contribute at all). To partially (though not entirely) remedy this problem, we choose to follow a recommended procedure in one of the original adaptive KDE papers [32] and "clip" the pilot density estimate:

$$\hat{f}(x_i) \to \max(\hat{f}(x_i), 0.1 \cdot \hat{f}(0)), \tag{6.9.1}$$

where $\hat{f}(0)$ is the maximum pilot density estimate obtained from all the MC entries. This has the effect of forcing a maximum value of h to be $\sqrt{10}$ times the value of the smallest h (which occurs in the peak of the distribution). The tails of our distribution are then not quite as smooth and we have unfortunately added another arbitrary parameter to our method, but it makes our machinery more robust and the density estimates less nonlocal.

6.10 Signal Density Estimates

Figures 6.3 and 6.4 show 2d density estimates obtained from $t\bar{t}$ MC with $\mathrm{M_{top}} = 170\ \mathrm{GeV}/c^2$ and the nominal values of Δ_{JES}. All the above effects, including adaptive

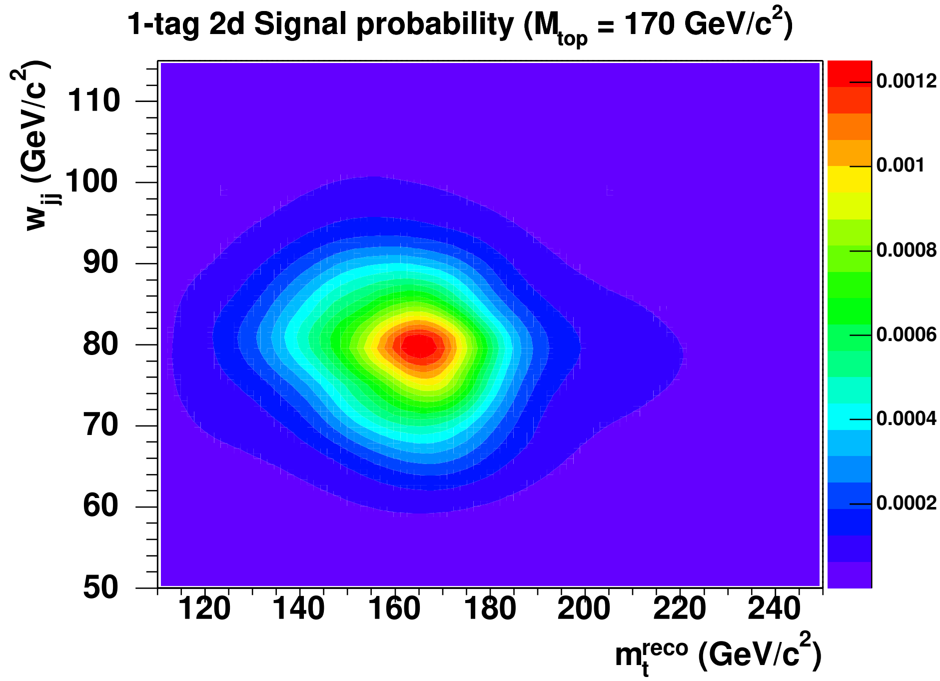

Figure 6.3: Two-dimensional PDF used for 1-tag events at $\mathrm{M_{top}} = 170$ GeV/c^2 and $\Delta_{\mathrm{JES}} = 0.0\sigma_{\mathrm{c}}$.

density estimates, clipped kernels and boundary conditions, are applied.

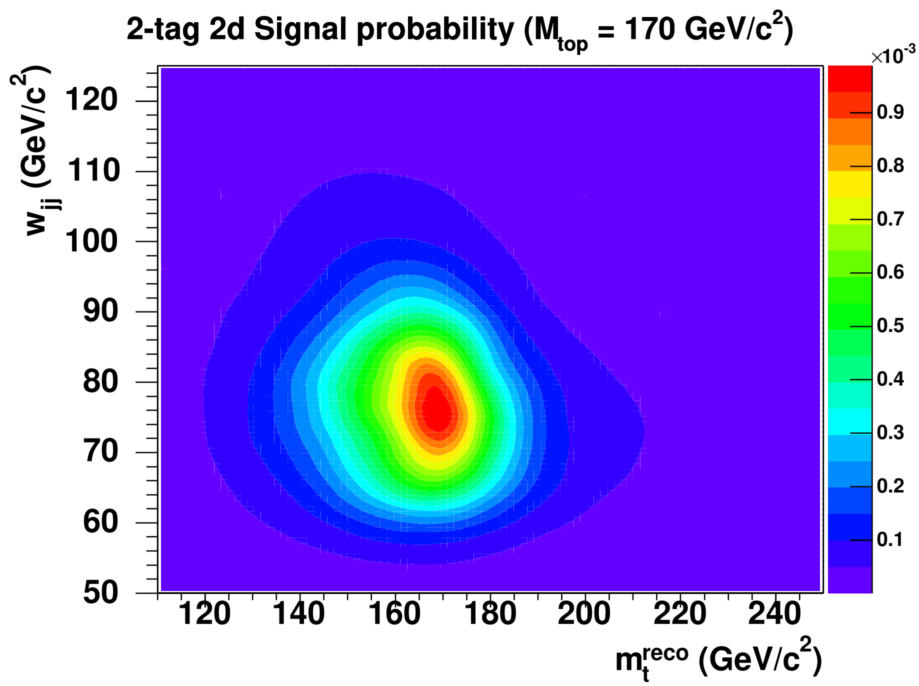

Figure 6.4: Two-dimensional PDF used for 2-tag events at $\mathrm{M_{top}} = 170$ GeV/c^2 and $\Delta_{\mathrm{JES}} = 0.0\sigma_{\mathrm{c}}$.

CHAPTER 7

BACKGROUNDS

Any measurement of the top quark mass needs to model the non-negligible background events that mimic the signature of $t\bar{t}$ events. Requiring the presence of at least one b-tag in the event significantly reduces, but does not eliminate, the background. In this analysis, templates for each of the individual backgrounds are combined with the appropriate relative weights to make single templates for both M_t^{reco} and W_{jj}. Estimates for the expected number of background events in the two subsamples are used as Gaussian constraints in the likelihood, improving on the statistical power of the method. The background estimation is based on CDF's Method II background estimation machinery.

7.1 Method II overview

Constraining or measuring backgrounds using data is typically preferred to using Monte Carlo to model the background. Unfortunately, data-based techniques are not entirely possible for all the backgrounds to Lepton+Jets $t\bar{t}$ events. Method II is a hybrid of data and Monte Carlo measurements - data are used whenever possible, and Monte Carlo-based techniques are applied to fill in any remaining gaps in knowledge. In particular, the overall rate of events with real W bosons and additional jets (W+jets), which dominate the background sample, is found in the data. Such a data-based measurement is important, as many higher-order diagrams contribute to this background, and the Monte Carlo is not expected to accurately reproduce the rate for such events. On the other hand, the fractions of these events with one or two charm quarks (Wc and W$c\bar{c}$ events) or two bottom quarks (W$b\bar{b}$ events) are taken from Monte Carlo.[1] The assumption is that many of the uncertainties on the rate of W+jets events are common to all types of W+jets events, and cancel when calculating ratios. Thus, overall normalizations of W+jets events come from the data after subtracting off an estimate for the fraction of W events that do not contain a real W boson but come from QCD processes (non-W), and separating out an estimate for other types of backgrounds (WW/WZ/ZZ, $t\bar{t}$ and single-top production).[2] The rate for tagging these events is then estimated from a combination of Monte Carlo- as well as data-based measurements.

The expected number of W pretag events can be written as:

$$N_{\text{pre}}^{W} = N_{\text{pre}}^{data} \cdot (1 - F_{\text{non-W}}) - N_{\text{pre}}^{t\bar{t}} - N_{\text{pre}}^{\text{diboson}} - N_{\text{pre}}^{\text{single-top}}. \tag{7.1.1}$$

The number of W+jets events with heavy flavor in the sample is then given by

$$N_{\text{HF}}^{W} = N_{\text{pre}}^{W} \cdot F_{\text{HF}} \cdot \epsilon_{\text{HF}}, \tag{7.1.2}$$

where F_{HF} is the fraction of W+jets events with heavy flavor quarks, and ϵ_{HF} is the efficiency for such an event to pass event selection.

1. At typical momentum fractions of interest, the proton and anti-proton have non-negligible charm parton distribution functions, making Wc a real background that must be taken into consideration. The bottom parton distribution function is almost zero, so that events with a real W and a bottom quark are only found when a gluon splits into a $b\bar{b}$ pair, and not as W+b.

2. These samples contain events with a real W, but need to be sutracted off since they include other diagrams that would not accurately be accounted for in the ratio.

7.2 Tagging efficiency

Any estimate of the number of tagged background events requires a measurement of the tagging efficiency in the data. The Monte Carlo is not necessarily trusted to fully reproduce tracking efficiencies and resolutions in the tails of distributions, so a data-based measurement of the efficiency to tag heavy flavor jets is preferred. The difference between the data and Monte Carlo-based tagging efficiencies is denoted by the b-tag Scale Factor (SF), which is the tagging efficiency in the data over the tagging efficiency in the MC. The SF is measured in two independent samples. The first sample consists of dijet events in which one of the jets contains an electron, often signaling the semi-leptonic decay of a heavy flavor hadron. The sample is split in two depending on whether the electron has a partner consistent with a photon conversion to an e^+e^- pair; events in which the electron has a conversion partner are dominated by light flavor, and events in which the electron does not have a partner are dominated by heavy flavor from gluon splitting. The difference between the two results in a measurement of the tagging efficiency, and thus of the SF. The second measurement of the SF comes from dijet events in which one of the jets contains a muon. The distributions of the muon p_T relative to the jet axis are fit to templates derived from heavy flavor and light flavor MC; the muons from heavy flavor decays are significantly harder in this variable. The fraction of the events containing real heavy flavor, as derived from the fits, gives an estimate for the SF.

The two methods are combined to measure $\mathrm{SF} = 0.95 \pm 0.01$ (stat.) ± 0.05 (syst.), giving a average per-jet bottom tag efficiency of $40 \pm 2\%$ and a charm tag efficiency of $8 \pm 1\%$ in $t\bar{t}$ events. The charm SF is not measured, but is taken as the same value as the SF for b-jets, with the uncertainty tripled to be conservative.

7.3 Mistag rates

The number of background events passing event selection cuts also depends on the rate to mistag light-quark jets. These fake tags come from a variety of sources: interactions with material in the detector, mismeasured tracks, or the real decays of long-lived light-flavor particles such as K_s and Λ. The rate for mistagging jets is encapsulated in a mistag matrix, which gives the probably to improperly tag a light flavor jet given the jet energy and η_{det}, the number of tracks in the jet, the number of primary vertices found in the event, the event's primary vertex position, and the sum of the transverse energy in the event. The matrix is derived from the number of negative tags in generic jet trigger samples. The negative tag rate, which is the rate of tags found behind the interaction point (in the wrong direction relative to that for a heavy flavor decay), is a good first-order estimate of the mistag rate due to mismeasured tracks. This rate is then modified to account for real heavy flavor quarks in the generic jet trigger sample and the possibility of mistagging a jet due to material interactions or real long-lived light flavor decays. The measured mass of the tracks associated with the secondary vertex in positively tagged and negatively tagged jets with the mistag matrix applied is fit to MC-derived templates for charm, bottom and light flavor quarks. The full correction to the mistag rate depends on jet E_T and varies from 30 to over 50 %. The combined probability to mistag any light-flavor jet in a $t\bar{t}$ event is $1.0 \pm 0.2\%$.

7.4 QCD events

The requirement of an energetic charged, isolated lepton and missing energy in the detector significantly enchances the number of events with real W bosons in the event sample. Purely QCD events with no real W's can still pass the event selection, however. Energetic charged leptons from W's can be faked by photon conversions (electrons) or misidentified pions and kaons (muons), as well as from semi-leptonc heavy-flavor decays. In such events, the $\not{E}_T$ requirement can also be passed when jets are mismeasured or partially fall into cracks. On the other hand, non-W events typical do fail the $\not{E}_T$ requirement, so the $\not{E}_T$ distribution can be used to fit for the fraction of events with fake W bosons. Templates for $t\bar{t}$ from MC, constrained to the expected cross section, are subtracted from the distribution. The remaining data are then fit to templates derived from W+jets MC, dijet events with leptons (dominated by $b\bar{b}$), and a sample to mimic the $\not{E}_T$ signature of QCD events. The last template is filled with "antielectron" events, in which a high p_T electron candidate fails two of the tight electron geometric selection requirements. The fits give the non-W fraction, F_{non-W}, with uncertainties that are increased by a factor reflecting the fit χ^2 probability and by a systematic to account for the contamination of real W electrons in the non-W sample.

7.5 Monte Carlo backgrounds

Backgrounds with Monte Carlo-based estimates are more straightforward to calculate. For each background, the expected number of events is:

$$N_{\mathrm{pre}}^{MC} = \sigma^{MC} \cdot \int \mathcal{L} dt \cdot \epsilon^{MC}, \tag{7.5.1}$$

as the number of events expected from a background is given by the theoretical cross section multiplied by the integrated luminosity for the measurement and the efficiency for an event to pass the event selection. The number of tagged events from one of these backgrounds is N_{pre}^{MC} multiplied by the tag efficiency for real b quarks, and the mistag probability for light quarks.

7.6 Preliminary Background Model

The CDF top group provides numbers for the above background model. W+jets MC of various flavors is modeled by ALPGEN version 2.10 [40], and showered with PYTHIA version 6.325 [28]. The $t\bar{t}$ selection requires 4 or more jets, so the MC is divided into exclusive n-parton samples with less than 4 partons, and inclusive samples with 4 or more partons. A matching scheme between quarks at the matrix element and hadrons after fragmentation ensures no double-counting of phase space across samples, as it is otherwise possible for events with hard hadronic shower evolution to give states already described by events at the matrix element level [41, 42].

Similarly, events with heavy flavor after fragmentation are checked to ensure no double-counting of phase space across samples with different flavor types at the matrix element level. This can occur if samples generated at the matrix element level with light-flavor partons produce charm or bottom quarks during fragmenation, or if events generated at the matrix element level with charm quarks produce bottom quarks during fragmentation.

W+$b\bar{b}$ + n-partons samples are produced for n = 0,1 and 2 partons; the 2-parton sample is an inclusive 2 or more parton sample, and the other samples are exclusive. Similiar n-parton samples are produced for W$c\bar{c}$ events. W+c+n-parton MC is exclusive for n=1,2, and inclusive for n=3 or more. W+n-light partons is exclusive for n=2, 3, and inclusive for n=4 or more. Single-top MC is separated for s-channel and t-channel production. Diboson MC samples include WW, WZ and ZZ production, and the QCD background shape is modeled by events in the data passing all cuts except the isolation cut on the charged lepton; an anti-isolation cut is imposed, requiring more than 20% of the energy in a cone of $\Delta R = 0.4$ around the lepton not to be associated with the object.

Figures 7.1 and 7.2 show the 1d $\mathrm{M}_t^{\mathrm{reco}}$ density estimates for the main individual background components at the nominal Δ_{JES}, weighted by the expected contributions to the total background. Figures 7.3 and 7.4 show the 1d W_{jj} distributions. These distributions are not explicitly used in the analysis, but show that the different backgrounds largely have similar shapes and peak at lower values of M_t^{reco} than the signal.

7.7 Final background model

The background model provided by the top group at CDF comes close to the final event selection for this analysis, but it is not identical. The number of background events and expected fractions from each sample are scaled to account for several effects: the $\chi^2 < 9.0$ cut on the best solution from the kinematic fitter, the boundary cuts on both M_t^{reco} and W_{jj}, and the fact that this analysis allows for tags on only the leading 4 jets (this last effect matters only for the 2-tag sample, since the 1-tag sample requires exactly 4 tight jets).

