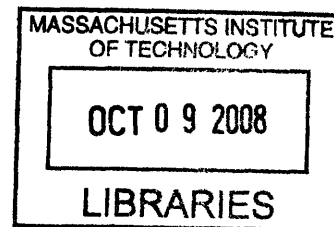# Model Independent Search For New Physics At The Tevatron

by

Georgios Choudalakis

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of   **ARCHIVES**

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

April 2008

Author . . . .
Department of Physics
April 24, 2008

Certified by. . . . .
Peter Fisher
Professor of Physics
Thesis Supervisor

Accepted by . . . .
/ Thomas Greytak
Associate Department Head for Education

Bruce Knuteson supervised much of the work.

Model Independent Search For New Physics At The Tevatron

by

Georgios Choudalakis

Submitted to the Department of Physics
on April 24, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

The Standard Model of elementary particles can not be the final theory. There are theoretical reasons to expect the appearance of new physics, possibly at the energy scale of few TeV. Several possible theories of new physics have been proposed, each with unknown probability to be confirmed. Instead of arbitrarily choosing to examine one of those theories, this thesis is about searching for any sign of new physics in a model-independent way. This search is performed at the Collider Detector at Fermilab (CDF).

The Standard Model prediction is implemented in all final states simultaneously, and an array of statistical probes is employed to search for significant discrepancies between data and prediction. The probes are sensitive to overall population discrepancies, shape disagreements in distributions of kinematic quantities of final particles, excesses of events of large total transverse momentum, and local excesses of data expected from resonances due to new massive particles.

The result of this search, first in 1 fb^{-1} and then in 2 fb^{-1}, is null, namely no considerable evidence of new physics was found.

Thesis Supervisor: Peter Fisher
Title: Professor of Physics

Στους γονείς μου, Χρήστο και Ευαγγελία, στην αδελφή μου Σοφία, και στον Επίκουρο.

To my parents, Christos and Evangelia, my sister Sophia, and to Epicurus.

Acknowledgments

.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Standard Model

Our current understanding of nature on its most fundamental level is encoded in the "Standard Model" of elementary particles.

The building blocks of matter are categorized into three families of fermions and four gauge bosons, shown in Fig. 1-1.

The Standard Model is a local gauge invariant quantum field theory, which describes electromagnetic, weak and strong interactions. Interactions are introduced for free with the assumption that nature is symmetric under local gauge transformations of the $U(1)_Y \times SU(2)_L \times SU(3)_c$ group [1]. Electromagnetic and weak interactions are aspects of a unified electroweak interaction, which are distinguishable in result of electroweak symmetry breaking via the Higgs mechanism. Elementary particles acquire bare mass by coupling to the same Higgs field that is responsible for the electroweak symmetry breaking. The success of this model of electroweak interactions in describing experimental data from the last 35 years builds confidence in the existence of the Higgs boson, though it has not been directly observed as of today.

The Standard Model carries 26 free parameters, which are determined experimentally. Depending on how one counts, they are the 6 lepton masses, the 6 quark masses, 4 parameters from CKM plus 4 from PMNS matrix, the strong coupling α_s, the QCD angle θ_{QCD}, the electromagnetic coupling α, Weinberg angle θ_w, the vacuum
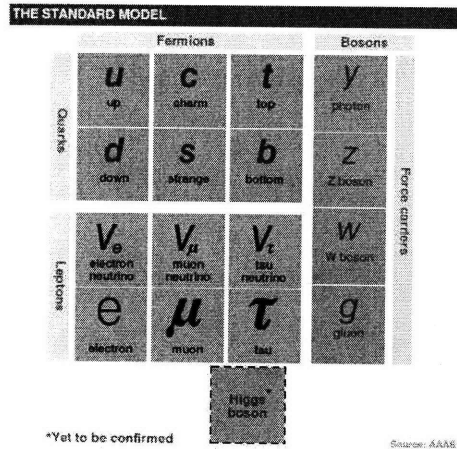
Figure 1-1: Elementary particles in the Standard Model.

expectation value (v) and the mass (m_H) of the Higgs.

The success of the Standard Model is certainly among the greatest achievements in physics. At the same time, it is bound to not be the final theory. Some reasons are explained in Section 1.1.1.

1.1.1 Limitations

The most obvious shortcoming of the Standard Model, as it stands, is that it does not describe gravity [2, 3]. Its domain is limited to energies much smaller than Planck mass (M_{Pl}), where from dimensional analysis gravity is expected to be comparable to the other three known interactions.

Another nuisance is the presence of 26 free parameters. Past successful theories have established in our minds some notion of scientific aesthetics, according to which the fundamental theory should be able to derive, from first principles, numbers such as the mass of the electron, or the amount of CP violation observed in systems like K^0 and B^0 mesons. Otherwise one can not claim to understand those effects. Grand Unification Theories try to address these issues by embedding the Standard Model into larger symmetry groups (Sec. 1.2.1).

There is overwhelming evidence (from observations of the cosmic microwave background radiation, galaxy rotations, gravitational lensing, spectroscopy of clusters and super-novae) that dark matter and dark energy dominate the mass-energy density of

the universe [4]. Currently, the Standard Model fails to provide a good candidate for either.

Another puzzle is the so-called "hierarchy problem", namely why the electroweak symmetry is broken at energy $\lesssim 1$ TeV, so much smaller than M_{Pl}, where gravity becomes significant. Theories involving extra dimensions propose some answers (Sec. 1.2.3).

Related to hierarchy is the the problem of "naturalness" in the Standard Model. A small parameter in a theory is "natural" when setting it to zero increases some symmetry of the theory, therefore its smallness can be attributed to that very symmetry. For instance, the masslessness of a vector field such as the photon can be related to the gauge invariance of the theory. However, for a scalar field, such as the Standard Model Higgs, no symmetry is there to protect its mass from acquiring quadratically divergent corrections at the loop level (Fig. 1-3), unless the theory is highly fine-tuned (Fig. 1-2). The required precision of fine-tuning depends on how far one wishes to extend the validity of the Standard Model. If one wishes it account for loop corrections up to the Planck scale, while keeping the Higgs lighter than 1 TeV, as required by electroweak measurements, then the required fine-tuning is so precise that it seems unnatural (hence the connection between naturalness and hierarchy). A solution to this can be either to abandon the concept of fundamental scalars, as in technicolor models (Sec. 1.2.4), or to search for a theory where quadratic divergences cancel, as in Supersymmetry (Sec. 1.2.2).

1.2 Beyond the Standard Model

Let me summarize the main proposals which address the limitations explained in Sec. 1.1.1, and what observable implications each suggests.