Table 7.1 shows the expected number of background events in the 1-tag and 2-tag samples after both the χ^2 cut and the boundary cut, as well as the expected number of signal events based on the theoretical cross-section (6.7 pb) at $\mathrm{M}_{\mathrm{top}} =$ 175 GeV/c^2. The combined background template uses the backgrounds fractions after the χ^2 cut, but as described in Chapter 9, to accurately model the background in MC psuedoexperiments, the expectation before the χ^2 cut is needed. Uncertainties on these numbers are derived by taking the uncertainties given by the CDF top group and scaling for the appropriate differences as above. Uncertainties across jet bins and between electron and muon backgrounds are taken as fully correlated. To be conservative, the uncertainties on the W+jets backgrounds are assumed to have 100%

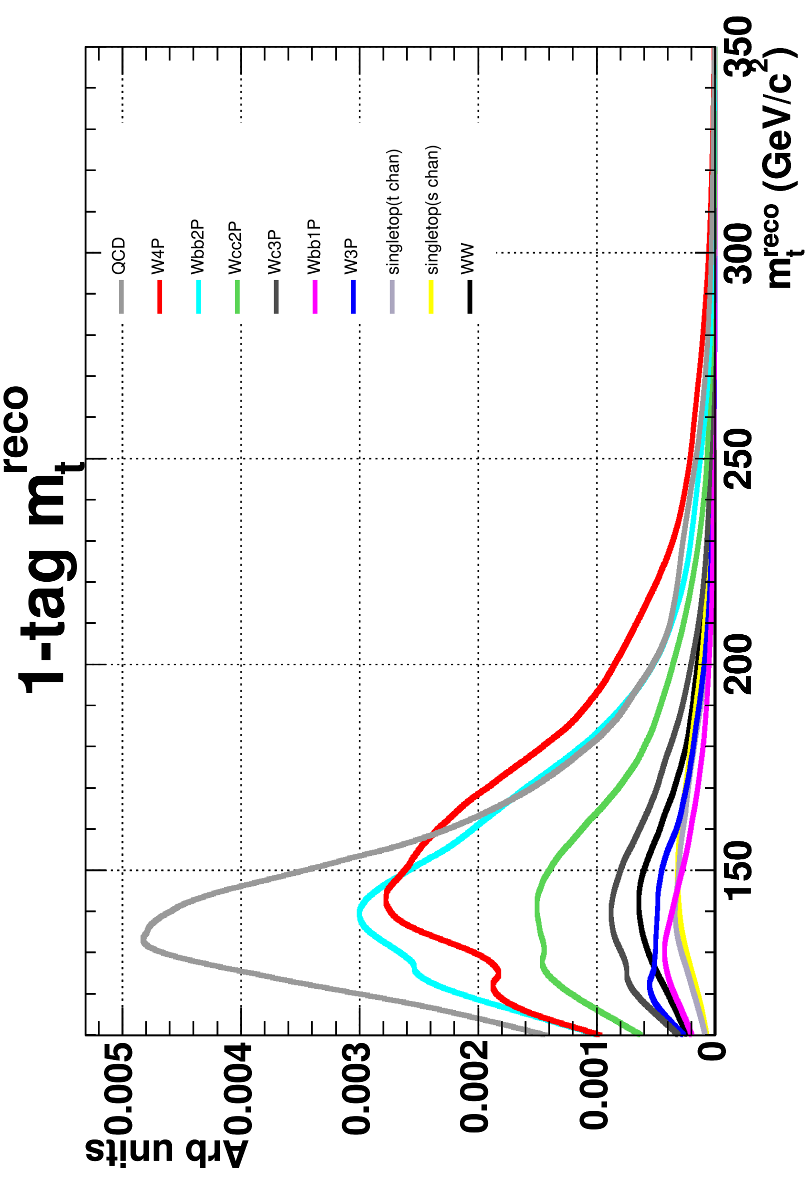

Figure 7.1: 1-tag 1d M_t^{reco} PDFs separated for individual backgrounds. The nominal Δ_{JES} was used.

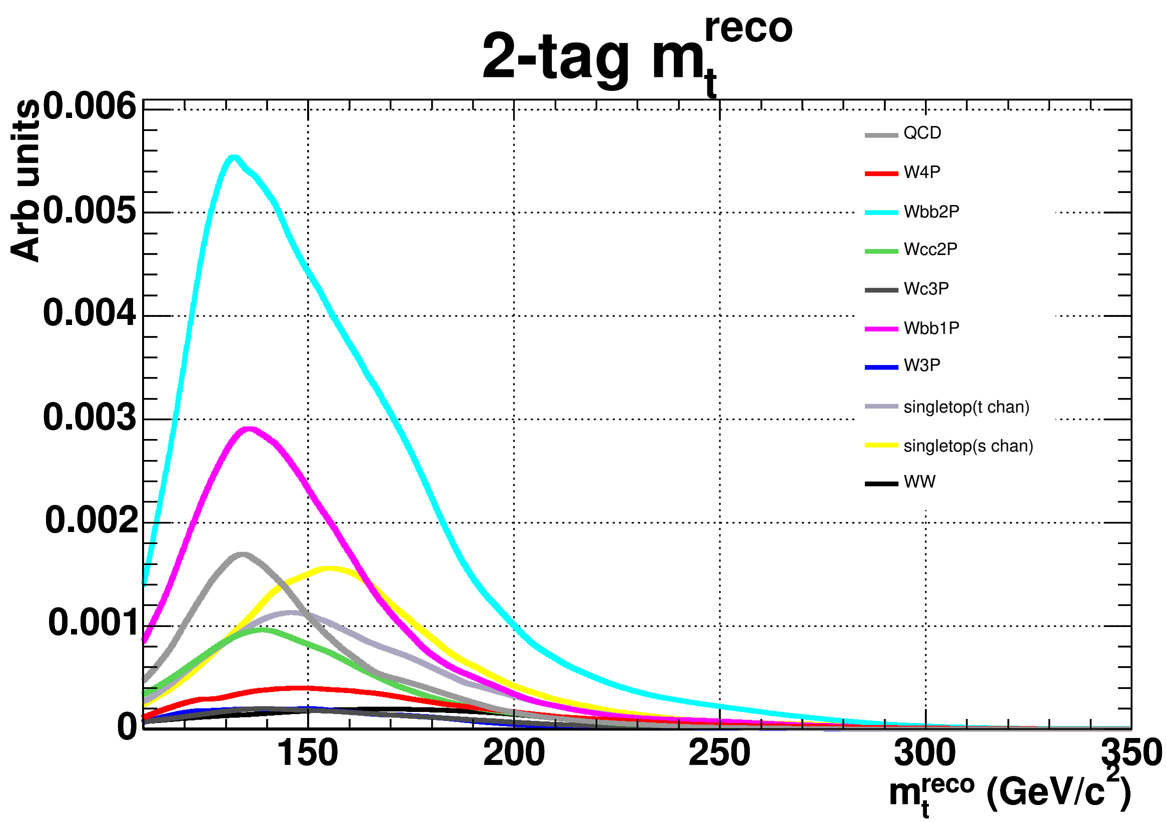

Figure 7.2: 2-tag 1d M_t^{reco} PDFs separated for individual backgrounds. The nominal ΔJES was used.

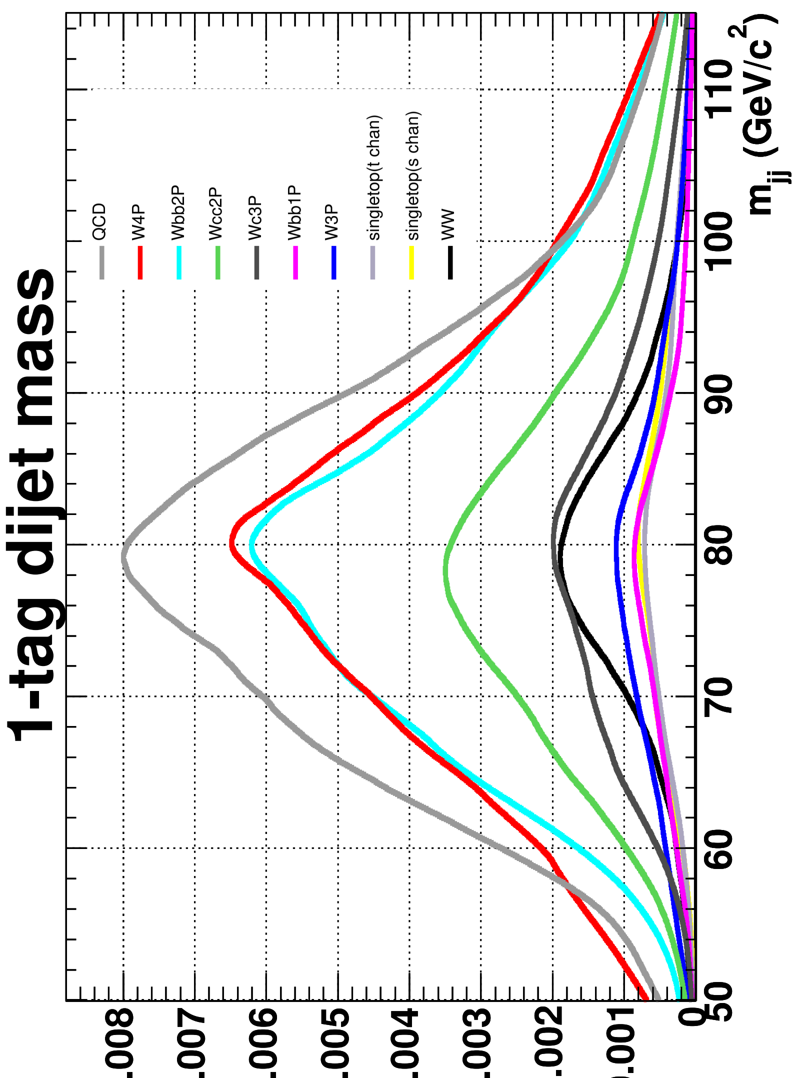

Figure 7.3: 1-tag 1d W_{jj} PDFs separated for individual backgrounds. The nominal Δ_{JES} was used.

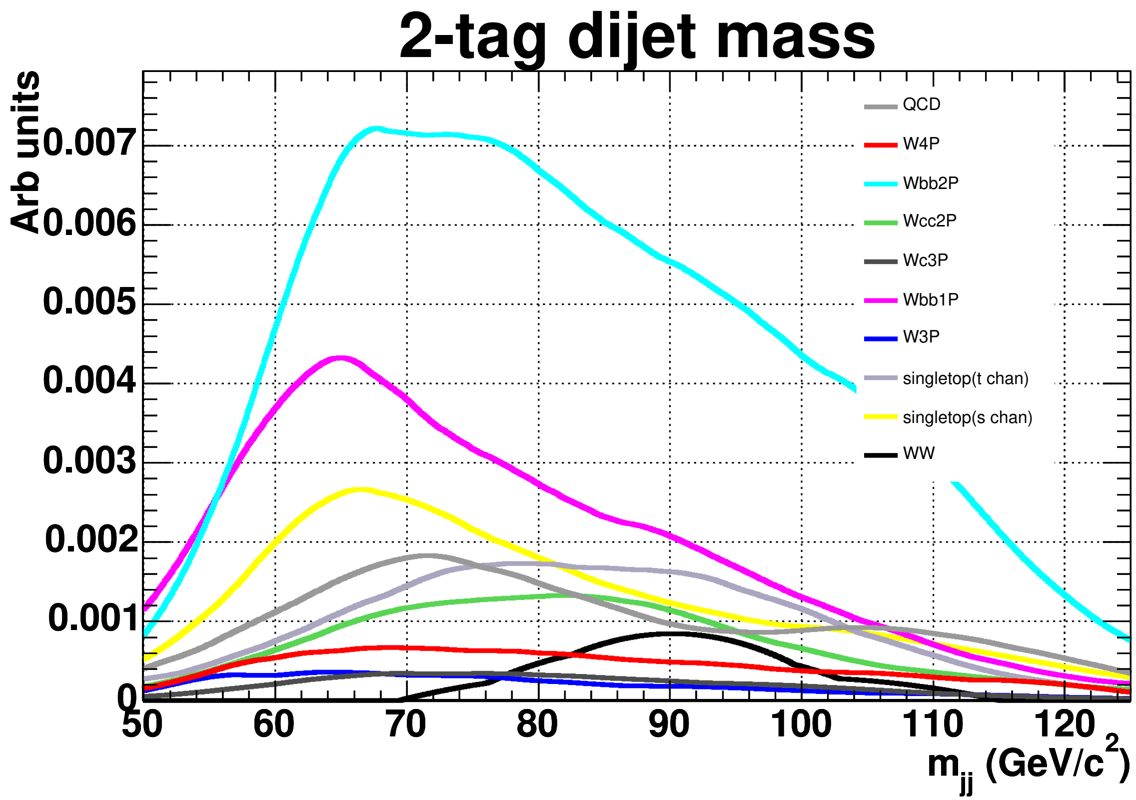

Figure 7.4: 2-tag 1d W_{jj} PDFs separated for individual backgrounds. The nominal ΔJES was used.

Table 7.1: Expected number of background events. Both sets of numbers are after the boundary cuts and include corrections for tagging only on the leading jets.

	After χ^2 cut		Before χ^2 cut	
	1-tag	2-tag	1-tag	2-tag
W$b\bar{b}$+ $\geq$2P	7.89	1.41	9.76	2.05
W$b\bar{b}$+1P	0.96	0.65	1.17	0.90
W$b\bar{b}$+0P	0.23	0.06	0.30	0.11
W$c\bar{c}$+ $\geq$2P	4.37	0.23	5.34	0.33
W$c\bar{c}$+1P	0.5	0.05	0.61	0.08
W$c\bar{c}$+0P	0.08	0.16	0.08	0.16
Wc+ $\geq$3P	2.49	0.06	3.10	0.10
Wc+2P	0.63	0.01	0.79	0.02
Wc+1P	0.21	0.07	0.23	0.07
W+ $\geq$4P	8.60	0.14	10.73	0.22
W+3P	1.43	0.06	1.73	0.09
W+2P	0.41	0.01	0.47	0.01
s-channel single top	1.03	0.42	1.26	0.61
t-channel single top	0.95	0.32	1.22	0.42
WW	1.83	0.08	2.14	0.19
WZ	0.56	0.12	0.66	0.17
ZZ	0.05	0.01	0.06	0.02
QCD	10.44	0.33	11.66	0.43
Total Background	42.7 $\pm$12.5	4.2 $\pm$1.9	51.3 $\pm$15.0	6.0 $\pm$2.7
$t\bar{t}$ (6.7 pb)	156.7	76.6	173.7	93.0

correlation. The 6% luminosity uncertainties are taken as fully correlated across MC backgrounds; the remaining uncertainties on the MC-based backgrounds are assumed to have no correlation. Any other uncertainties are added in quadrature.