1.2.1 Grand Unification

The motivation behind Grand Unification Theories (GUTs) [6, 7] are questions such as "why protons and electrons have exactly opposite charge", or "why have three
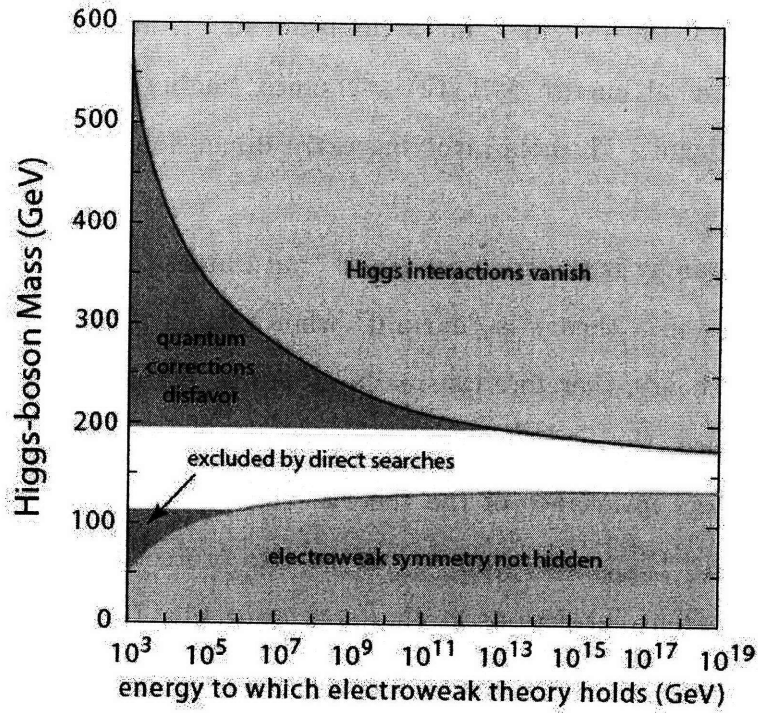
Figure 1-2: For any m_H there is an energy scale at which the Standard Model stops making sense. This happens in two ways [5]: In one case the Higgs potential runs to $-\infty$ resulting in a trivial theory without Higgs interactions. In the other case the Higgs potential has its minimum at 0, resulting in zero vacuum expectation value for Higgs, namely no electroweak symmetry breaking.
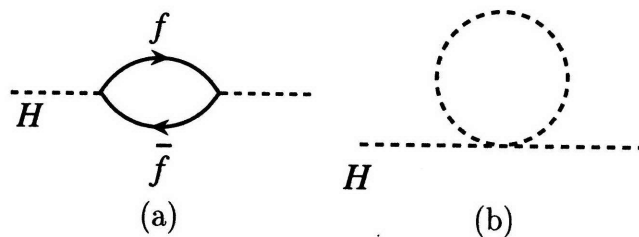
Figure 1-3: Quantum corrections to the Higgs m_H^2, through fermion loops (a) and Higgs's self-coupling (b).
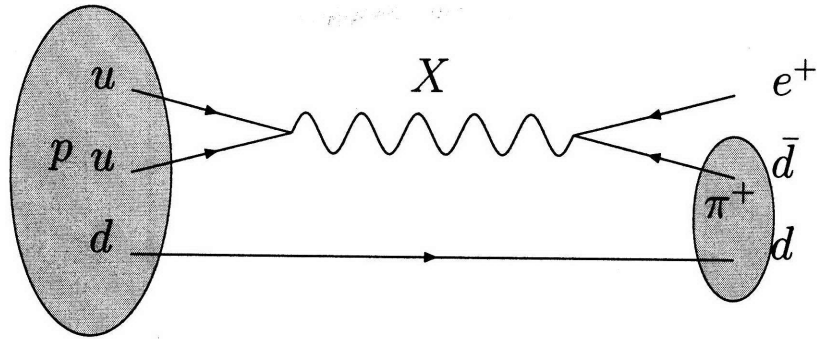
Figure 1-4: Diagram leading to proton decay in the context of $SU(5)$ Grand Unification.

generations of fermions and three interactions". These questions could become less thorny if instead of many we had just one symmetry group, which would make all particles look like components of just one particle, and all interactions like aspects of one force. Such a theory wouldn't only satisfy common taste, but more importantly could derive from mathematical principles the values of some constants, such as $\sin \theta_w$, which would be a significant advancement in our understanding nature from a reductionist's point of view.

Several Lie algebras have been studied; notably $SU(5)$, $SO(10)$, E_6 and more [2, 3]. Phenomenology varies significantly depending on the assummed symmetry. An effect predicted typically is proton decay, as new gauge bosons such as the one in Fig. 1.2.1, are predicted in breaking these hyper-symmetries at some large energy, typically $M_{\mathrm{GUT}} \simeq 10^{16}$ GeV.

1.2.2 Supersymmetry

Supersymmetric theories take the approach of solving the problem of naturalness (Sec. 1.1.1), by having a bosonic loop for each fermionic one, thus canceling out the quadratically divergent loop corrections.

SUSY introduces boson partners to Standard Model fermions, and fermion partners to gauge bosons. It introduces operators which transform fields into "superpartners" which differ from the original particles by half a unit of spin [8]. The superpartners of gauge bosons are called "gauginos", those of leptons "sleptons" and

23

Particle	Spin	Superpartner	Spin
Photon	1	Photino	1/2
Gluon	1	Gluino	1/2
W	1	Wino	1/2
Z^0	1	Zino	1/2
H	0	Higgsino	1/2
Graviton	2	Gravitino	3/2
Electron	1/2	Selectron	0
Muon	1/2	Smuon	0
Tau	1/2	Stau	0
Neutrino	1/2	Sneutrino	0
Quark	1/2	Squark	0

Table 1.1: Ordinary particles and their superpartners.

those of quarks "squarks" (Table 1.1).

SUSY can have additional favorable features, which increase interest in it. With the extra assumption of a conserved multiplicative quantum number (R-parity), which is +1 for ordinary particles and -1 for superpartners, the lightest superpartner becomes stable, serving as a cold dark matter candidate [9]. Furthermore, a theory of local supersymmetry should lead to invariance under general coordinate transformations, which may be the road to incorporating General Relativity into the Standard Model. Finally, SUSY can affect the running of couplings to make them exactly equal at some energy, in compliance with Grand Unification Theories.

If supersymmetry were exact, then each Standard Model particle would have a superpartner of equal mass. Since this is not observed, SUSY has to be broken at some energy scale [3]. It is non-trivial to construct models where SUSY is broken in ways that avoid contradicting observation, and simultaneously do not destroy its desirable features.

Higgs mass is predicted to be of order 10^2 GeV/c^2, so for SUSY to secure it from divergences it has to be introduced at energy $\lesssim 1$ TeV. That happens to be also the energy scale where it needs to be introduced in order to equalize couplings at the scale of 10^{16} GeV, associated with Grand Unification. These elements hint that, if SUSY is a correct theory, it may be within reach for current experiments.

Most SUSY signatures involve large missing energy accompanied by multiple lep-

tons and jets. Missing energy would be the effect of stable and elusive superpartners, while jets and leptons would result from long decay chains of unstable ones.

1.2.3 Extra Dimensions

Theories of extra dimensions are motivated by the hierarchy problem.

One hypothesis is that of large extra dimensions, where the known 4 dimensions, i.e. our "brane", are embedded in a manifold of higher dimensionality, and gravity only appears to be feeble because part of it is projected onto our brane, while the rest propagates in the extra dimensions, often referred to as "the bulk". By adjusting the number of extra dimensions and their radius of curvature, one can make gravity appear significant at M_{Pl} and still lower its natural scale down to the electroweak scale [10].

Theories with universal extra dimensions exist too, where fermions and/or gauge bosons also propagate in the bulk [11].

Other theories assume wrapped extra dimensions. Hierarchy then emerges by exploiting the metric of the bulk space itself. For example, with one wrapped extra dimension periodically bounded by two 3-dimensional branes, Einstein's equations result in an anti de Sitter metric, whose exponential factor makes gravity appear feeble on one of the 3-branes, where the Standard Model fields are supposed to be confined [12].