7.8 Background templates

Figures 7.5 and 7.6 show 2d density estimates for the combined background at $\Delta_{\mathrm{JES}} = 0.0\ \sigma_{\mathrm{c}}$. All KDE procedures, including adaptive density estimates, clipped kernels and boundary conditions, are applied. Density estimates are derived independently for each of the backgrounds (such that the smoothing factors differ between

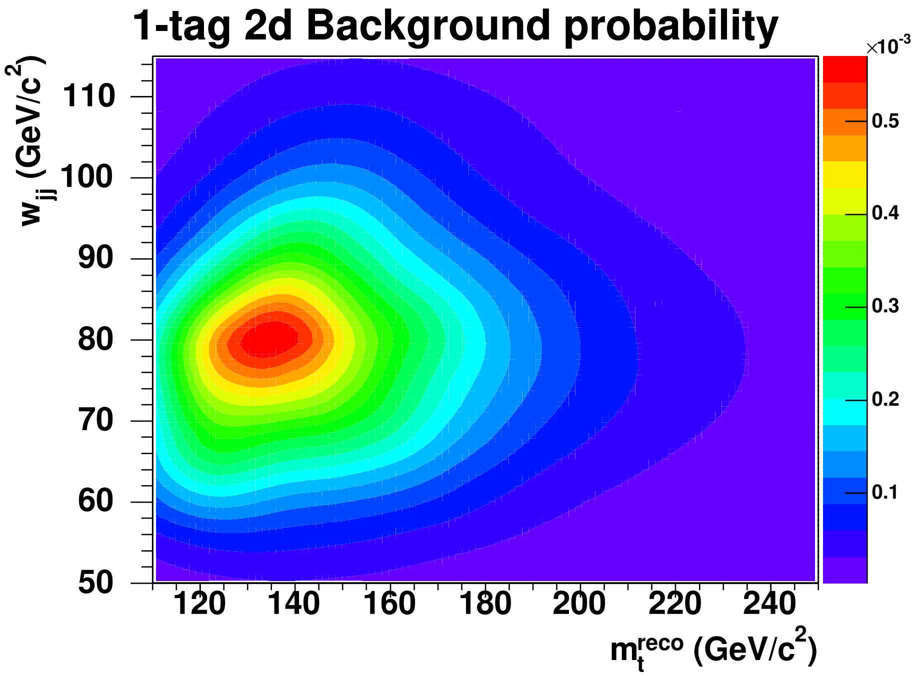

Figure 7.5: Two-dimensional PDF used for 1-tag background events at $\Delta_{\rm JES} = 0.0$ $\sigma_{\rm C}$.

backgrounds), and are combined with the appropriate normalizations. All events in a particular tag subsample and background type have the same weight; the only exception is for mistag events, which are weighted not only by the overall mistag normalization but also by the relative event-by-event mistag probabilities. The values of $\Delta_{\rm JES}$ are allowed to vary for all backgrounds except for the QCD model, which is based on data and thus has no $\Delta_{\rm JES}$ dependence. The relative normalizations of the different backgrounds do not vary with $\Delta_{\rm JES}$, as a very conservative (but small) background composition systematic is taken, as described in Chapter 10.

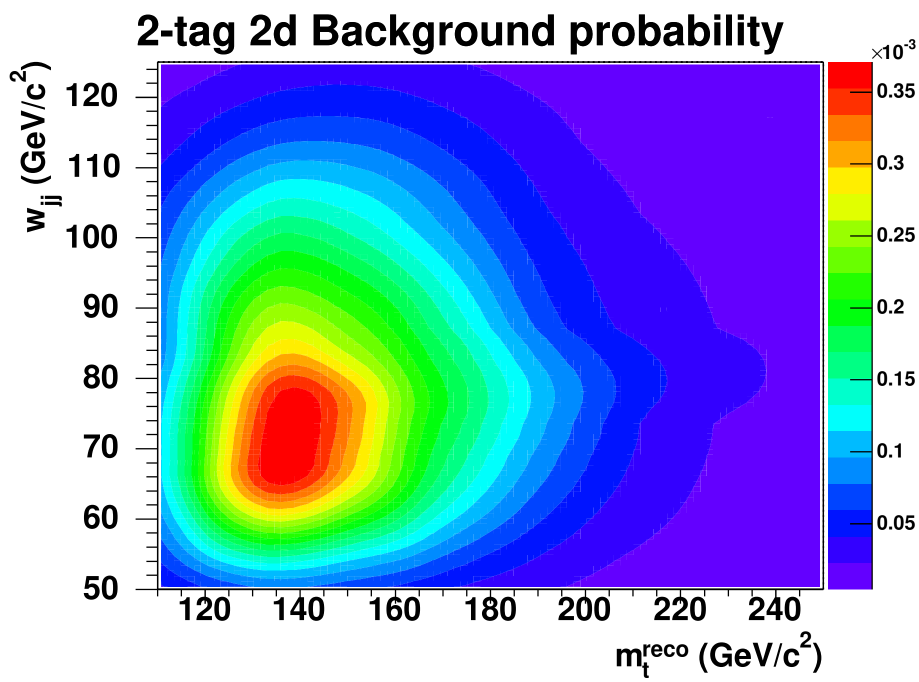

Figure 7.6: Two-dimensional PDF used for 2-tag background events at $\Delta_{\mathrm{JES}} = 0.0$ σ_{C}.

CHAPTER 8

LIKELIHOOD

Kernel Density Estimation provides the probability of observing values of M_t^{reco} and W_{jj} for a given value of $\mathrm{M_{top}}$ and Δ_{JES}, but we require the likelihood of observing an ensemble of M_t^{reco} and W_{jj}. The derivations that follow are largely taken from [43].

8.1 Definitions

Before beginning, it is convenient to introduce some notation:

$$\begin{aligned} n_s &\equiv \text{Expected number of signal events in the data,} \\ n_b &\equiv \text{Expected number of background events in the data,} \end{aligned} \tag{8.1.1}$$

where these numbers are expectation values given by the unknown, true cross sections, acceptances and luminosity. Also of interest are the unknown, actual number of signal and events in the data:

$$N_s \equiv \text{True number of signal events in the data,} \tag{8.1.2}$$

$$N_b \equiv \text{True number of background events in the data,} \tag{8.1.3}$$

$$N \equiv N_s + N_b \equiv \text{Total number of events in the data.} \tag{8.1.4}$$

Ultimately we are interested in the probability of observing the data. Refer to the ensemble of data (all the values of M_t^{reco} and W_{jj}) for events passing all cuts as D. Then the quantity of interest is the probability of observing D given some possible values of $\mathrm{M_{top}}$, Δ_{JES}, n_s and n_b, $P(D|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}})$.

8.2 Extended Maximum Likelihood Derivation

Start off by summing over all possible values of N_s (there is no additional sum over N_b because N_b is uniquely defined for a given number of events N in D and N_s):

$$P(D|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) = \sum_{N_s=0}^{N_s=N} P(D, N_s, N_b|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}), \tag{8.2.1}$$

$$\begin{aligned} &\sum_{N_s=0}^{N_s=N} P(D, N_s, N_b|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) = \\ &\sum_{N_s=0}^{N_s=N} P(D|n_s, n_b, N_s, N_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) P(N_s, N_b|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}). \end{aligned} \tag{8.2.2}$$

Given n_s and n_b (which depend on $\mathrm{M_{top}}$ and Δ_{JES}), the probability of observing N_s and N_b is independent of $\mathrm{M_{top}}$ and Δ_{JES}. Thus, the second factor in equation 8.2.2 is just two factors of the probability of having N events in the data given n expected events, which is given by a Poisson probability:

$$q(N, n) \equiv \frac{e^{-n} n^N}{N!}, \tag{8.2.3}$$

so that we can write the probability to observe the data as:

$$P(D|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) = \sum_{N_s=0}^{N_s=N} P(D|n_s, n_b, N_s, N_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) \cdot q(N_s, n_s)q(N_b, n_b). \quad (8.2.4)$$

The probability of observing the data given N_s and N_b does not depend on n_s and n_b, so we can simplify this as:

$$P(D|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) = \sum_{N_s=0}^{N_s=N} P(D|N_s, N_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) \cdot q(N_s, n_s)q(N_b, n_b). \quad (8.2.5)$$

Focusing on $P(D|N_s, N_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}})$, there are 2^N different assignments of signal or background to the N events in D. Note that these are all possible assignments, as we don't know which events are signal and which events are background. Let's focus on one such particular assignment, for example the first N_s events coming from $t\bar{t}$ and the last N_b events arising from background. The probabilities for the values of M_t^{reco} and W_{jj} in the data are then given by the 2d KDE described in Chapter 6:

$$P(D|N_s, N_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) = \prod_{Z=1}^{Z=N_s} f_s({M_t^{\mathrm{reco}}}_Z, {W_{\mathrm{jj}}}_Z|\mathrm{M_{top}}, \Delta_{\mathrm{JES}}) \prod_{K=N_s+1}^{K=N} f_b({M_t^{\mathrm{reco}}}_K, {W_{\mathrm{jj}}}_K|\Delta_{\mathrm{JES}}), \quad (8.2.6)$$

where $Z(K)$ are indices ranging over the first N_s (last N_b) events, and f_s and f_b are the signal and background probabilities given by the 2d density estimates obtained from KDE.

The above is one way to assign N_s signal events to the N total events, but other ways are possible. For example, another assignment puts the first N_b events as background, and the last N_s events as signal. In total, there are $\binom{N}{N_s}$ ways for us to assign N_s signal events to the N events. Each of these ways is equally likely, so the probability for any one of them to occur is $\frac{1}{\binom{N}{N_s}}$. Summing over all these possible ways, we can write Equation 8.2.5 as a double sum:

$$
\begin{aligned}
P(D|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) = \sum_{N_s=0}^{N_s=N} P(D|N_s, N_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) q(N_s, n_s) q(N_b, n_b) = \\
\sum_{N_s=0}^{N_s=N} \sum_{\text{assignments}} \frac{q(N_s, n_s) q(N_b, n_b)}{\binom{N}{N_s}} \times \\
\prod_{Z \in \mathrm{sig}} f_s({M_t^{\mathrm{reco}}}_Z, {W_{\mathrm{jj}}}_Z | \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) \times \prod_{K \in \mathrm{bkgd}} f_b({M_t^{\mathrm{reco}}}_K, {W_{\mathrm{jj}}}_K | \Delta_{\mathrm{JES}}),
\end{aligned}
\tag{8.2.7}
$$

where the sum over assignments is a sum over all possible ways for us to assign N_s out of N events in D as signal events. The double sum over all possible values of N_s and all possible ways to assign a given N_s can be reinterpreted as a single sum over all possible 2^N assignments. Additionally, the fraction immediately after the sums can be rewritten as well:

$$
\frac{q(N_s, n_s) q(N_b, n_b)}{\binom{N}{N_s}} = \frac{e^{-n_s} n_s^{N_s}}{N_s!} \frac{e^{-n_b} n_b^{N_b}}{N_b!} \frac{(N - N_s)! N_s!}{N!}. \tag{8.2.8}
$$

Using the fact that $N - N_s = N_b$, rearranging and multiplying by a factor of

$1 = \frac{(n_s+n_b)^N}{(n_s+n_b)^N}$ gives a much simpler form:

$$\frac{q(N_s, n_s)q(N_b, n_b)}{\binom{N}{N_s}} = \frac{e^{-(n_s+n_b)}(n_s+n_b)^N}{N!}\frac{n_s^{N_s}n_b^{N_b}}{(n_s+n_b)^N}$$
$$= q(N, n_s+n_b)\frac{n_s^{N_s}n_b^{N_b}}{(n_s+n_b)^N}. \quad (8.2.9)$$

Plugging this in to Equation 8.2.7 and swapping the double sum for a single sum gives:

$$P(D|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) = \sum_{2^N \text{assignments}} q(N, n_s+n_b)\frac{n_s^{N_s}n_b^{N_b}}{(n_s+n_b)^N}\times$$
$$\prod_{Z\in\mathrm{sig}} f_s(M_t^{\mathrm{reco}}{}_Z, W_{\mathrm{jj}_Z}|\mathrm{M_{top}}, \Delta_{\mathrm{JES}}) \times \prod_{K\in\mathrm{bkgd}} f_b(M_t^{\mathrm{reco}}{}_K, W_{\mathrm{jj}K}|\Delta_{\mathrm{JES}}).$$
$$(8.2.10)$$

As for the sum over the two sets of products, this can be simplified as the product over all events of the sum of signal and background probabilities if we rearrange the factors of $n_s^{N_s}$ and $n_b^{N_b}$ and place them inside the products:

$$\sum_{2^N \text{assign.}}\left[\prod_{Z\in\mathrm{sig}} n_s f_{sZ}\right]\times\left[\prod_{K\in\mathrm{bkgd}} n_b f_{bK}\right] \to \prod_{i=1}^{N}(n_s f_{si} + n_b f_{bi}). \quad (8.2.11)$$

Equation 8.2.10 then reduces to:

$$P(D|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) = \frac{q(N, n_s + n_b)}{(n_s + n_b)^N} \times$$
$$\prod_{Z=1}^{N} \left[n_s f_s({M_t^{\mathrm{reco}}}_Z, {W_{\mathrm{jj}}}_Z | \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) + n_b f_b({M_t^{\mathrm{reco}}}_Z, {W_{\mathrm{jj}}}_Z | \Delta_{\mathrm{JES}}) \right]. \tag{8.2.12}$$

Rearranging by moving the $(n_s + n_b)$ term inside the product gives the form of the extended maximum likelihood that most people are accustomed to seeing:

$$P(D|n_s, n_b, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) = q(N, n_s + n_b) \times$$
$$\prod_{Z=1}^{N} \frac{\left[n_s f_s({M_t^{\mathrm{reco}}}_Z, {W_{\mathrm{jj}}}_Z | \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) + n_b f_b({M_t^{\mathrm{reco}}}_Z, {W_{\mathrm{jj}}}_Z | \Delta_{\mathrm{JES}}) \right]}{n_s + n_b}. \tag{8.2.13}$$

8.3 Fit and Background Constraint

We extend the likelihood defined in Equation 8.2.13 by including prior information about the expected number of background events. As discussed in Chapter 7, there is an *a priori* estimate for the background expectation. Using this information in the likelihood fit should improve the measurement. The background expectation is included as a Gaussian constraint on the expected number of background events. No such constraint is included on the expected number of signal events, as that depends on $\sigma_{t\bar{t}}$, which itself depends on $\mathrm{M_{top}}$. The likelhood for the kth subsample is defined as:

$$\mathcal{L}^k = P^k(D^k|n_s^k, n_b^k, \mathrm{M_{top}}, \Delta_{\mathrm{JES}}) \cdot \mathrm{Exp}\left(\frac{-(n_b^k - n_{0b}^k)^2}{2(\sigma_b^k)^2}\right), \tag{8.3.1}$$

where k refers to either 1-tag or the 2-tag subsample and the Gaussian term constrains the fit for the expected number of background events (n_b^k) to remain close to the *a priori* estimate (n_{0b}^k), within the uncertainty (σ_b^k).

8.4 Multiple subsamples

The likelihood given in Equation 8.3.1 is applicable for one particular subsample of events, for example, the 1-tag subsample. The machinery fits for the two tag subsamples by fitting independently for the values of n_b^k and n_s^k for both k. The likelihood forms are multipled together so that the likelhood over the two subsamples is given as:

$$\mathcal{L}_{\mathrm{shape}} = \prod_{Q=1}^{2} \mathcal{L}^Q. \tag{8.4.1}$$

8.5 $\Delta_{\mathbf{JES}}$ prior and log likelihood

The final likelihood is a product over all subsample likelihoods as defined in Equation 8.4.1 with an additional Gaussian constraint on the measured Δ_{JES}. This prior constrains the Δ_{JES} to the nominal CDF value within its uncertainty, 1 unit of σ_c:

$$\mathcal{L} = \mathcal{L}_{\mathrm{shape}} \cdot \mathrm{Exp}\left(\frac{-\Delta_{\mathrm{JES}}^2}{2\sigma_c^2}\right). \tag{8.5.1}$$

In practice, what is used inside MINUIT is not Equation 8.5.1, but $-\ln(\mathcal{L})$. The ln of the likelihood is easier and faster to compute, since the exponentials convert to polynomials. Negating this value allows for the use of a minimization routine instead of a maximization routine.

8.6 Local Polynomial Smoothing

Figure 8.1 shows example likelihoods for the 2-tag 1d fit (using M_t^{reco} as the only observable and fitting only for $\mathrm{M_{top}}$, not Δ_{JES}). The distribution of $-\ln(\mathcal{L})$ has the expected parabolic shape, so that analytically fitting a quadratic near the minimum does a reasonably good job of extrapolating the likelihood in between mass points. Similarly, the 2d likelihood can be fit to a 2d quadratic with a cross-term to account for $\mathrm{M_{top}} - \Delta_{\mathrm{JES}}$ correlations. On the other hand, the likelihood is not expected to be completely Gaussian; for example, the errors can be asymmetric, and without infinite statistics in our pseudoexperiments, the tails of the likelihood may also exhibit non-Gaussian behavior. Perhaps more importantly, the values of the likelihood have errors associated with finite MC statistics. This can be seen in Figure 8.1 as the jitter from point-to-point from the expected smooth shape. Fitting can remove the jitter, but a better approach is to avoid the assumption of a particular functional form for the likelihood and use another non-parametric technique.

One such technique goes by the name of local polynomial smoothing (LPS), which is a variant of a standard least-squares polynomial fit [44]. LPS models the relationship between a predictor variable x and the response variable Y. There are n pairs of observations $(x_1, Y_1), (x_2, Y_2), ...(x_n, Y_n)$. In the case of the likelihood fit, the x are the observables M_t^{reco} and W_{jj}, and the Y is the value of the likelihood

In the limit of infinite statistics, the values of Y are uniquely defined by x:

$$Y_i = \mu(x_i), \tag{8.6.1}$$

so that μ_i is the true values of the likelihood at x_i. Locally, assume that μ can be approximated by a Taylor expansion around x. Unlike in a typical least squares fit, LPS weights the MC values by how far away they are from the point of interest. Similar to KDE, it is necessary to define a bandwidth h and a smoothing window, such that only points $x_i \in [x-h, x+h]$ contribute to the LPS estimate at x. Also similar to KDE, a weighting function is defined:

$$w_i(x) = W\left(\frac{x_i - x}{h}\right). \tag{8.6.2}$$

We use the tribcube weighting function:

$$W(t) = (1 - \left|t^3\right|)^3. \tag{8.6.3}$$

Applying a Taylor expansion and using some compact vector notation gives:

$$\mu(t) \approx a_0 + a_1(t-x) + \frac{a_2}{2}(t-x)^2 \equiv < a, A(t-x) > . \tag{8.6.4}$$

Following a least squares fit, the coefficients of the vector a can be obtained by minimizing the sum of the distances from the observed values. In LPS, the distances are weighted via the weighting function, so that the function that is minimized is:

$$\chi^2_{\mathrm{LPS}} = \sum_{i=1}^{n} W_i(x)(Y_i - < a, A(x_i - x) >)^2. \tag{8.6.5}$$

If $W_i(x)$ is set to a constant, the normal least squares fit is recovered. In 2d LPS, the response variable is defined as:

$$Y_i = \mu(x_{i,1}, x_{i,2}), \tag{8.6.6}$$

and the Taylor expansion is:

$$\mu(t_1,t_2) \approx a_0 + a_1(t_1 - x_1) + a_2(t_2 - x_2) + \frac{a_3}{2}(t_1 - x_1)^2 + \\ \frac{a_4}{2}(t_2 - x_2)^2 + a_5(t_1 - x_1)(t_2 - x_2) =< a, A(t-x) > . \tag{8.6.7}$$

In practice, signal MC is generated at 76 mass points ranging from $\mathrm{M_{top}} = 150-240$ GeV/c^2. The spacing between mass points is small (0.5 GeV/c^2) in the region of interest ($\mathrm{M_{top}} = 160-185$ GeV/c^2), and gets larger in the tails of the mass grid. Each signal MC and MC-based background is processed using 29 different values of Δ_{JES} from -3.0 σ_{c} to +3.0 σ_{c}. Near the range of interest of nominal Δ_{JES}, the spacing is 0.2 σ_{c}.

After studying different bandwidths and optimizing for both bias and expected error, we use $\mathrm{h}_{\mathrm{M_{top}}} = 10$ GeV/c^2 and $\mathrm{h}_{\Delta_{\mathrm{JES}}} = 0.8\sigma_{\mathrm{c}}$ for the signal events. To smooth out backgrounds, which have a high degree of jitter across Δ_{JES} due to low statistics and effects arising from selecting events near threshold, we use $\mathrm{h}^{\mathrm{bkgd}}_{\Delta_{\mathrm{JES}}} = 3.0\sigma_{\mathrm{c}}$.

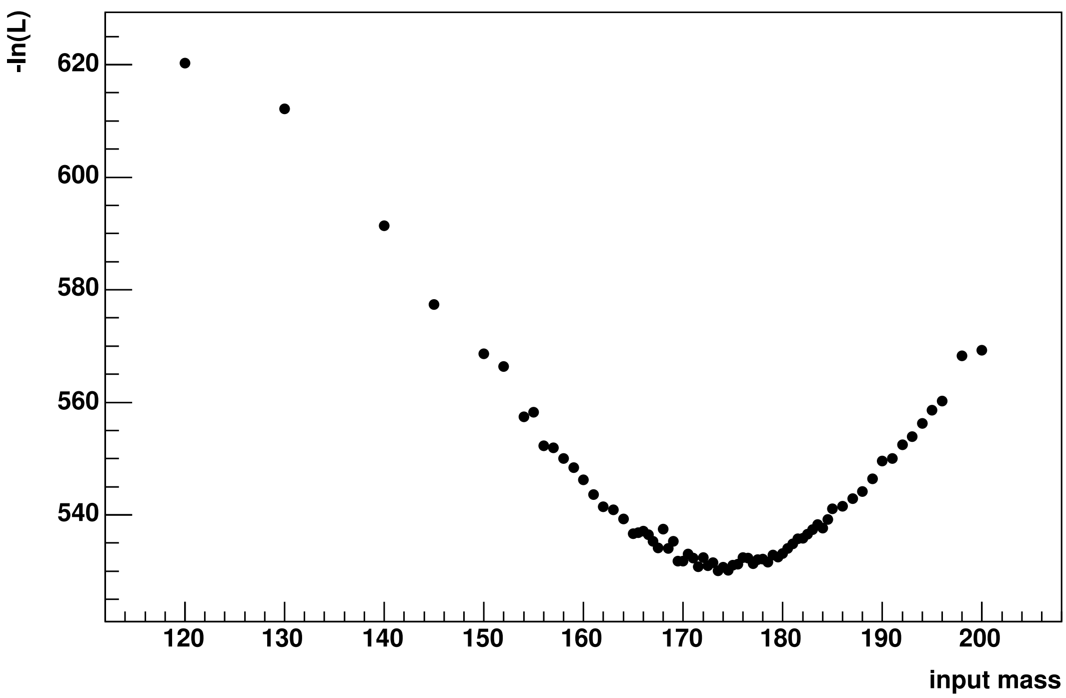

Figure 8.1: Example 1d $-ln(\mathcal{L})$ for a single toy MC experiment.

CHAPTER 9

CHECK OF MACHINERY

To validate the machinery of this analysis, a number of checks for possible sources of bias are examined in the form of pseudoexperiments (PEs). For each bias check, pseudodata are drawn from fully simulated $t\bar{t}$ MC with a certain value of $\mathrm{M_{top}}$ and Δ_{JES}. Background events are drawn with the same value of Δ_{JES} as the signal events. The number of signal events drawn in each pseudoexperiment for the separate tag categories is subject to Poisson fluctuatations around the expected number of signal events from the theoretical cross section, as given in Table 7.1. The number of background events is subject to similar Poisson fluctuations around expectations. Background events are drawn with weights according to their relative normalizations, including the event-by-event mistag normalizations. For historical reasons, all events are drawn before the χ^2 cut, and are rejected before input to the likelihood if they fail the cut.

To account for fluctuations of Δ_{JES} in the Gaussian constraint inside the likelihood, the central value of the constraint is varied from PE-to-PE by randomly drawing it from a Gaussian with a 1 σ_{c} width. When values of Δ_{JES} other that the nominal value are used for pseudodata, the constraint is drawn from a Gaussian centered at the input value. To account for fluctuations in n_b in the Gaussian constraint inside the likelihood, the values of the two constraints (one for each subsample) are also varied inside the likelihood from PE-to-PE. The constraints are drawn from Gaussians centered around the expectations, with widths equal to the values of the uncertainties in the *a priori* background estimates.

9.1 Bias checks

The legend for bias check plots is shown in Figure 9.1. The most important result of the bias checks is shown in Figure 9.2, which plots the mass residual (distance from the expected value as a function of M_t^{reco}). The output mass is taken as the mean value of M_{top} from 3000 PEs. Based on this plot, we conclude that the method is unbiased. The widths of pull distributions, defined using the asymmetric errors from MINOS after minimization, are shown in Figure 9.3. The RMS of the pull distribution is used to define the width. Based on the pull width distribution, we conclude the need to inflate errors from the data fit by 2.9 % to guarantee the 68.3 % coverage as typically defined by $\Delta \ln \mathcal{L} = 0.5$ uncertainties.

The bias check plots contain points with varying Δ_{JES}, but the fits are only to points using the nominal Δ_{JES}. The additional points come from the same MC samples, and thus PEs with varying Δ_{JES} but the same input M_{top} are highly correlated.

The value of Δ_{JES} is not a physical quantity and has no meaning outside of CDF, but it nice to see that its measurement is unbiased. Figure 9.4 gives the Δ_{JES} residual as a function of M_{top}, showing an unbiased result to 1 %. Figures 9.5 and 9.6 show the mass residual and pull width as a function of Δ_{JES}. As before, no fits are applied, since the points are highly correlated, though there is no clear trend with Δ_{JES}. For some mass points, the residual goes slightly down and then up with increasing Δ_{JES}, but the trend is reversed for other masses. In either case the changes are small, and we conclude that this is low-level noise and the method is unbiased across Δ_{JES}. Figure 9.7 shows the Δ_{JES} residual as a function of Δ_{JES}.

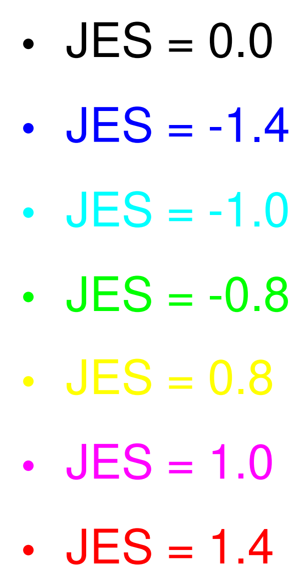

Figure 9.1: Legend for bias checks.

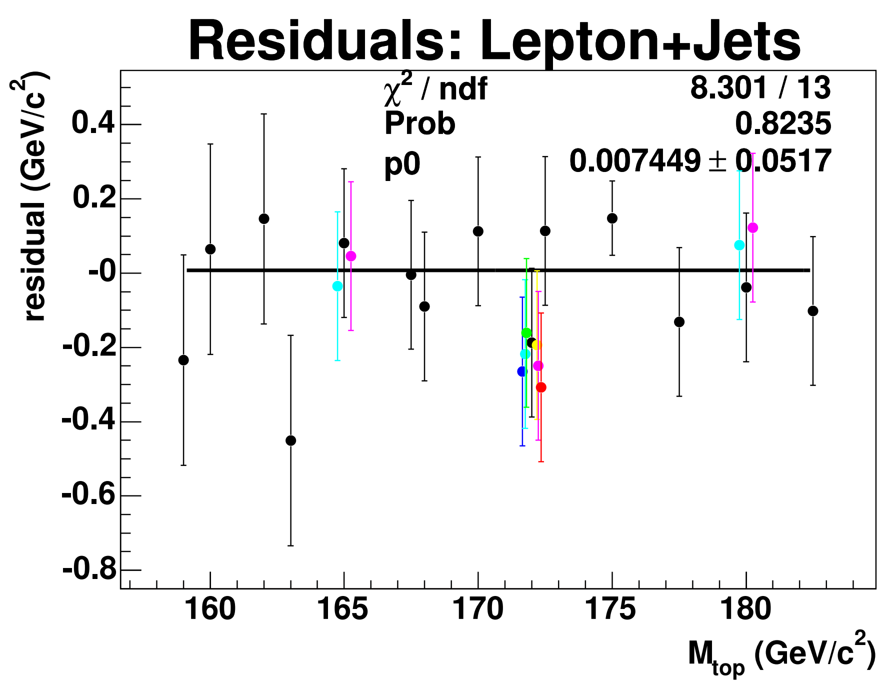

Figure 9.2: Residual top mass bias as a function of M_{top}. The fit is only to points with nominal Δ_{JES}.

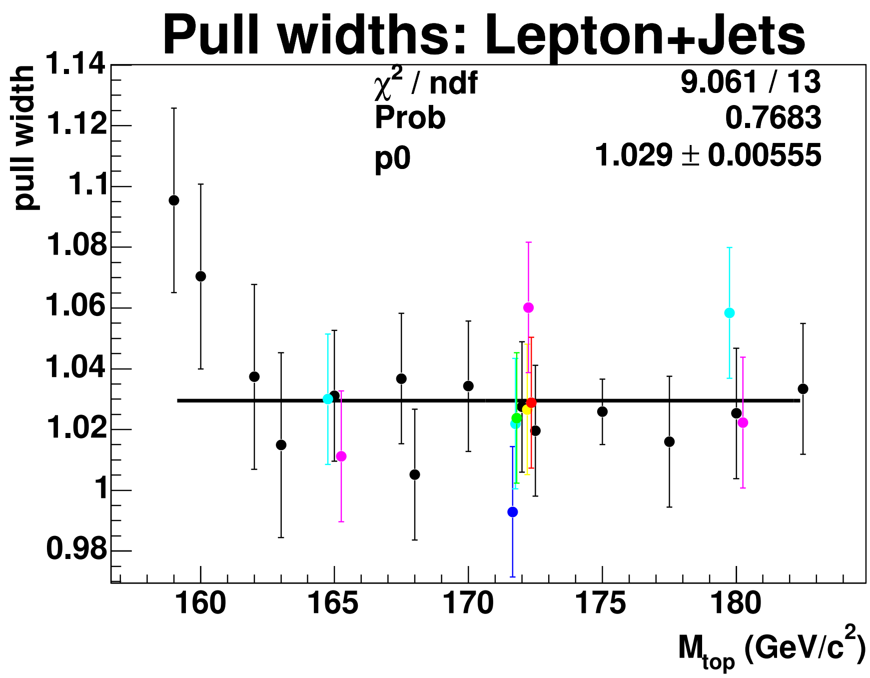

Figure 9.3: Pull width as a function of $\mathrm{M_{top}}$. The fit is only to points with nominal Δ_{JES}.

9.2 Bootstrap

The uncertainties on the the bias checks and systematics due to limited MC statistics are derived using the bootstrap method. The uncertainties do not follow simple $\sqrt{N}$ rules, since the typical 3000 PE ensemble is grossly oversampled; if the ensembles were to be derived from completely independent signal samples, a typical mass point would hold only 50-100 PEs. Additionally, the uncertainty due to limited background MC statistics, while unknown and largely correlated across mass points, must be understood.