If at small distances gravity is not as feeble as suggested macroscopically by M_{Pl}, then collider experiments could reveal the coupling of gravitons. For example, a signature could be $p\bar{p} \to gG_n$, i.e. mono-jet events with large missing energy due to the graviton G_n escaping in the bulk (Fig. 1-5). Another signature of the graviton could be the Standard-Model-forbidden $gg \to G_n \to \ell^+\ell^-$ [3]. In the case of universal extra dimensions one may observe the Kaluza-Klein higher states of fermions and bosons, through $Z' \to t\bar{t}$ for instance.
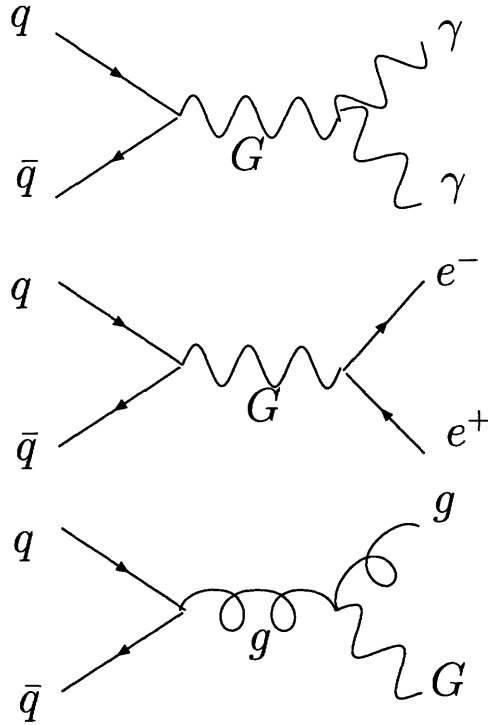
Figure 1-5: Possible signatures of graviton.

1.2.4 Technicolor

An alternative approach to electroweak symmetry breaking, which avoids the introduction of fundamental scalar fields, is new strong dynamics. With the introduction of a new non-abelian gauge symmetry and additional fermions (*"technifermions"*) which have this new interaction, it becomes possible to form a technifermion condensate that can break the chiral symmetry of fermions, in a way analogous to QCD where the $q\bar{q}$ condensate breaks the approximate $SU(2) \times SU(2)$ symmetry down to $SU(2)_{isospin}$. The breaking of global chiral symmetries implies the existence of Goldstone bosons, the "technipions" (π_T), in analogy with QCD pions. Three of the Goldstone bosons are absorbed through the Higgs mechanism to become the longitudinal components of the W and Z, which then acquire mass proportional to the technipion decay constant.

Experimental signatures of technicolor are model dependent. For example, they can be the resonance of a Standard Model gauge boson into an excited technivector

(a)   (b)
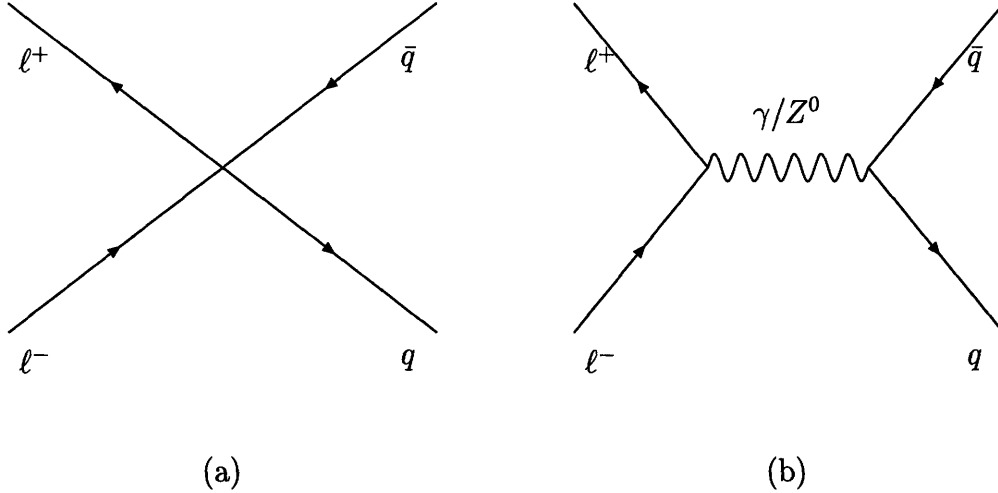
Figure 1-6: (a) Contact interaction allowed in the case of compositeness. (b) Tree-level SM diagram with the same initial and final state, where the interaction is mediated by a gauge boson.

meson, like a technirho (ρ_T), which subsequently decays into W and π_T, with π_T possibly decaying to regular quarks [3]. For example, assuming that π_T couples preferably to the third generation, such a process could be $\rho_T^{\pm} \to W^{\pm}\pi_T^0 \to \ell^{\pm}\nu b\bar{b}$, or $\rho_T^0 \to W^+\pi_T^- \to \ell^+\nu_\ell b\bar{c}$.

1.2.5   Compositeness

Compositeness is the idea that the Higgs and possibly other bosons and fermions contain substructure. Compositeness addresses the problem or naturalness similarly with technicolor, namely by avoiding the assumption of a fundamental scalar particle.

If quarks and leptons are not elementary, then they are predicted to have excited states (q^*, ℓ^*). For example, excited leptons could appear via $\ell^* \to \ell\gamma$ or $\ell^* \to W\nu$..

More importantly, if quarks and leptons have structure, new interactions should appear between them at the energy scale of their binding energy. They would be contact interactions, allowing processes such as $\ell^+\ell^- \to \ell^+\ell^-$ and $\ell^+\ell^- \to q\bar{q}$ to occur in ways additional to those of the SM (Fig. 1-6) [13, 3].

27

1.3 Current standpoint - Motivation

In 1995, the discovery of the top quark was announced [14], leaving Higgs as the only unobserved Standard Model particle. We now enter the Large Hadron Collider (LHC) era with some confidence that the Higgs will be observed to complete the Standard Model pantheon of particles. At the same time, there is hope that even what has to lie beyond the Standard Model will be revealed soon. If such a groundbreaking discovery is made, it will be different from the top quark or even a possible Higgs discovery, in the sense that it will signify the opening to a new continent of unexplored physics.

Nature has proven its capacity to surprise us. There are many ideas of what the new physics may be, but there is no need for any of them to be right. So, especially in this historical time when we expect to overcome the current impasse, it makes sense to search for any sign of discrepancy between the data and the Standard Model, without introducing any bias in what it may look like. This is the motivation behind performing a model-independent and global search.

Tevatron stands at the current high energy frontier, producing $p\bar{p}$ collisions at energy 1.96 TeV and constantly increasing luminosity. Although the size and reach of the Tevatron are inferior to those of LHC, there is still a window of opportunity in the former, until the latter has collected data and understood systematic effects specific to it. It would be undesirable to discover something at the LHC and then look back only to realize that it had been overlooked at the Tevatron. On the other hand, performing a global, model-independent analysis of the Tevatron data has the potential of revealing evidence of new physics that can be cross-checked at the LHC. This hope motivates the present work.

Chapter 2

Experimental apparatus

The present search for new physics is performed in data collected with Collider Detector at Fermilab (CDF), a general scope detector for particles generated at high energy $p\bar{p}$ collisions produced by the Tevatron accelerator. Tevatron and the Fermi National Accelerator Laboratory (FNAL) are shown in Fig. 2-1.

This chapter describes the production of $p\bar{p}$ collisions and the CDF detector. For the many acronyms used, please consult Appendix D.

2.1 Beam Production

Either due to CP violation or some other unknown reason, free protons outnumber antiprotons, which makes it easier to obtain the former, and use them to generate the latter. In this section, the procedure leading to the production of the p and $\bar{p}$ beams is outlined.