If the estimation is not done correctly, error bars that are too small can falsely indicate bad machinery or a bias. On the other hand, error bars that are too large can hide problems with an analysis. For detailed information on the bootstrap method,

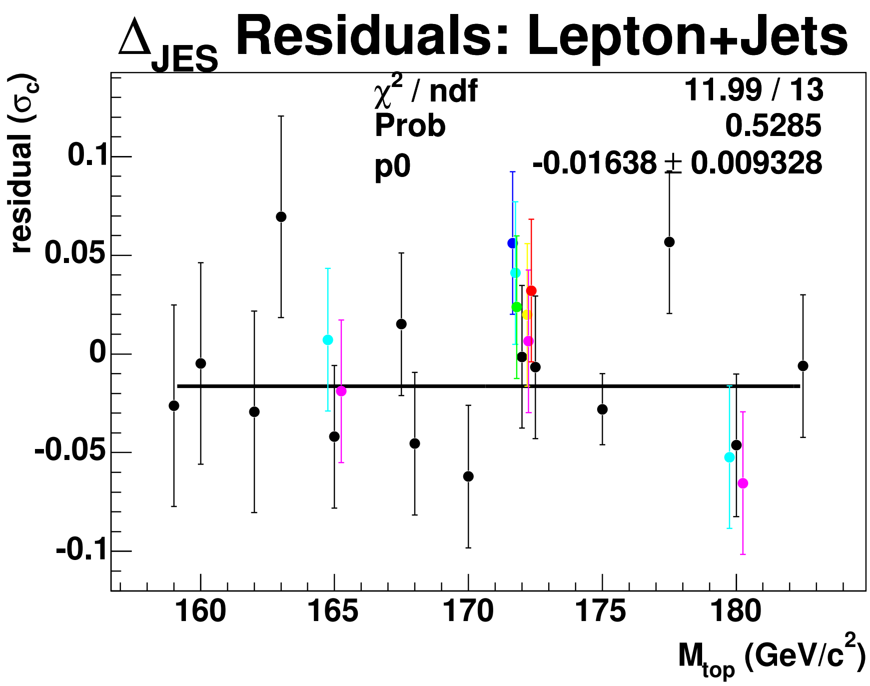

Figure 9.4: Residual Δ_{JES} bias as a function of $\mathrm{M_{top}}$. The fit is only to points with nominal Δ_{JES}.

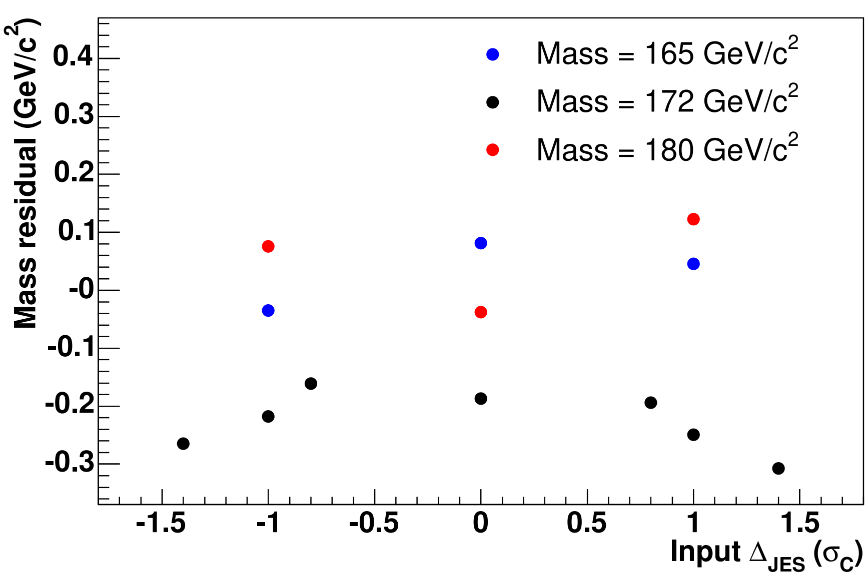

Figure 9.5: Residual top mass bias as a function of Δ_{JES}.

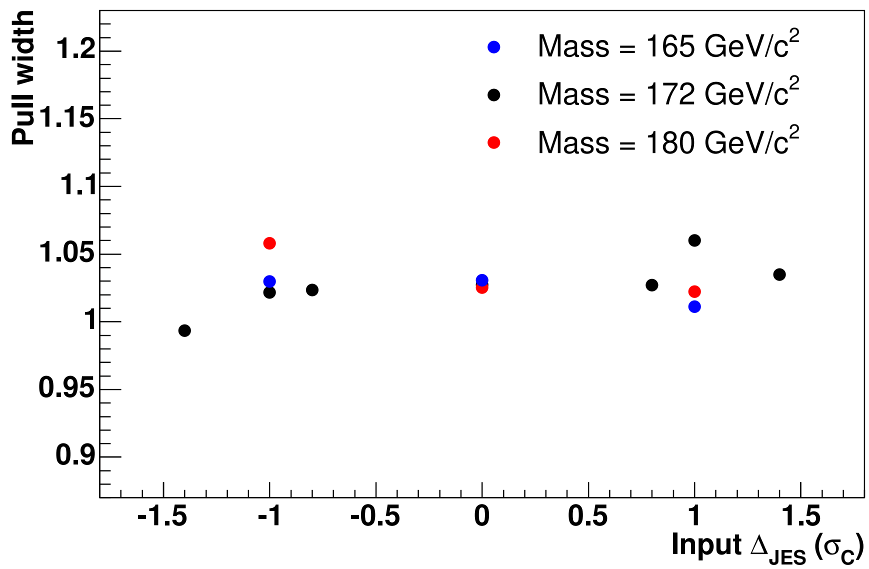

Figure 9.6: Pull width as a function of Δ_{JES}.

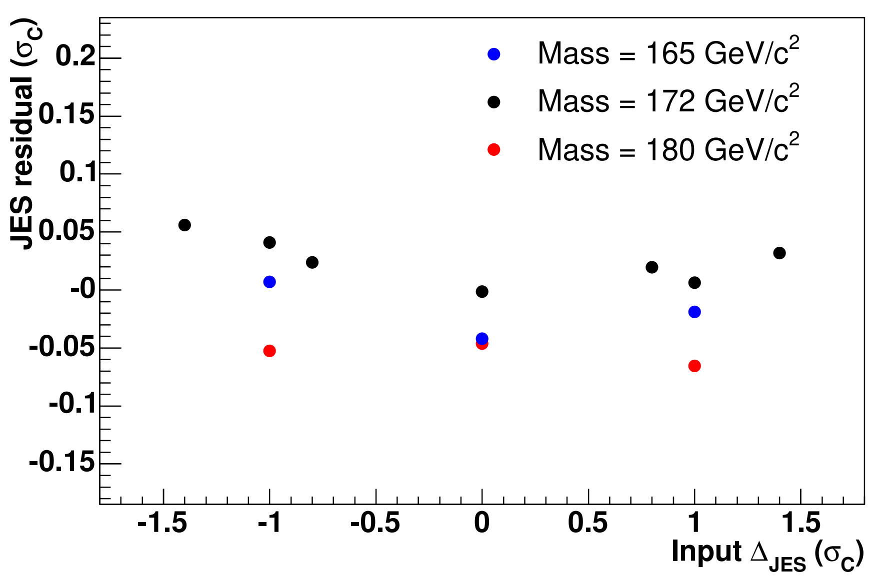

Figure 9.7: Residual Δ_{JES} bias as a function of Δ_{JES}.

see [45] and [46]. A brief summary of the application of the bootstrap technique to this analysis is as follows:

- Start with an ntuple with n signal events.
- Sample with replacement from the ntuple n times to produce a new set of pseudoedata derived from the original sample. In other words, we have a new quasi-ntuple with the same number of events (n) as the original sample. All events in the new quasi-ntuple come from the original ntuple, but some events may be duplicated multiple times, and others may not appear at all. Run a full set of pseudoexperiments on this altered sample.
- Repeat the above sampling with replacement and running of pseudoexperiments N times.
- Plot any quantity of interest (pull mean, pull width, mass RMS, etc.) from the N ensembles. The RMS of this distribution is a reasonable estimate of the uncertainty on the quantity.

In practice, the MC sample with nominal Δ_{JES} and $\mathrm{M}_t^{\mathrm{reco}} = 170$ GeV/c^2 is bootstraped 65 times ($N = 65$) to give estimates of uncertainties. Uncertainties on quantities for samples genereted with a different number of MC events are derived by subsequently scaling by $1/\sqrt{N}$. For example, MC samples generated with twice the statistics as the MC at 170 GeV/c^2 are estimated to have statistical uncertainties $\sqrt{2}$ times smaller. Bootstrapped output masses, which give the error on the residual bias, are shown in Figure 9.8. Bootstrapped pull widths are shown in Figure 9.9.

Independently of the above PE ensembles, the background samples are also bootstraped to evaluate the effect of limited statistics in the background pseudodata. The

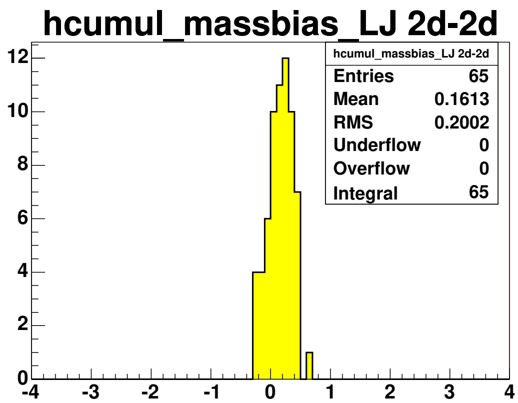

Figure 9.8: Output mass residual for bootstrapped samples.

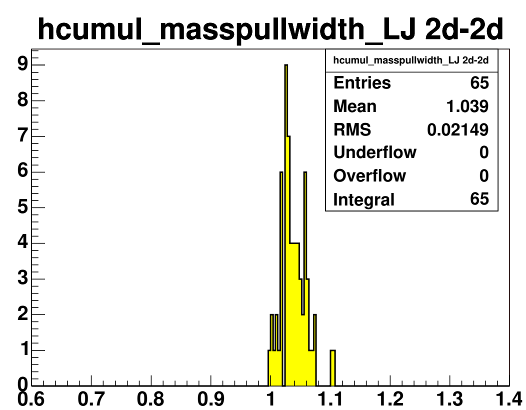

Figure 9.9: Output pull widths for bootstrapped samples.

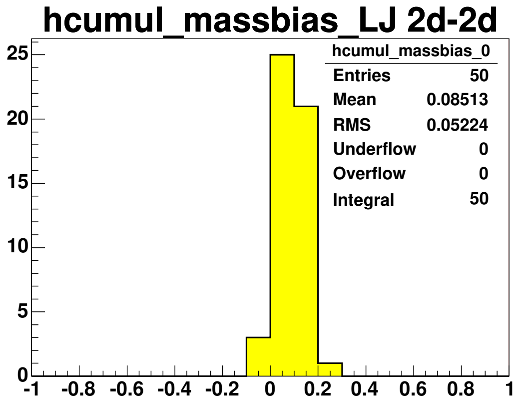

Figure 9.10: Output mass residual for background bootstrapped samples.

bootstrap is performed without taking into account the event weights so that each event has equal chance to enter the bootstrap set, but still carries its appropriate weight for drawing pseudodata. The output mass of 50 background bootsrap ensembles is shown in Figure 9.10, leading to a 50 MeV/c^2 systematic for limited background pseudodata statistics (though the background pseudodata often has low statistics, the systematic is small because most of the backgrounds have similar shapes).

CHAPTER 10

SYSTEMATIC UNCERTAINTIES

Any mismodeling of the detector response and event reconstruction has the possibility to systematically shift the top quark mass measurement. Similarly, mismodeling of $t\bar{t}$ and background production and decay properties can also lead to systematic shifts in the measurement. Uncertainities due to unknown systematic parameters or effects are estimated by running pseudoexperiments in which the pseudodata is altered from the nominal set described in Chapter 9. The pseudoexperiments use the same number of events as for the bias checks. To be conservative, when a systematic shift has an associated statistical uncertainty (due to finite MC statistics and estimated using the bootstrap method) larger than the shift itself, we take the larger value as the systematic uncertainty. When comparing highly correlated samples resulting from reweighting events, we do not use the statistical uncertainty on the samples.

10.1 Residual JES

As a single nuisance parameter, the Δ_{JES} that is measured in the likelihood does not fully capture the complexities of possible jet energy scale uncertainties, particularly those with different η and p_T dependence. Fitting for the global JES removes most of these effects, but not all of them. To study any possible remaining (residual) systematic uncertainty due to the JES, we run PEs where we shift both the signal and background pseudodata up and down by 1σ for each of the separate JES levels, as described in Section 4.4. The JES prior is removed for these pseudoexperiments so

Table 10.1: Mean output from pseudoexperiments used to evaluate residual JES systematic effects.

Sample	M_{top} (GeV/c^2)	Δ_{JES} (σc)
PYTHIA M_{top} = 175.0 GeV/c^2	175.10	-0.03
$+1\sigma$ Relative correction	175.14	0.16
-1σ Relative correction	175.17	-0.21
$+1\sigma$ Multiple interaction correction	175.25	0.04
-1σ Multiple interaction correction	175.11	-0.06
$+1\sigma$ Absolute correction	175.49	0.51
-1σ Absolute correction	174.65	-0.53
$+1\sigma$ Underlying event correction	175.12	0.07
-1σ Underlying event correction	175.17	-0.11
$+1\sigma$ Out-of-cone correction	174.78	0.78
-1σ Out-of-cone correction	175.40	-0.79
$+1\sigma$ Splash-out	175.14	0.12
-1σ Splash-out	175.23	-0.16

as not to bias the fits, and half of the difference between the $+1\sigma$ and -1σ shifts is taken as the systematic uncertainty for each of the JES levels. The input mass for the signal pseudodata is 175.0 GeV/c^2. Results from these pseudoexperiments are shown in Table 10.1. The residual systematic effect due to uncertainties in the absolute and OOC corrections dominate. For the purposes of evaluating systematic effects, the uncertainty due to OOC corrections is divided into two parts - the uncertainties due to energy falling out of the jet cone but within $\Delta R < 1.3$ is called the out-of-cone uncertainty, and any remaining, small leakage energy is referred to as splash-out. The total sum in quadrature for the residual JES systematic uncertainty is 0.53 GeV/c^2.

10.2 Background modeling uncertainties

This analysis is mostly insensitive to uncertain modeling of the background, as the background templates tend to peak at much lower values of M_t^{reco} than the signal templates. To examine possible systematic effects due to how the backgrounds are stitched together (and because the background weights relative to one another have uncertainties), we run pseudoexperiments drawing only from certain types of backgrounds. In particular, we run PEs drawing only from: W$b\bar{b}$, W$c\bar{c}$, Wc, and W+light jets, single top, QCD and the diboson samples. When a particular type of background consists of multiple n-parton samples, the entire n-parton set is used and the relative weights between them are kept as in normal PEs. For single top-only PEs, both the s-channel and t-channel samples are used. For diboson PEs, WW, WZ and ZZ are used. Results are included in Table 10.2. The largest shift (0.16 GeV/c^2) from the nominal value when using the background model comes from PEs drawing only from the mistag (W+light flavor jets) samples. This very conservative number (the background surely does not consist of only W+light flavor) is assigned as a background composition systematic.

To study possible mismodeling of the background, we run pseudoexperiments drawing from only W$b\bar{b}$ samples that have Q^2 (the square of the momentum transfer in the hard scatter process) changed by factors of 2^2 and 0.5^2. When drawing from these backgrounds, we weight the samples by the cross sections given by ALPGEN, which partially compensates for the changes in scale by giving different relative weights to the high- and low-parton samples. When the Q^2 scale is increased (decreased), ALPGEN increases the relative weight for the low- (high-) parton samples such that the net effect of changing scales is reduced.

Results are included in Table 10.2. We find a shift of 0.09 GeV/c^2, which is taken as the background shape systematic uncertainty.

Finally, possible mismodeling of the QCD background is also studied. The templates show very little difference when the anti-isolation cut is varied from > 0.2 to

Table 10.2: Mean output from pseudoexperiments used to evaluate background-related systematic effects.

Sample	M_{top} (GeV/c^2)	Δ_{JES} (σ_c)
PYTHIA M_{top} = 175.0 GeV/c^2	175.10	-0.03
W$b\bar{b}$ for bkgd	175.01	-0.01
W$c\bar{c}$ for bkgd	175.02	-0.04
Wc for bkgd	175.23	-0.03
W+light for bkgd	175.26	0.00
single top for bkgd	175.21	0.04
QCD for bkgd	175.07	-0.09
diboson for bkgd	175.24	-0.02
W$b\bar{b}$ Qfact=2.0 for bkgd	174.95	0.00
W$b\bar{b}$ Qfact=0.5 for bkgd	175.04	-0.01
anti-electron for QCD	175.16	-0.01

0.15 and 0.3. As a possible second model, templates are formed from antielectron samples, which consist of events where the isolated electron is replaced by objects in the detector that look almost, but not quite, like electrons. These antielectrons fail two of the kinematic cuts normally used to remove QCD events and clean up the electron sample. Very few of the antielectron events pass contain b-tags, so the mistag matrix is applied and events are weighted by their mistag probabilities. As shown in Figures 10.1 and 10.2, the non-isolated lepton and anti-electron templates do differ, though M_t^{reco} peaks at similar locations for both selections. Pseudoexperiments using the antielectrons in place of the normal QCD pseudodata show a shift of 0.06 GeV/c^2, which is taken as a QCD modeling uncertainty.

10.3 b-JES

Possible systematic effects due to uncertain modeling of b quark jets are also examined. These jets have different fragmentation than light quarks jets, as well as semi-leptonic decays and different color flow, effects which can result in a systematic

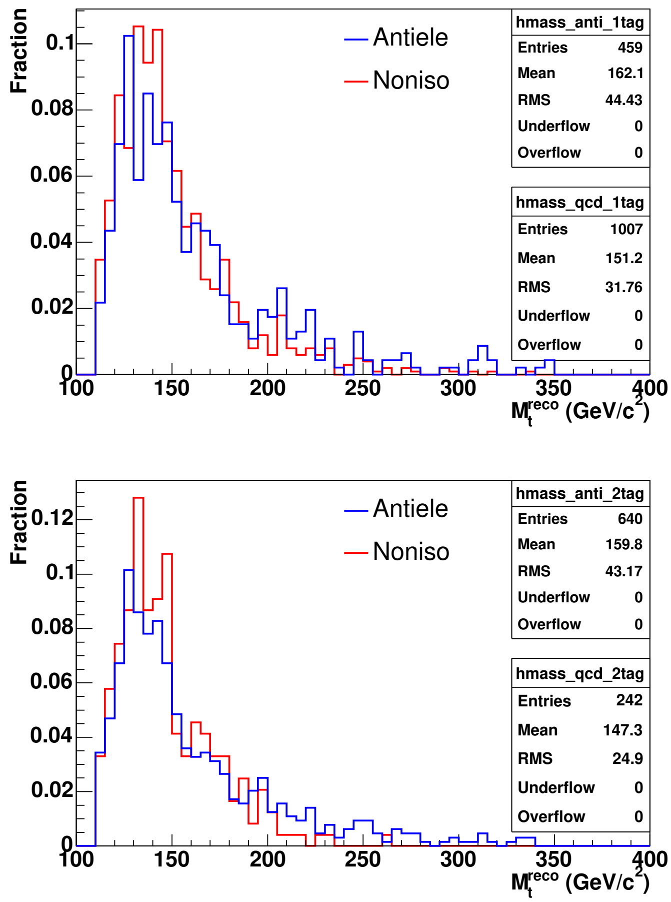

Figure 10.1: Comparison of $M_t^{\rm reco}$ for the nonisolated lepton sample (with real tags) and antielectrons (with the mistag matrix) for 1-tag (top) and 2-tag (bottom) events.

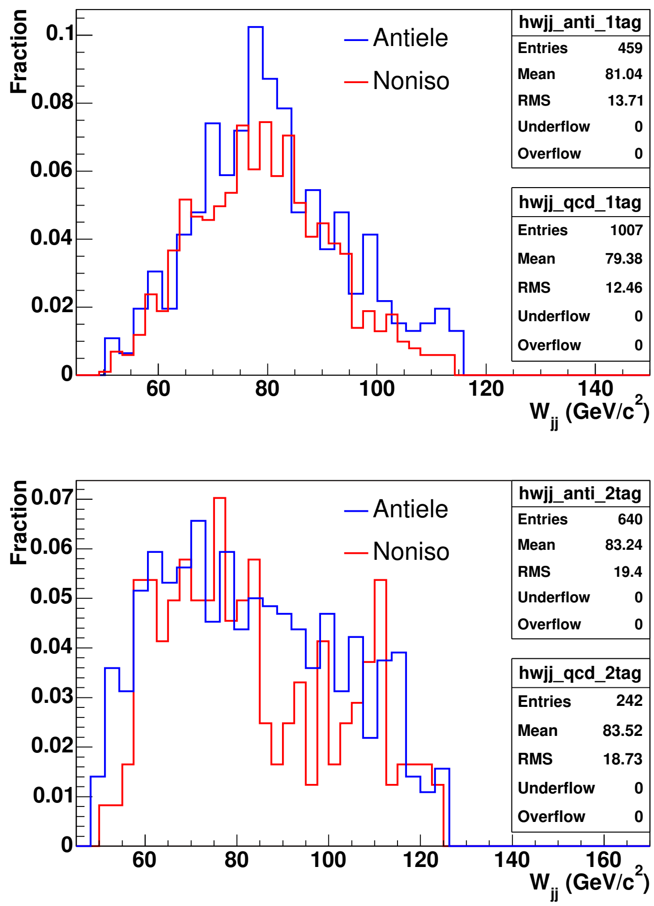

Figure 10.2: Comparison of W_{jj} for the nonisolated lepton sample (with real tags) and antielectrons (with the mistag matrix) for 1-tag (top) and 2-tag (bottom) events.

shift in the top quark mass if not correctly modeled. An estimate for uncertainties on these effects is 0.6% in b-specific energy scale [29]. To gain a bit in lever arm, we multiply the energies of jets matched to b quarks by factors of 1.01 and 0.99, corresponding to 1.0% shifts, and then try to scale uncertainties by 1.0/0.6. Using this prescription and taking half of the difference between the estimated $\pm$ 0.6% shifts, we find a systematic uncertainty of 0.35 GeV/c^2. To be conservative we use the old systematic uncertainty of 0.60 GeV/c^2 quoted in [29].

10.4 Parton Distribution Functions

The systematic uncertainty due to the parton distribution functions (PDFs) [47, 48] that provide the probability to find gluons and quarks of a given flavor in the proton and anti-proton with a certain momentum is estimated by reweighting PYTHIA events at M_{top} = 175.0 GeV/c^2. We compare PDFs from two different groups (CTEQ5L vs MRST72) and take the absolute difference as a systematic uncertainty. We also compare PDFs with different values of the strong coupling constant, corresponding to different QCD energy scales: MRST72 (Λ_{QCD} = 228 MeV) vs MRST75 (Λ_{QCD} = 300 MeV), again taking the absolute difference as a systematic. Finally, we examine the uncertainties in M_{top} due to shifting up and down by 1σ the 20 eigenvectors that define the CTEQ6M set, taking half of the difference between the $+1$ σ and -1σ shifts for each eigenvector pair. We find systematic shifts of 0.10 GeV/c^2 for the different groups, 0.09 GeV/c^2 for the different QCD scales, and 0.22 GeV/c^2 for the uncertainty on the eigenvectors, giving a total systematic uncertainty on PDFs of 0.25 GeV/c^2. We also test the effect of reweighting MC to increase the fraction of $t\bar{t}$ events initiated by gg (vs qq) from the 6% in the leading order MC to 20%. Results from these pseudoexperiments are summarized in Table 10.3.

Table 10.3: Mean output from pseudoexperiments used to evaluate parton distribution function-related systematic effects.

Sample	M_{top} (GeV/c^2)	Δ_{JES} (σ_c)
PYTHIA $M_{top} = 175.0$ GeV/c^2	175.10	-0.03
gg fraction reweighted	175.06	0.00
CTEQ5L	175.27	0.01
MRST72	175.18	0.01
MRST75	175.26	0.02

10.5 Other systematics

A systematic uncertainty due to using PYTHIA to model signal events is studied by comparing PEs that instead use HERWIG [49] at $M_{top} = 175.0$ GeV/c^2 for the signal pseudodata. In particular, HERWIG includes spin correlations between the top and antitop quarks, and also has a different fragmentation model. We find a difference of 0.10 GeV/c^2 in output mass from pseudoexperiments - the uncertainties on this difference dominate, so 0.24 GeV/c^2 is quoted as the generator systematic uncertainty.

Uncertainties in the modeling of gluon ISR and FSR are studied by extrapolating uncertainties in the p_T of Drell-Yan data to the $t\bar{t}$ region [29]. Unlike in [29], settings affecting radiation are shifted coherently for ISR and FSR, as the same parameters govern both types of radiation. The difference between the samples with more and less gluon radiation, 0.22 GeV/c^2, is smaller than uncertainty on this difference, 0.28 GeV/c^2, which is quoted as the radiation systematic uncertainty.

We also separately shift the energy scales for electrons and muons by moving them up and down 1% before our event selection. When muons have energies shifted, the $\not{E}_T$ is recalulated accordingly. Compared to the nominal sample, the largest shifts are 0.05 GeV/c^2 for each of the lepton types. We also add a 0.05 GeV/c^2 systematic uncertainty due to mismodeling of luminosity, and 0.05 GeV/c^2 uncertainties each

Table 10.4: Mean output from pseudoexperiments used to evaluate additional systematic effects.

Sample	M_{top} (GeV/c^2)	Δ_{JES} (σ_c)
PYTHIA $M_{top} = 175.0$ GeV/c^2	175.10	-0.03
HERWIG $M_{top} = 175.0$ GeV/c^2	175.00	0.26
More gluon radiation	175.32	0.13
Less gluon radiation	175.10	0.03
electron energy scale · 0.99	175.15	-0.04
electron energy scale · 1.01	175.14	-0.03
μ energy scale · 0.99	175.15	-0.03
μ energy scale · 1.01	175.12	-0.02
ALPGEN $t\bar{t}$ 0p + $t\bar{t} \geq$ 1P	175.46	0.27
ALPGEN $t\bar{t}$ inclusive	175.42	0.28

for finite signal and background MC statistics, as given by the bootstrap method described in Section 9.2.

We also examine ALPGEN pseudodata derived from the exclusive $t\bar{t}$ 0-parton sample in combination with the inclusive $t\bar{t} \geq$ 1-parton sample. The samples are weighted according to the acceptance and cross sections and showered with PYTHIA. We keep this as a cross-check, but do not use it to evaluate any additional systematics, as the shifts can largely be correlated with PDFs and ISR effects. This sample also gives very similar pseudoexperiments when compared to an entirely inclusive ALPGEN $t\bar{t}$ sample, indicating that the addition of extra hard partons in the matrix element does not affect the mass measurement.

Results from these PEs are shown in Table 10.4.

10.6 Summary of systematics

Table 10.5 summarizes the systematic uncertainties on this measurement. The total systematic uncertainty is 0.94 GeV/c^2.

Table 10.5: Summary of systematic uncertainties.

Systematic effect	Size of effect (GeV/c^2)
Generator	0.24
gg fraction	0.04
Radiation	0.28
ele energy scale	0.05
μ energy scale	0.05
Parton Distribution Functions	0.25
Residual JES	0.53
b-JES	0.60
MC signal statistics	0.05
MC bkgd statistics	0.05
Bkgd composition	0.16
Bkgd shape	0.09
QCD Model	0.06
Luminosity effects	0.05
Combined	0.94

CHAPTER 11

DATA FIT

Confident in our machinery and understanding of systematics, we search the data to find events passing all cuts. The number of events after kinematic selection in 1.9 fb^{-1} of data is summarized in Table 11.1. A comparison with Table 7.1 shows that we find $\sim 20\%$ more events than expected, indicating that the measured cross section from this sample would be larger than the 6.7 pb expected at 175.0 GeV/c^2. Our analysis does not depend on signal normalization, so we are not sensitive to this effect. The full machinery is applied to the data, giving:

$$\mathrm{M_{top}} = 171.76^{+1.80}_{-1.83}\ \mathrm{GeV}/c^2\ \text{(unscaled, statistics-only)}, \tag{11.0.1}$$

$$\mathrm{M_{top}} = 171.76^{+1.85}_{-1.88}\ \mathrm{GeV}/c^2\ \text{(pull width scaled, statistics-only)}. \tag{11.0.2}$$

The likelihood contours for the measurement are shown in Figure 11.1.

Table 11.1: Observed number of candidate events in data before χ^2 cut, after χ^2 cut, and after both χ^2 cut and boundary cut.

	1-tag	2-tag
Pre-χ^2, pre-boundary cuts	284	152
Post-χ^2, pre-boundary cuts	237	101
Post-χ^2, post-boundary cuts	233	99

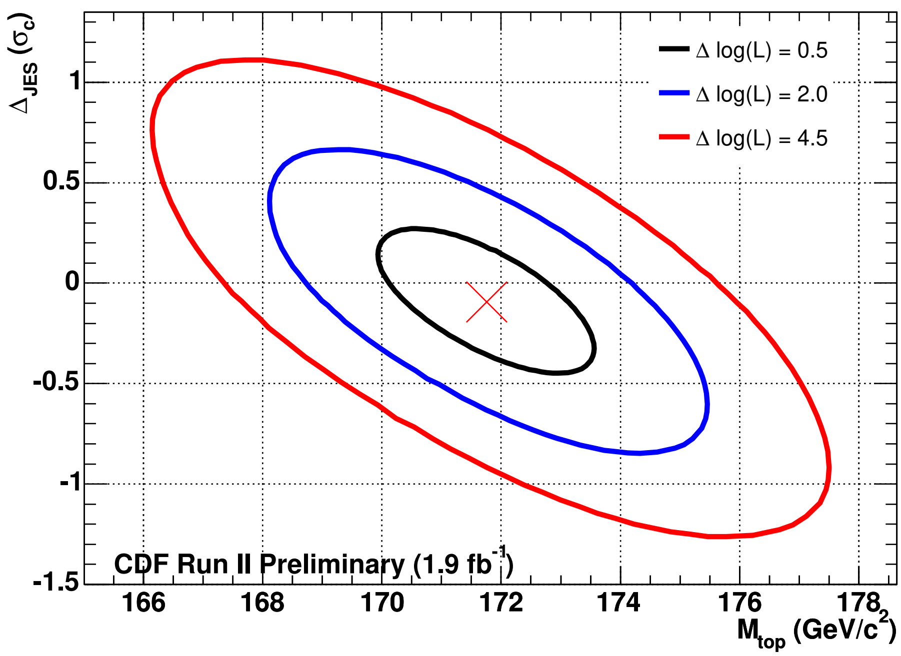

Figure 11.1: Likelihood contours for the combined fit. In the limit of full Gaussian behavior, the contours would correspond to regions of 1, 2 and 3 σ uncertainties.

CHAPTER 12

CROSS-CHECKS

We run a variety of cross-checks on the measurement. Results are summarized in Table 12.1. All uncertainties are uncorrected for pull width effects. Note that the delineation of the 1.9 fb^{-1} of data into different periods is typically a result of Tevatron shutdowns or significant changes to the CDF detector. As such, the different periods are not expected to correspond to equally sized datasets, and the errors between the different periods are not expected to be constant. With the exception of the nominal fit and the 1-tag- and 2-tag-only fits, the background constraint is turned off, as the *a priori* background estimation is not applicable.