2.1.1 p Source

The production starts with storing hydrogen gas (H_2) in a **Cockroft-Walton** chamber [15], in which a 750 kV DC voltage causes electric discharges which produce negative hydrogen ions (H^-). The H^- are separated from the rest of the gas by use of a magnetic transport system and are channeled to the Linac.
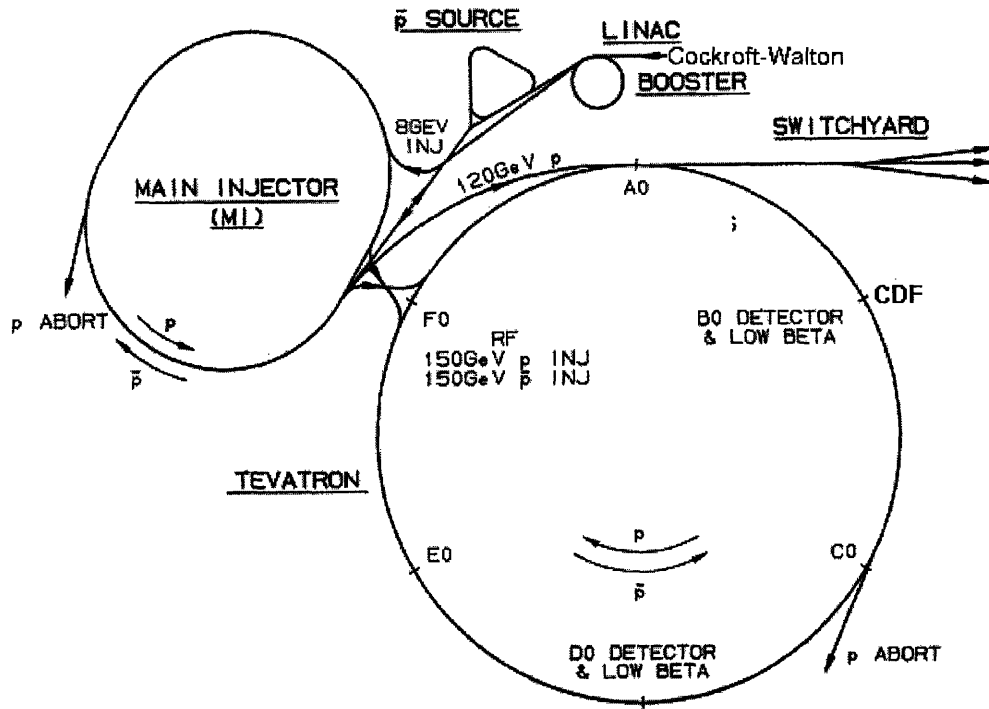
Figure 2-1: Sketch of the FNAL accelerator complex.

The **Linac** [16] is a 130 m long Alvarez linear accelerator that transfers the H^- from the Cockroft-Walton to the Booster, accelerating them from 750 keV to 400 MeV.

The **Booster** [17] is a 475 m long synchrotron that accelerates the H^- from 400 MeV to 8 GeV in just 67 ms, hence its name. One Linac load is 40 μs long and the rotation period of the beam in the Booster during injection is 2.22 μs, which means that in principle it could take $\frac{18 \times 2.22}{40} = 99.9\%$ of the Linac's load in 18 turns. Operationally however, only 5 or 6 turns get used for maximum intensity, and the rest (66.7%) of the Linac's load is dumped. At the entrance, the H^- ions pass through a carbon foil, which strips off the electrons, transforming H^- into H^+, viz. protons. It is important that the H^- pass through the carbon foil at their entrance to the ring, as they meet with the circulating H^+. This technique, named CEI, allows for higher beam brightness, avoiding limitations that would have otherwise followed from Liouville's theorem [18]. A full Booster "batch" contains a maximum of 5×10^{12} protons at 8 GeV, coalesced into 84 bunches, ready to be delivered to the Main Injector.

2.1.2 Main Injector

The **Main Injector** [19] is a 3.319 km long non-circular synchrotron, serving not only the Tevatron, but also providing protons for the production of the NuMI neutrino beam and the proton beam in the Fixed Target area. Its operations that relate to the Tevatron are:

1. $\bar{p}$ production: A single Booster batch is injected into the MI at 8 GeV. These protons are accelerated to 120 GeV and extracted in a single turn for delivery to the $\bar{p}$ production target. The produced antiprotons will eventually return to the MI for acceleration to 150 GeV, before they are delivered to the Tevatron.

2. Collider mode: Accelerate protons or antiprotons to 150 GeV and deliver them to the Tevatron.

3. End of store: Accept 150 GeV antiprotons and decelerate them to 8 GeV for storage in the Recycler.

2.1.3 $\bar{p}$ Source

At the $\bar{p}$ **production area**, the 120 GeV protons coming from the MI are directed onto a nickel target [20]. Before the collision, the bunch undergoes some modulation called *RF bunch rotation*, so as to be shorter in time and, in agreement with Liouville's theorem, contain a wider spectrum of momenta. Its being more sudden maximizes the phase-space density of antiprotons produced as secondary products of the collision with the nickel target. First, the cone of particles produced at the collision is rendered parallel by means of a lithium lens [21]. Then, a dipole magnet selects 8 GeV antiprotons, as that is the standard MI injection energy, and directs them into the Debuncher.

At the **Debuncher** [20], which is a "ring" of rounded triangular shape, the 8 GeV antiprotons are subjected to a RF bunch rotation, this time in the reverse direction, so that their beam contains a narrower spectrum of momenta and, in agreement with Liouville's theorem, spans a longer time interval. This reduction in momentum

spread is done to improve the Debuncher-to-Accumulator transfer, because of the limited momentum aperture of the Accumulator at injection. The Debuncher makes use of the time between MI cycles to reduce the beam transverse size and longitudinal momentum spread through betatron and momentum stochastic cooling respectively. This further improves the efficiency of the Debuncher-to-Accumulator transfer.

The **Accumulator** [20] is a rounded triangular "ring", similar to the Debuncher. The reason for that is that it also applies stochastic cooling to the $\bar{p}$ beam, which requires linear segments along the ring to accommodate pickups and kickers. The main purpose of the Accumulator is to hold antiprotons until they are needed by the Tevatron. The antiprotons are stored in the Accumulator for hours or days, while they augment as more are produced at the nickel target. When a new pulse of antiprotons enters the Accumulator, it circulates along a trajectory of greater "radius" than the antiprotons that have already been cooled down. The RF decelerates the recently injected pulses of antiprotons from the injection energy to the edge of the stack tail. The stack tail momentum cooling system sweeps the beam deposited by the RF away from the edge of the tail and decelerates it towards the dense portion of the stack, known as the core. Additional cooling systems keep the antiprotons in the core at the desired momentum and minimize the transverse beam size.

There is yet another ring, the **Recycler** [22], which has a role similar to that of the Accumulator. It is a 3.3 km long ring along the MI, being therefore much longer than the Accumulator, which means that if the Accumulator is getting full it can use the Recycler to hold some antiprotons too. Spread over a longer ring, the antiprotons in the Recycler are easier to maintain stable, since the beam is less dense and the dispersive forces weaker. In addition to being longer, the Recycler employs the electron cooling method to reduce the momentum spread of the antiprotons. Electron cooling is a more modern technique than stochastic cooling, in which a cold (small momentum spread) beam of electrons travels parallel to the hot antiproton beam, serving as a heat sink, where the heat of the antiproton beam is dumped, since the two beams interact electromagnetically and from thermodynamics it is known that heat goes from the hotter system to the cooler. Once the electron beam heats

up, it is discarded for a new, cold electron beam to take over. The Recycler does not only accept antiprotons that the Accumulator can not hold, but also those that the Tevatron does not need any more. Since antiprotons are so hard to produce, the Recycler keeps them to be reused in the next "store", hence its name. When the stored antiprotons reach adequate quantity, the **Tevatron** is ready to start $p\bar{p}$ collisions.

2.1.4   Tevatron

For over two decades, the Tevatron [23, 3] has been the largest hadron collider, to be soon succeeded by the Large Hadron Collider (LHC) at CERN. It is a synchrotron accelerator with radius 1 km. Along its ring are 774 dipole and 216 quadrupole superconductive magnets, providing magnetic field of intensity 4.4 T. The magnets operate in superconductive state, with cooling from liquid helium.

The Tevatron receives p and $\bar{p}$ bunches from the MI, where they have been accelerated from 8 to 150 GeV. The filling takes about 30 minutes, much longer than the acceleration period that is only 86 seconds. It accelerates the p and the $\bar{p}$ beam to the energy of 980 GeV, producing head-on collisions at $\sqrt{s} = 1.96$ TeV in the reference frame of CDF [3]. The proton and antiproton beams are both separated in 3 trains, each containing 12 bunches, therefore there are 36 p and 36 $\bar{p}$ bunches traveling in opposite directions at the same energy. Each bunch is about 18 ns (57 cm) long, which is the length of one RF bucket[1] at the Tevatron. The interval between successive bunch crossings is 396 ns (21 buckets), which is of course equal to the interval between successive bunches in a train. Successive trains are separated by longer (2621 ns or 139 buckets) intervals, called *abort gaps*.

Each p and $\bar{p}$ bunch counts about 24×10^{10} and 6×10^{10} particles respectively. As of today, the beam's optical properties allow for instantaneous luminosity that is over 2×10^{32} cm^{-2}s^{-1} at CDF, and about 15% lower at DØ [24, 25].

---

[1]A RF bucket is a slot defined by the RF electromagnetic waves, in which a bunch may be accommodated.
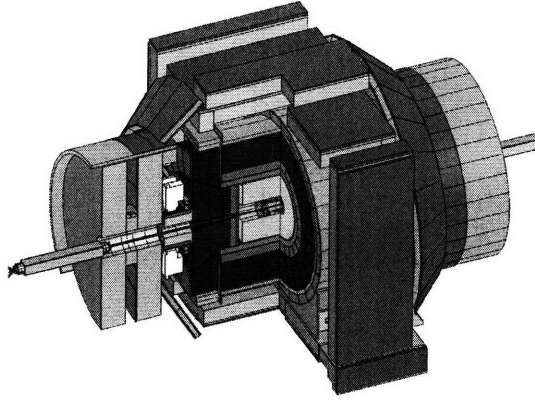
Figure 2-2: Cut-away view of the CDF detector.

2.2 The CDF detector

CDF is a ~5,000 ton detector [26] enveloping the B0 collision point of the Tevatron (Fit. 2-1). Externally, it looks forward-backward symmetric (Fig 2-2), mostly made of steel, of dimensions that are approximately 16 m × 13 m × 13 m. It is underground, shielded behind tons of concrete, which keeps it somewhat insulated from environmental sources of noise and prevents potentially hazardous radiation from leaking into its immediate surroundings. A three story building houses in its basement the detector and its assembly site, while in the superjacent levels it accommodates the data acquisition devices and the Control Room, from where operations are managed.

The CDF detector allows for a broad range of physics searches, from heavy flavor physics to searches of exotic new phenomena. It combines a variety of features, i.e. tracking, timing, calorimetry and muon detection systems, all seamed together with powerful trigger and DAQ systems.

By 1996, when the Run I period of Tevatron was over, about 90 pb^{-1} of data had been collected, in which the long-sought t-quark had eventually been discovered [14]. In preparation for the even more ambitious Run II era, which started in 2001, CDF was decisively upgraded [26], with new tracking and calorimetry capabilities and a much more efficient muon detection system. The DAQ system had to be upgraded too, to respond to the expected instantaneous luminosity of up to 5×10^{32} cm^{-2}s^{-1}. In the following sections, the current status of CDF will be described.
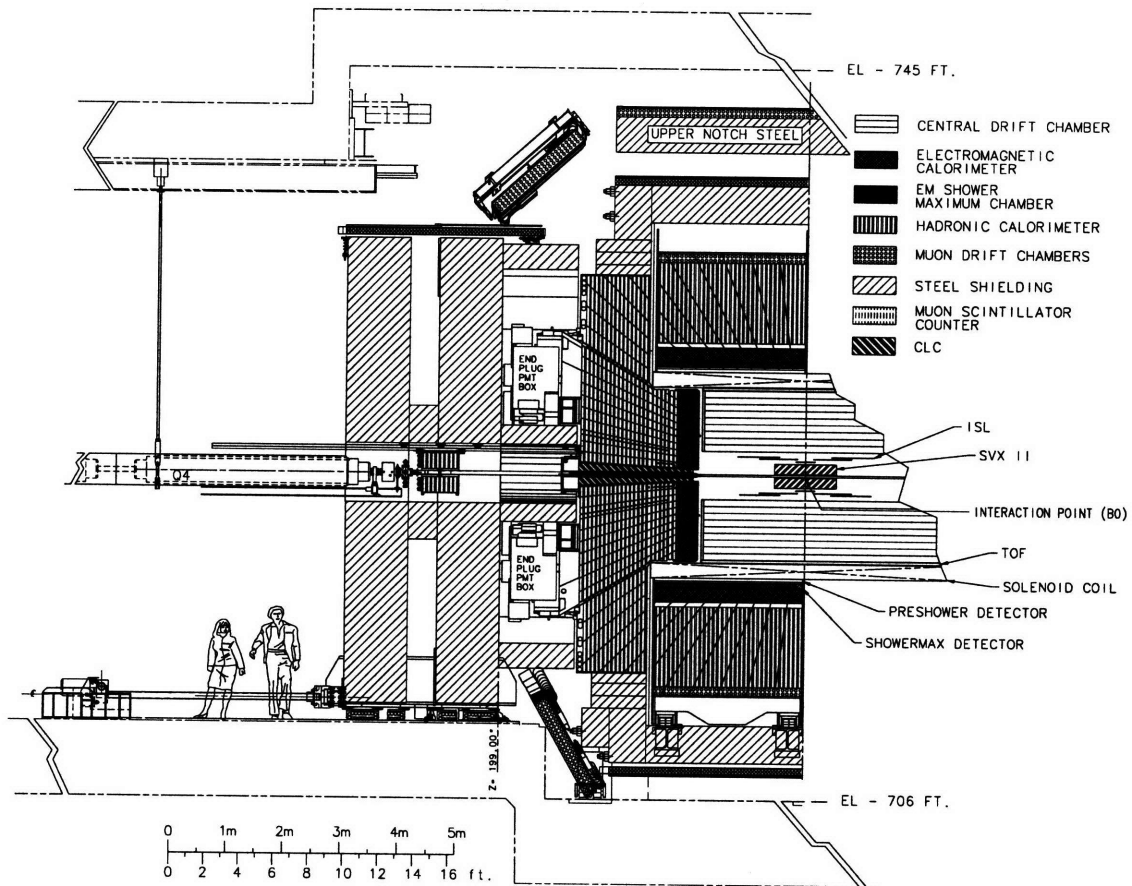
Figure 2-3: Transverse section of half of the CDF detector in Run II.