Removal of the background and Δ_{JES} priors does not significantly alter the results, though the fit in the 1-tag subsample prefers approximately 1σ more background than the *a priori* estimation and the 2-tag subsample prefers to fit no background. The luck (ie p-value) of the data fits is examined by comparing the measured, symmetrized error obtained in data with those expected from pseudoexperiments. For this examination, the number of background events in each subsample is fluctuated around the expectation, with the total number of events fluctuated around the observed number and not the expected number. We use MC with the nominal Δ_{JES} and $\mathrm{M}_{\mathrm{top}} = 172.0\ \mathrm{GeV}/c^2$. Results of this check are summarized in Figure 12.1. Given an expected number of events based on the data and the *a priori* background estimation, 18 % of PEs show a smaller error than the one observed in data.

Figures 12.2, 12.3 and 12.4 show distributions from the kinematic fitter, comparing data to the fitted signal and background expectations using the full background

Table 12.1: Cross-checks on the data, uncorrected for pull width effects.

Fit type	M_{top} (GeV/c^2)		Δ_{JES} (σ_c)	
	1tag n_s	2tag n_s	1tag n_b	2tag n_b
Nominal	$171.76^{+1.80}_{-1.83}$		$-0.09^{+0.36}_{-0.36}$	
	184.0	96.4	47.5	3.4
No JES prior	$171.80^{+1.86}_{-1.88}$		$-0.11^{+0.39}_{-0.38}$	
	184.1	96.4	47.5	3.4
No bkgd prior	$171.78^{+1.82}_{-1.81}$		$-0.06^{+0.36}_{-0.36}$	
	175.7	99.0	57.3	0.0
1-tag only	$169.11^{+3.06}_{-2.55}$		$-0.17^{+0.48}_{-0.57}$	
	187.6	–	44.8	–
2-tag only	$173.57^{+2.64}_{-2.32}$		$0.20^{+0.47}_{-0.50}$	
	–	96.3	–	3.5
Data period 1 (no bkgd const)	$174.58^{+3.73}_{-3.53}$		$-0.18^{+0.68}_{-0.82}$	
	30.6	14.0	15.4	0.0
Data period 2 (no bkgd const)	$172.97^{+4.17}_{-3.63}$		$0.30^{+0.70}_{-0.68}$	
	29.0	21.0	22.0	0.0
Data period 3 (no bkgd const)	$171.06^{+4.44}_{-4.21}$		$1.05^{+0.69}_{-0.72}$	
	30.3	13.0	4.7	0.0
Data period 4 (no bkgd const)	$171.06^{+3.41}_{-3.38}$		$-0.60^{+0.55}_{-0.65}$	
	55.0	29.0	7.0	0.0
Data period 5 (no bkgd const)	$164.91^{+4.62}_{-4.42}$		$-0.23^{+0.74}_{-0.77}$	
	25.0	14.0	0.0	0.0
Data period 6 (no bkgd const)	$173.64^{+10.3}_{-7.8}$		$0.0^{+1.06}_{-1.01}$	
	5.0	8.0	9.0	0.0
ele only, no bkgd const	$172.22^{+2.66}_{-2.75}$		$-0.09^{+0.51}_{-0.51}$	
	95.1	54.0	35.9	0.0
μ only, no bkgd onst	$171.32^{+2.38}_{-2.26}$		$-0.04^{+0.46}_{-0.47}$	
	81.0	45.0	21.0	0.0
(-) lepton, no bkgd const	$169.68^{+2.51}_{-3.31}$		$0.19^{+0.72}_{-0.45}$	
	106.4	51.0	25.6	0.0
(+) lepton, no bkgd const	$174.95^{+3.25}_{-3.07}$		$-0.48^{+0.62}_{-0.62}$	
	106.4	51.0	25.6	0.0

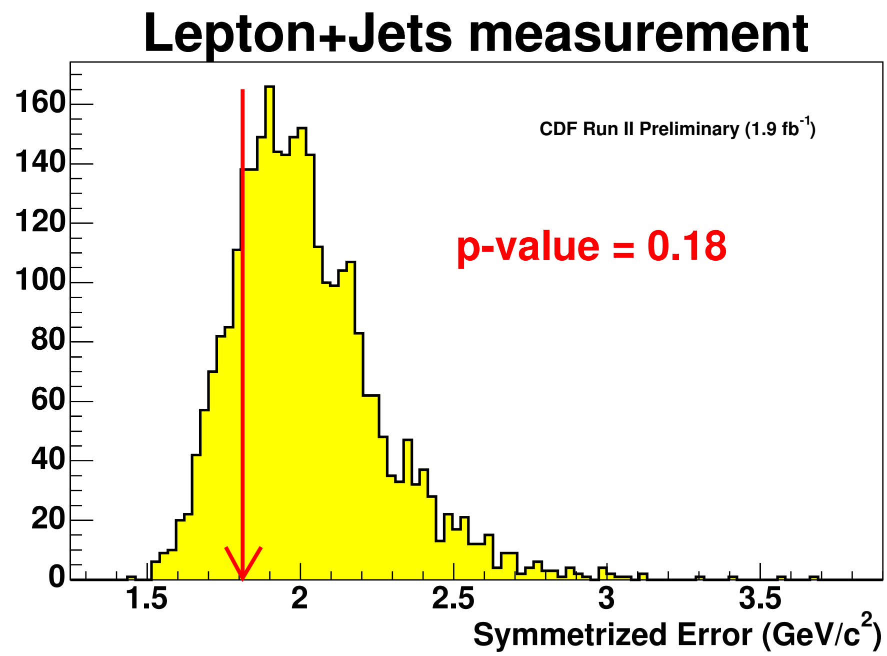

Figure 12.1: Expected errors from PEs using the observed number of events, nominal Δ_{JES} and $\mathrm{M}_t^{\mathrm{reco}} = 172.0\ \mathrm{GeV}/c^2$.

model. The signal model used in these plots is PYTHIA MC at $\mathrm{M}_{\mathrm{top}} = 172.0\ \mathrm{GeV}/c^2$. There is a wide range of agreement (or disagreement) from the very good to the very bad, with the p_T of the $t\bar{t}$ system and the p_T of the b jets having low KS probabilities. By looking at enough distributions, we expect a few to stand out as having poor probabilities. The distributions of the p_T of the $t\bar{t}$ system also look better in Figure 12.5, which uses HERWIG to model the signal. As described in Section 10.5, we include a systematic uncertainty on the difference between HERWIG and PYTHIA.

Figures 12.6 and 12.7 show the 1d data with the best-fit 1d signal and background distributions overlaid on top.

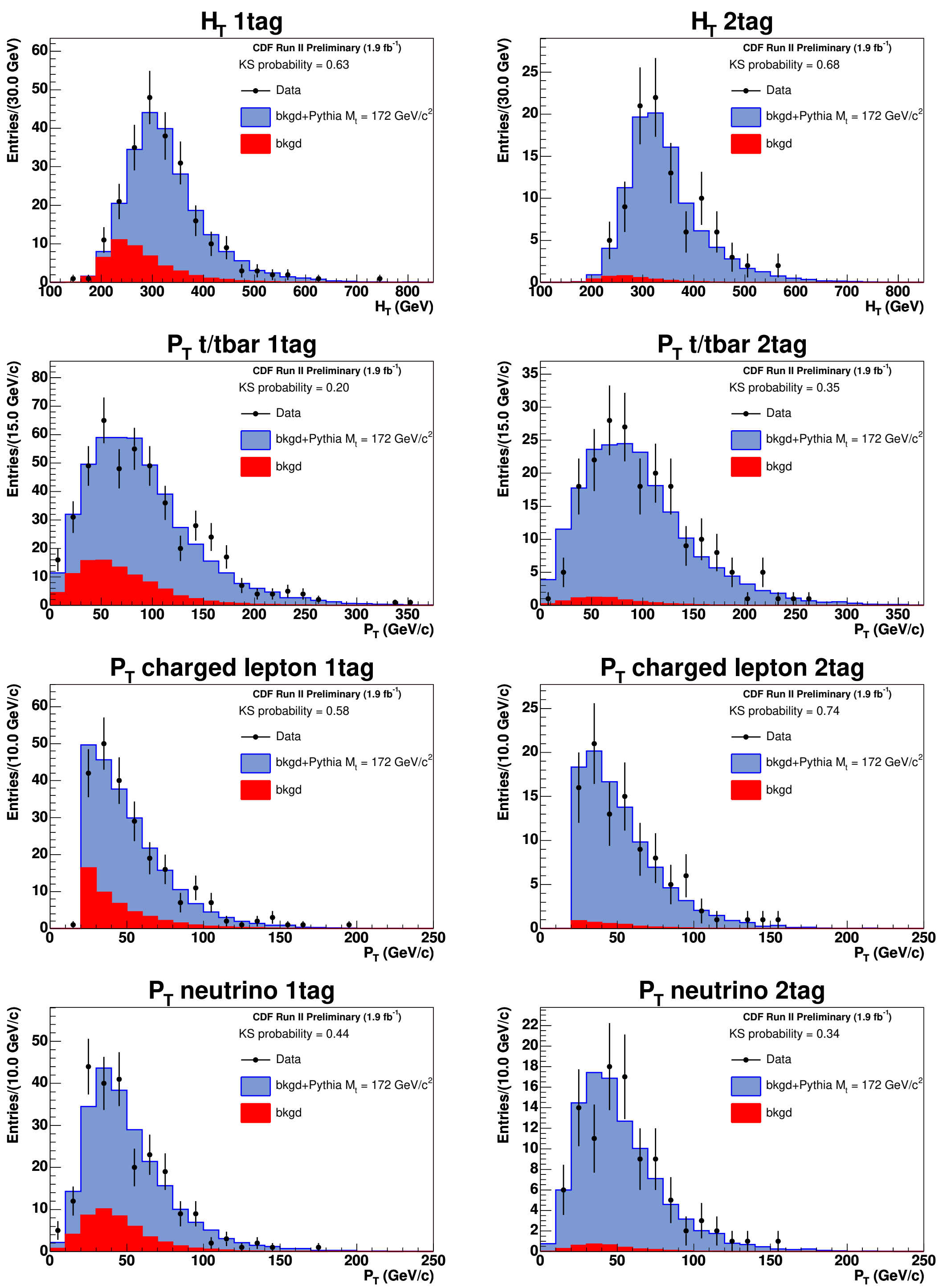

Figure 12.2: Distributions for fitter output (1).

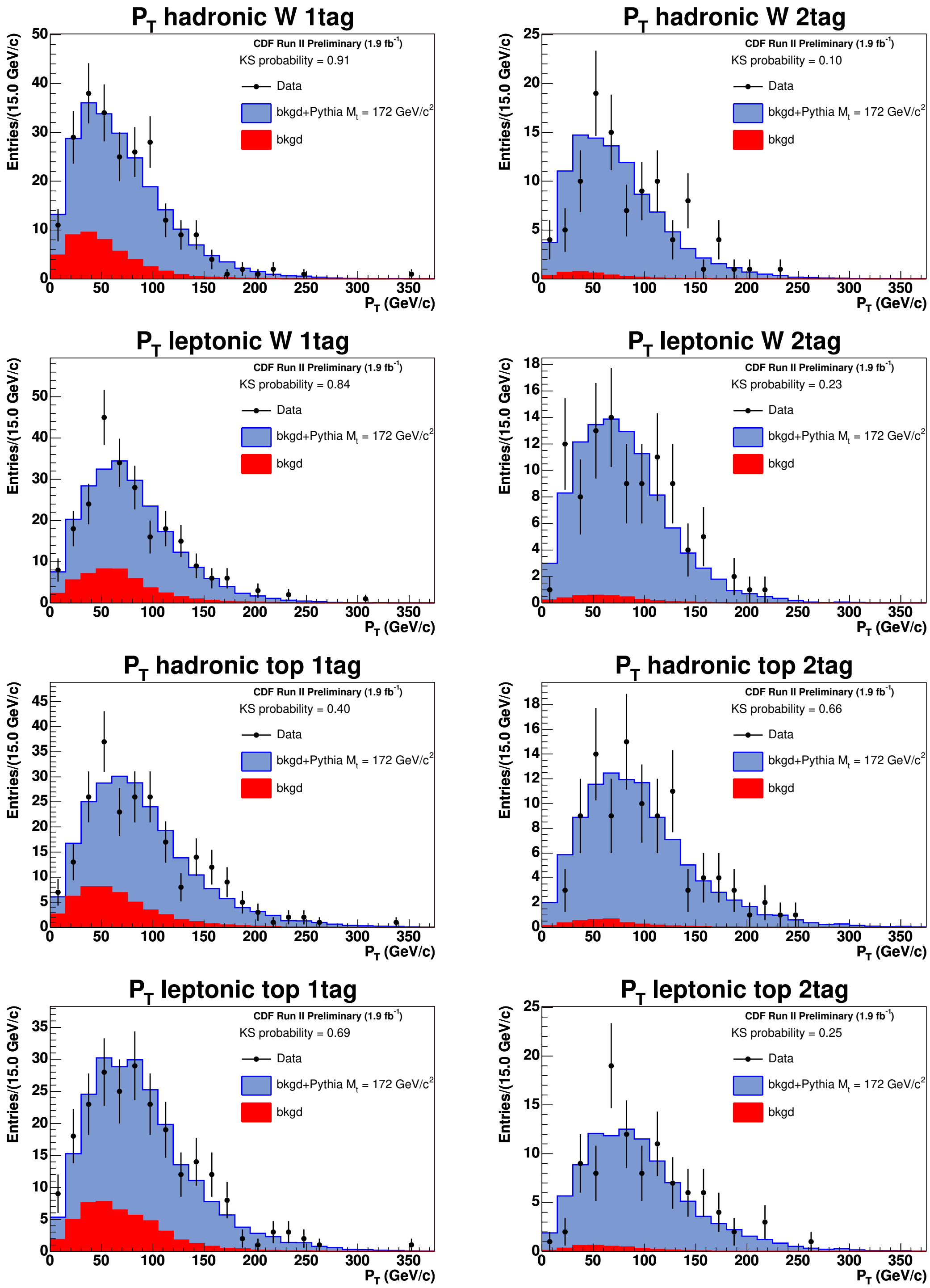

Figure 12.3: Distributions for fitter output (2).

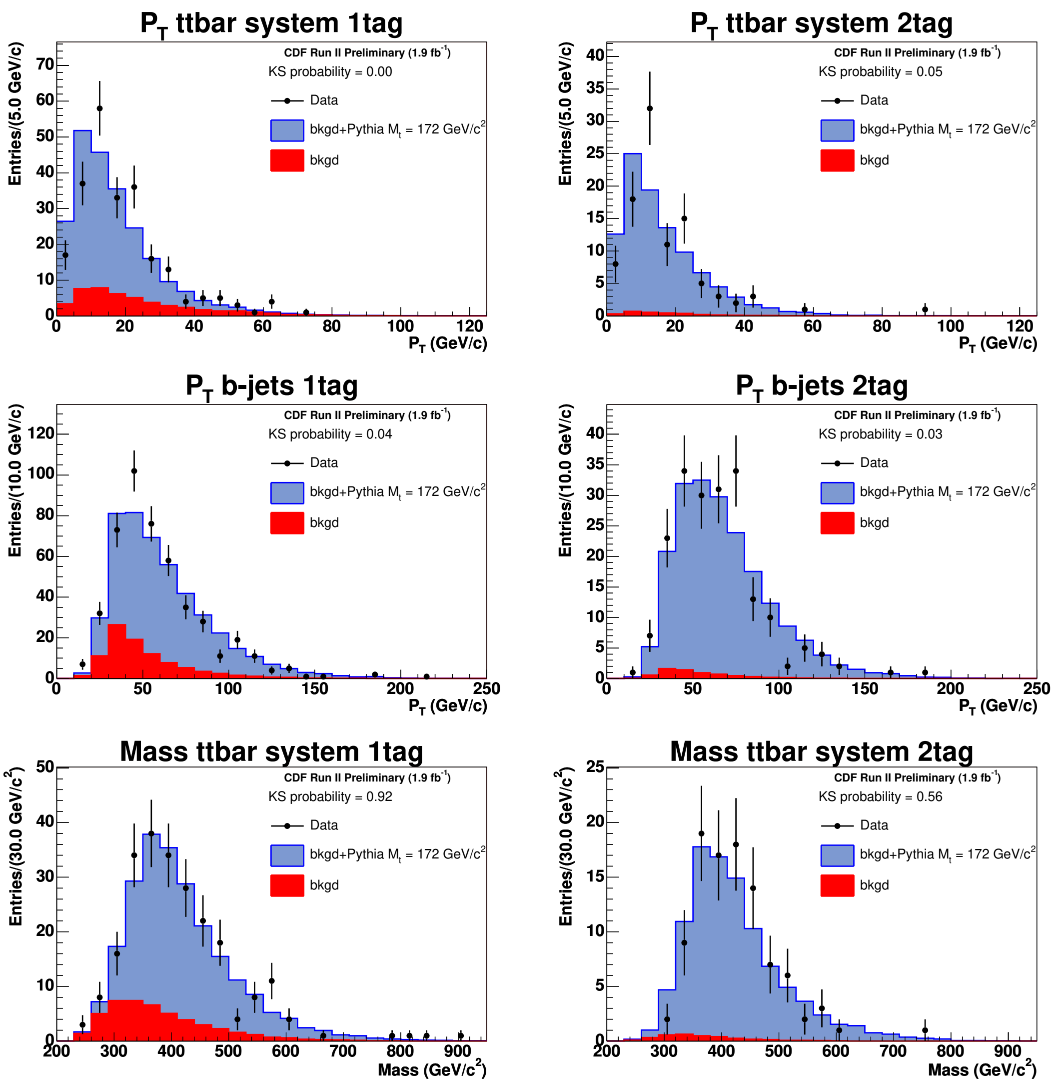

Figure 12.4: Distributions for fitter output (3).

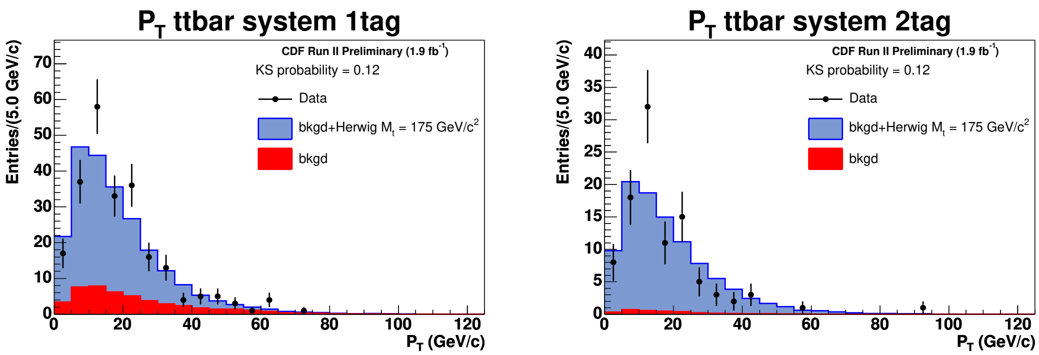

Figure 12.5: Comparison of the p_T of the $t\bar{t}$ system using HERWIG at a mass of 175.0 GeV/c^2 to model the signal.

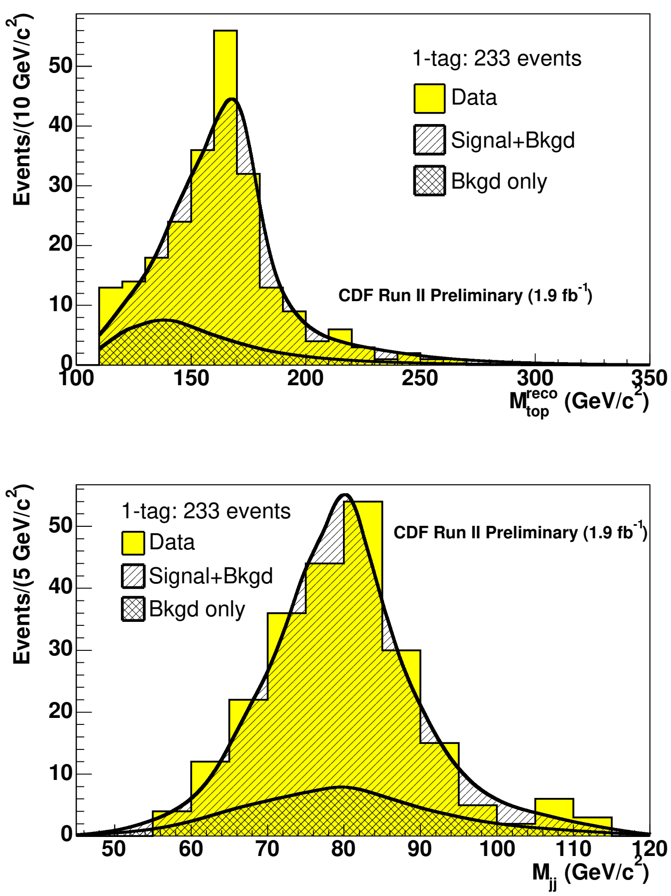

Figure 12.6: One-dimensional 1-tag data templates with PDFs from M_{top} = 172.0 GeV/c^2 and full background models overlaid. The nominal JES is used. The expected numbers of events are set to the values from the constrained fit.

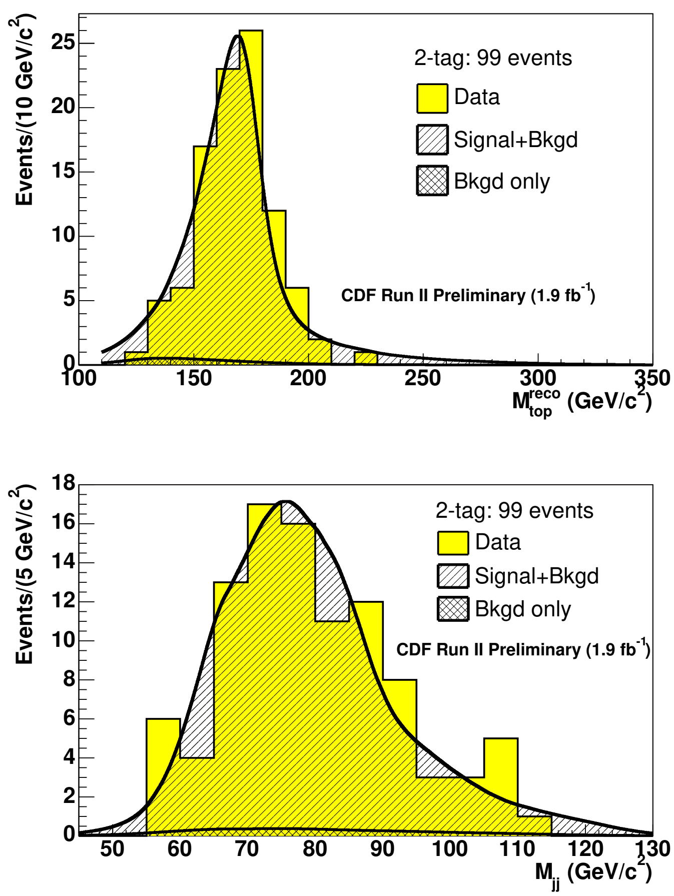

Figure 12.7: One-dimensional 2-tag data templates with PDFs from $M_{top} = 172.0$ GeV/c^2 and full background models overlaid. The nominal JES is used. The expected numbers of events are set to the values from the constrained fit.

CHAPTER 13

CONCLUSIONS

We present a measurement of the mass of the top quark in the Lepton+Jets channel at $\sqrt{s}$ =1.96 TeV. In 1.9 fb^{-1} of data collected using the CDF detector in Run II at the Tevatron, the top quark mass is measured to be $\mathrm{M_{top}} = 171.8\ ^{+1.9}_{-1.9}$ (stat.) $\pm$ 1.0 (syst.) GeV/c^2. This compares to the value measured in the CDF Run I top quark evidence paper [50], $\mathrm{M_{top}} = 174 \pm 10$ (stat.) $^{+13}_{-12}$ (syst.) GeV/c^2. By the end of Run II, this analysis alone should obtain better than 1% precision on the top quark mass, and have comparable statistical and systematic uncertainties if systematic uncertainties are not reduced.

REFERENCES

[1] W.-M. Yaho. Review of particle physics. *J. Phys.*, G 33:1, 2006.

[2] Walter R. Innes et al. Observation of structure in the υ region. *Phys. Rev. Lett.*, 39:1240, 1977.

[3] F. Abe et al. Observation of top quark production in anti-p p collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995. hep-ex/9503002.

[4] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995. hep-ex/9503003.

[5] Dhiman Chakraborty, Jacobo Konigsberg, and David L. Rainwater. Review of top quark physics. hep-ph/0303092, 2003.

[6] S. Willenbrock. Studying the top quark. *Rev. Mod. Phys.*, 72:1141–1148, 2000. hep-ph/0008189.

[7] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[8] S. Schael et al. Measurement of the W boson mass and width in e^+ e^- collisions at LEP. *Eur. Phys. J.*, C47:309–335, 2006.

[9] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys. Lett.*, B565:61–75, 2003.

[10] Peter B. Renton. Electroweak fits and constraints on the Higgs mass. 2004.

[11] P. B. Renton. Electroweak fits and the Higgs mass. *AIP Conf. Proc.*, 792:555–558, 2005.

[12] LEP electroweak working group (LEP EWWG). http://lepewwg.web.cern.ch/LEPEWWG/, April 2008.

[13] Christopher T. Hill and Stephen J. Parke. Top production: Sensitivity to new physics. *Phys. Rev.*, D49:4454–4462, 1994.

[14] Sven Heinemeyer. http://quark.phy.bnl.gov/~heinemey/uni/plots/.

[15] D. Stockinger A.M. Weber S. Heinemeyer, W. Hollik and G. Weiglein. Precise prediction for m_w in the mssm. hep-ph/0604147, 2006.

[16] W. Hollik S. Heinemeyer and G. Weiglein. Electroweak precision observables in the minimal supersymmetric standard model. hep-ph/0412214, 2004.

[17] Jorgen D'Hondt, Steven Lowette, Jan Heyninck, and Stefan Kasselmann. Light quark jet energy scale calibration using the W mass constraint in single-leptonic ttbar events. Technical Report CMS-NOTE-2006-025. CERN-CMS-NOTE-2006-025, CERN, Geneva, Jan 2006.

[18] J. Heyninck. Kinematic fit in CMS & the use of top quarks for calibration. *PoS*, TOP2006:024, 2006.

[19] Matteo Cacciari, Gennaro Corcella, and Alexander D. Mitov. Soft-gluon resummation for bottom fragmentation in top quark decay. ((u)) ((v)). *JHEP*, 12:015, 2002.

[20] Aran Garcia-Bellido. Evidence for single top quark production at D0. 2007.

[21] Makoto Kobayashi and Toshihide Maskawa. CP violation in the renormalizable theory of weak interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[22] D. Acosta et al. Measurement of B($t \rightarrow wb$)/B($t \rightarrow wq$) at the Collider Detector at Fermilab. *Phys. Rev. Lett.*, 95:102002, 2005.

[23] D. Gross and F. Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett*, 30:1343, 1973.

[24] H. D. Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett*, 30:1346, 1973.

[25] F. Abe et al. The topology of three jet events in anti-p p collisions at s**(1/2) = 1.8 TeV. *Phys. Rev.*, D45:1448–1458, 1992.

[26] A. Bhatti et al. Determination of the jet energy scale at the Collider Detector at Fermilab. *Nucl. Instr. and Meth. A*, 566:375–412, 2006.

[27] T. Affolder et al. Measurement of the t anti-t production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. D*, 64(3):032002, Jul 2001.

[28] S. Mrenna T. Sjostrand and P. Skands. Pythia 6.4 physics and manual. *JHEP*, 05:026, 2006. hep-ph/0603175.

[29] A. Abulenci et al. Top quark mass measurement using the template method in the lepton+jets channel at CDF II. *Phys. Rev. D*, 73:032003, 2006.

[30] A. Abulenci et al. Precision top quark mass measurement in the lepton+jets topology in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$TeV. *Phys. Rev. Lett.*, 96:022004, 2006.

[31] David W. Scott. *Multivariate Density Estimation: Theory, Practive and Visualization.* Wiley-Interscience, 1992.

[32] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.

[33] Kyle Cranmer. Kernel estimation in high-energy physics. *Computer Physical Communcations*, 136(3):198–207, 2001.

[34] Stefan Sperlich Wolfang Härdle, Marlene Müller and Axel Werwatz. Nonparameteric and semiparametric models. http://www.quantlet.com/mdstat/scripts/spm//html/spmhtmlnode1.html-http://www.quantlet.com/mdstat/scripts/spm//html/spmhtmlnode8.html, June 2004.

[35] J. S. Marron and D. Nolan. Canonical kernels for density estimation. *Statistics and Probability Letter*, 7(3):195–199, 1988.

[36] George R. Terrell. *Journal of the American Statistical Association*, 85(410):470–477, 1990.

[37] Ian S. Abramson. On bandwidth variation in kernel estimates - a square root law. *The annals of Statistics*, 10(4):1217–1223, 1982.

[38] M. P. Wand and M. C. Jones. *Journal of the American Statistical Association*, 88(422):520–528, 1993.

[39] S. Sain. Multivariate locally adaptive density estimation, 1999.

[40] Fulvio Piccinini Roberto Pittau Michelangelo L. Mangano, Mauro Moretti and Antonello Polosa. Alpgen, a generator for hard multiparton processes in hadronic collisions. hep-ph/0206293.

[41] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, and Michele Treccani. Matching matrix elements and shower evolution for top-quark production in hadronic collisions. *JHEP*, 01:013, 2007.

[42] Stefan Hoche et al. Matching parton showers and matrix elements. 2006.

[43] Scott Stuart Snyder. *Measurement of the Top Quark Mass at D0*. PhD thesis, SUNY Stony Brook, Stony Brook, NY, 1995.

[44] Clive Loader. *Local Regression and Likelihood*. Springer, 1999.

[45] Bradley Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7:1–26, 1979.

[46] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap: Monographs on Statistics and Applied Probability 57*. Chapman and Hall, 1993.

[47] W. J. Stirling R. S. Thorne A. D. Martin, R. G. Roberts. Parton distributions: a new global analysis. hep-ph/9803445, 1998.

[48] J. Huston H. L. Lai P. Nadolsky W. K. Tung J. Pumplin, D.R. Stump. New generation of parton distributions with uncertainties from global QCD analysis. hep-ph/0201195, 2002.

[49] G. Marchesini S. Moretti K. Odagiri P. Richardson M.H. Seymour G. Corcella, I.G. Knowles and B.R. Webber. Herwig 6.5 release note. hep-ph/0210213.

[50] F. Abe et al. Evidence for top quark production in ppbar collisions at s**(1/2) = 1.8 TeV. *Phys. Rev. Lett.*, 73(2):225–231, 1994.