2.2.1 Coordinate Systems

Before describing the most important CDF components, it would be useful to present the established system of coordinates used at the experiment.

The Cartesian coordinate system has its axes starting at the detector's center, where the beams of p and $\bar{p}$ are supposed to collide. The y axis is defined to point vertically up, and the x to be perpendicular to the beam pipe and pointing in the direction away from the center of the Tevatron ring. In terms of $\hat{x}$ and $\hat{y}$, $\hat{z}$ is $\hat{x} \times \hat{y}$, which approximately coincides with the direction in which the p beam travels through the center of CDF.

The cylindrical coordinate system reflects the approximate axial symmetry of the tracker and the calorimeter around $\hat{z}$, which in cylindrical coordinates remains the same unit vector it was in Cartesian. The radial unit vector $\hat{r}$ at each point is

35

perpendicular to and pointing away from the z axis. The azimuthal angle ϕ is by definition 0 on the semi-infinite $z - x$ plane that contains the positive x axis and increases in the direction of $\hat{\phi} = \hat{z} \times \hat{r}$.

Spherical coordinates are used more often than the above two systems. The reason is that, to the physical event occurring in a $p\bar{p}$ scattering, the cylindrical or any other symmetry of the surrounding detector is irrelevant. The dynamics of the event recognize one special axis, viz. z, along which the p and $\bar{p}$ were traveling right before their collision. It is therefore convenient to define the angles of all outcoming particles with respect to $\hat{z}$. For any point in space, a radial unit vector $\hat{r}$ is defined to point in the direction away from the beginning of the coordinates. Also, a polar angle θ is defined, which is 0 along the positive z axis and increases in the direction of $\hat{\theta} = \hat{r} \times \frac{\hat{r} \times \hat{z}}{\sin \theta}$. Finally, the azimuthal angle ϕ is defined as in the cylindrical coordinates and increases along $\hat{\phi} = \hat{\theta} \times \hat{r}$.

Since the p and $\bar{p}$ beams are unpolarized, z has to be an axis of symmetry when examining a large set of events. In other words, based on the premise of isotropy of the universe which leaves z as the only axis special to the scattering, there can be no law of physics that would cause a non-uniform ϕ distribution of the particles coming out of the scattering.

It is common to not mention the polar angle θ per se, but instead a dimensionless quantity called "pseudorapidity", which is related to θ as

$$\eta = -\ln(\tan(\theta/2)). \qquad (2.1)$$

η is the $E \to |\vec{p}|$ limit of the quantity called "rapidity", which is[2]

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}, \qquad (2.2)$$

and has the beautiful property that for any pair of rapidities, the difference Δy is invariant under Lorentz boosts along the z axis.

---

[2]The rapidity y may not be confused with the Cartesian coordinate y.

2.2.2 Tracking

Tracking is crucial for particle identification; it has been so since the first experiments with wire and bubble chambers. Though technology has advanced, the principles remain:

- Only ionizing particles leave tracks, which distinguishes them from neutral ones.

- The curvature of a track under the influence of Lorentz force in the presence of a magnetic field $\vec{B}$ is a measure of the transverse momentum $\vec{p_T}$ of the particle, namely of the projection of its momentum $\vec{p}$ on the plane transverse to $\vec{B}$.

- The direction of the track can be used to estimate the direction (η, ϕ) in which a particle is produced.

- Being able to observe tracks improves our intuitive understanding of what particles are produced in an event. For example, the assembly of tracks within a cone is indicative of hadronic jet showers, while isolated tracks are more likely leptons[3].

- Extrapolating the tracks of an event down to their origin(s) indicates the position of the event. This can reveal the existence of displaced secondary vertices, indicative of the decay of a long-lived particle, such as a B^0 meson. It may also indicate the existence of multiple $p\bar{p}$ interactions in the same bunch crossing, by observation of multiple primary vertices in the same event.

Silicon Detector

The first tracking device particles pass through is the Silicon Detector. Silicon allows for a highly granular and radiation tolerant tracker that can survive as near as 1.5 cm from the collision point [26]. The operation principle of a silicon micro-strip is depicted in Fig. 2-4 [3, 27].

---

[3]Even though τ is a lepton, it is common to include only electrons and muons in the term "leptons", because they are easier to identify than τ which often decays hadronically, so they consist more "clear" leptons in the experimental sense.

About 722,000 read-out channels come from the Silicon Detector [28], by far more than from any other CDF component. It is separated in three subsystems: L00, SVX and ISL (Fig. 2-5, 2-6).

L00 is a single layer of single-sided silicon built directly onto the beam pipe, at 1.5 cm radius. It provides precision position measurement before the particles undergo multiple scattering.

SVX is the heart of the Silicon Detector, consisting of 12 identical wedges in ϕ. Each wedge contains 5 layers of double-sided silicon, oriented parallel to the beam pipe at radii from 2.5 to 10.6 cm. On one side, the silicon strips are aligned axially. The other side has $90°$ stereo strips for 3 of the layers, and $1.2°$ stereo strips for the remaining 2 layers. Obviously, the choice of aligning some strips non-axially was made to allow for three-dimensional track reconstruction.

The ISL envelops SVX. It carries $1.2°$ stereo double-sided silicon in a single layer for intermediate radius measurement of central[4] tracks and in two layers for tracking in the region $1 < |\eta| < 2$, which is not completely covered by the COT (Fig. 2-6).

The silicon embedded strips are 8 μm wide [29], which brings the hit's spatial resolution down to about 12 μm. This resolution makes it possible to measure the impact parameter of a track to 40 μm, with 30 μm uncertainty due to the beam width. The z_0, namely the z-coordinate of the primary vertex, can be measured with 70 μm accuracy.

Central Outer Tracker

The COT [30, 31] is a cylindrical multi-wire open-cell drift chamber surrounding the Silicon Detector (Fig. 2-6).

COT contains Argon-Ethane ($Ar - C_2H_6$) in a 1:1 mixture. When charged particles traverse the gaseous mixture they leave a trail of ionization electrons, which drift under the influence of an 1.9 kV/cm electric field. The latter is produced by field planes and homogenized by potential and shaper wires. After some time that

---

[4]Here and below the word "central" is used to describe objects with $|\eta_{det}| < 1.0$; "plug" is used to describe objects with $1.0 < |\eta_{det}| < 2.5$.
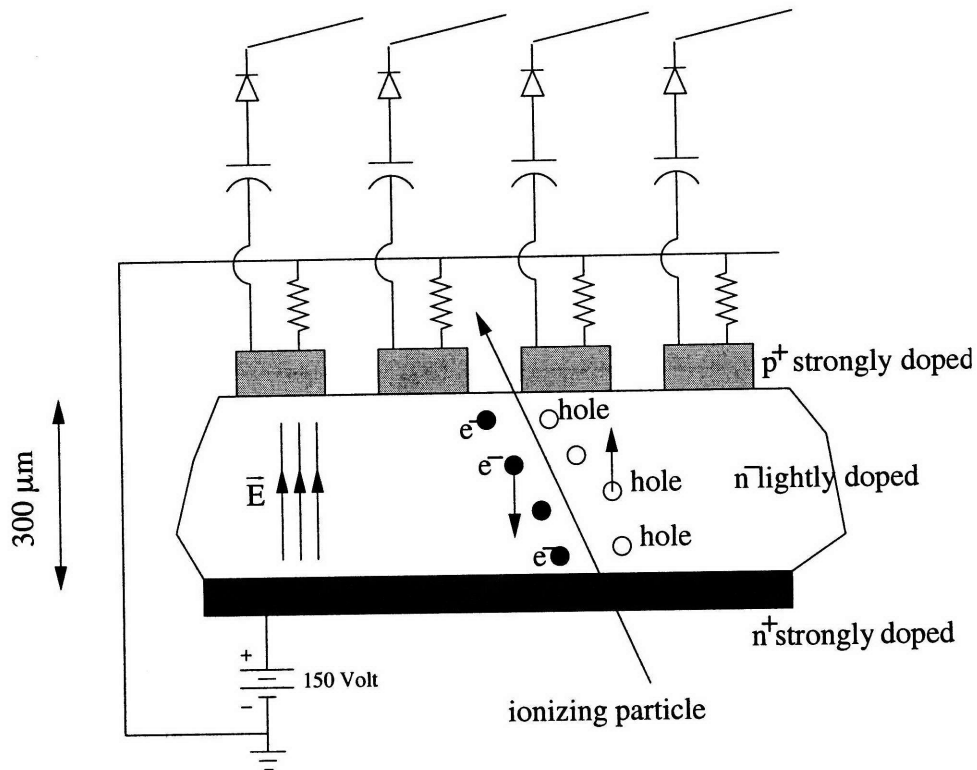
Figure 2-4: Schematic of a silicon particle sensor. An array of finely spaced p-type silicon strips is implanted in an n-type silicon substrate, typically 300 μm thick. The n-p contact is then reversely polarized, typically with a depletion voltage of 150 V. When an ionizing particle traverses the depletion zone it creates a localized stream of e^--hole pairs, which are collected by the nearest strips, where after amplification they are detected as small current signals. There are variations in the design of silicon strips, such as double-sided strips where signals are read from both sides. The spatial resolution of the most advanced silicon strip can be as fine as 2 - 4 μm, limited mostly by diffusion [3, 27].
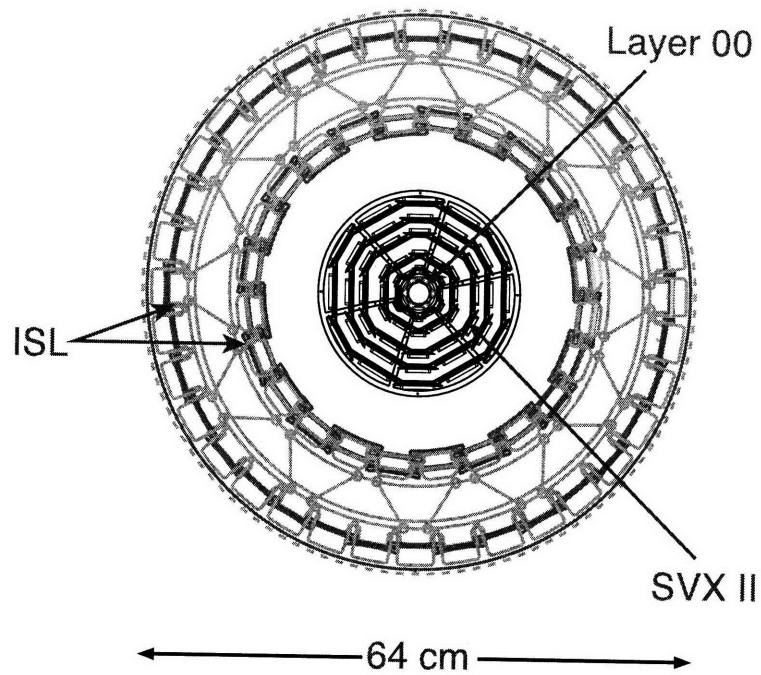
Figure 2-5: The CDF Silicon Detector (XY view) [28].
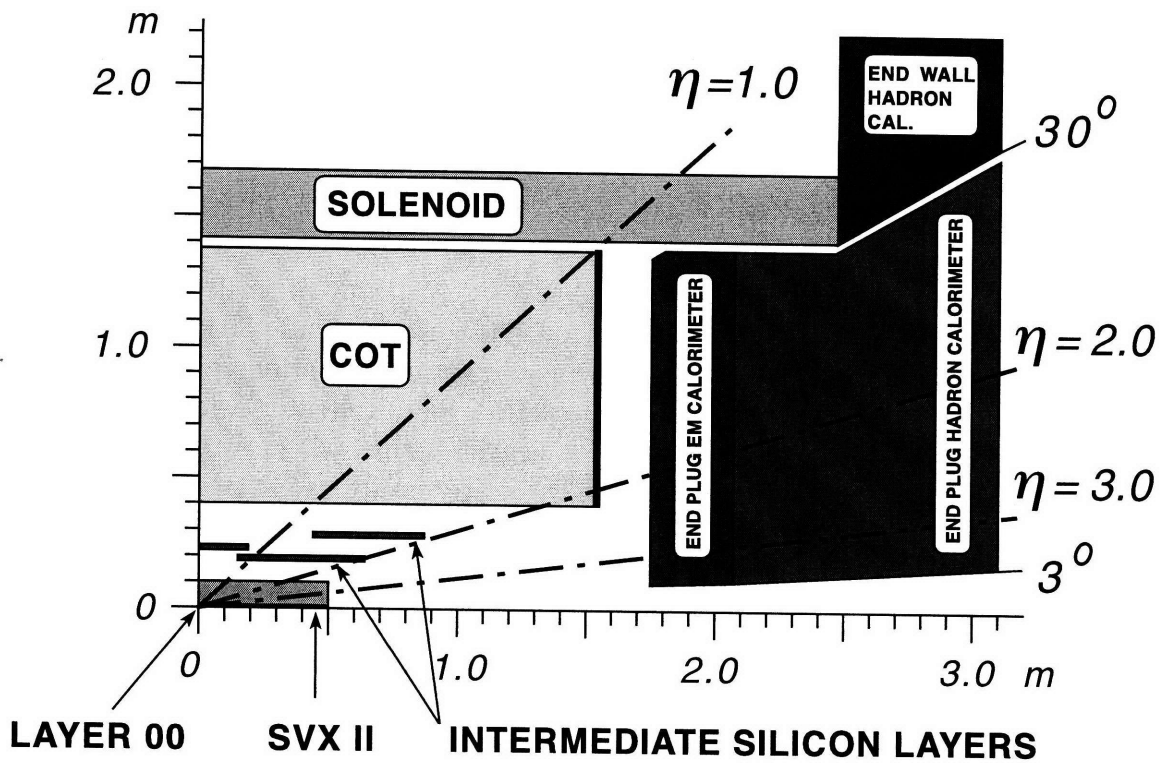
Figure 2-6: Schematic profile (RZ view) of the central part of the CDF detector [29]. The Time Of Flight detector (not shown) is between the COT and the solenoid. The central electromagnetic and hadronic calorimeter are not depicted either.
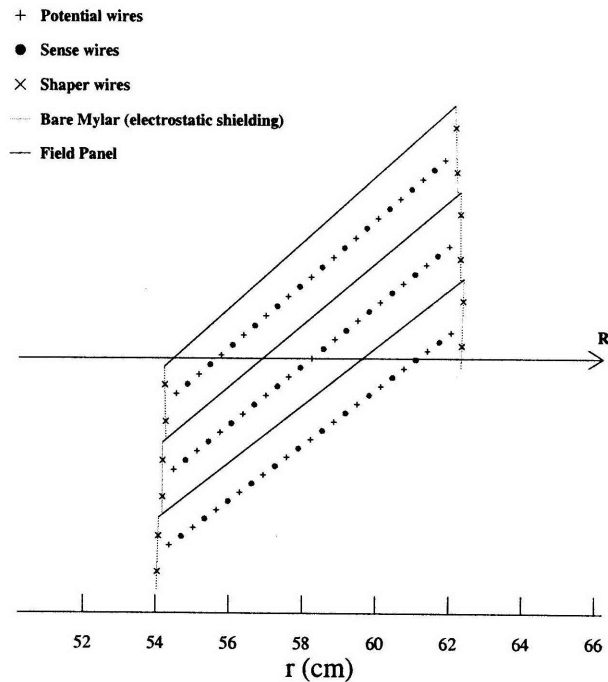
Figure 2-7: Three COT cells from the second superlayer (XY view). Their inclination with respect to the radial direction is equal to the Lorentz angle of 35° (see text).

depends on the distance they travel, the ionization electrons are collected by sense wires immersed in the gas producing a detectable[5] electric signal. The $r - \phi$ location of the track with respect to the sense wire is then estimated from the time it takes to detect the signal. The drift distance is less than 0.88 cm and is covered in less than 100 ns, which is less than the 396 ns between successive bunch crossings, therefore causes no pile-up of signals from different events.

The field panels, shape, potential and sense wires are all grouped in electrostatically shielded cells (Fig. 2-7). Each cell contains 12 sense, 13 potential and 4 shaper wires. Sense and potential wires alternate with successive sense wires being 7mm apart. Combining drift time information from several wires, the single hit resolution reduces to about 140 μm.

Cells are arranged in 8 superlayers (Fig. 2-8). The wires in the 1st and 5th superlayer are not oriented axially, but at a stereo angle of +3°. Similarly, there is a

---

[5] When an ionization electron approaches the 40μm thick sense wire it is accelerated by its rapidly increasing $(1/r)$ electric field, producing an "avalanche" of secondary ionization electrons and thus enhancing the signal.

1/6th East Endplate(s)
Units: centimeters [inches]

Optional split

+   R40.000 [15.75]   R137.000 [53.94]

Layer #	1	2	3	4	5	6	7	8
Cells	168	192	240	288	336	384	432	480
Radius (cm)	46	58	70	82	94	106	119	131
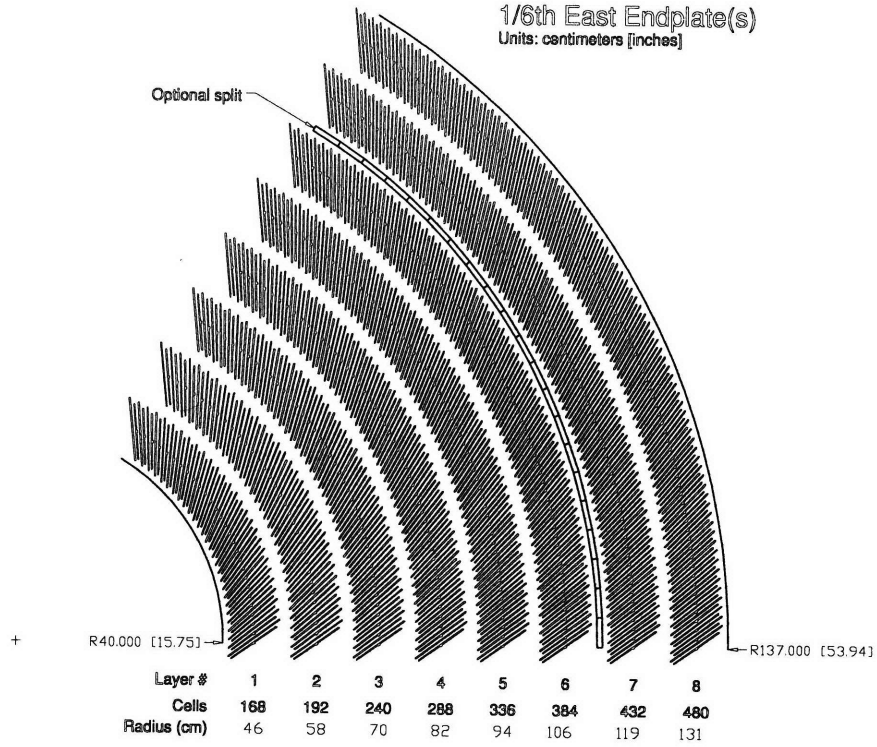
Figure 2-8: Part of the COT endplate (XY view). The wire-plane slots grouped into eight superlayers are shown.

stereo angle of $-3°$ in superlayers 3 and 7. Like in the case of the Silicon Detector, the reason that 4 out of the 8 superlayers are oriented non-axially is to allow for tracking in the three dimensions[6].

It was mentioned that ionization electrons drift under the influence of an electric field $\vec{E}$, but there is also a magnetic field $\vec{B}$ parallel to the z axis. So, as the force $-e\vec{E}$ accelerates the electron, the force $-e\vec{v} \times \vec{B}$ turns it on the $x-y$ plane (Fig. 2-9). At any time the velocity of the electron in the medium can be parametrized as $\vec{v} = \mu\vec{E}$, where μ is the *mobility* of the medium. Assuming that the $\vec{E}$ field is homogeneous on the $x-y$ plane and the electron is non-relativistic, the equilibrium is at an angle ψ with respect to $\vec{E}$ that is $\psi_L = \arctan \mu \left| \vec{B} \right|$. ψ_L is called the Lorentz angle and for the COT it is about $35°$. The wires in the COT cells are then arranged along the direction determined by the Lorentz angle, to minimize the drift time and maximize the COT efficiency and resolution (Fig. 2-7).

---

[6]If all COT wires were parallel to the z axis, then the z coordinate of hits would be unknown.
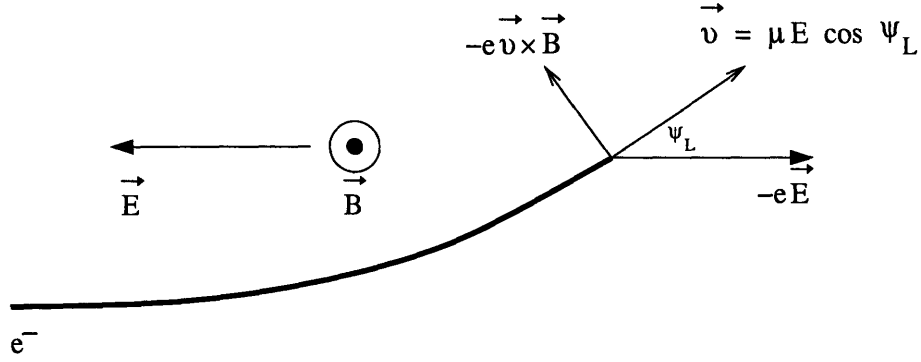
Figure 2-9: The trajectory of an ionization electron in the $\vec{E}$ and $\vec{B}$ field of the COT. The condition $eE \sin \psi_L = evB = e\mu E \cos \psi_L B$ determines the Lorentz angle $\psi_L = \arctan \mu B$.

Magnet

A 1.4 T magnetic field is produced in the $-\hat{z}$ direction by the superconductive solenoid surrounding the COT (Fig. 2-6 and 2-3).

The magnetic field is essential for the measurement of the transverse momentum (p_T) of ionizing particles. Greater magnetic field intensity and bigger tracking volume radius improve p_T resolution, which on the other hand is limited by the spatial resolution of the tracker and multiple scattering [3]. At CDF, the p_T resolution is $\delta(1/p_T) = \frac{0.15\%}{\text{GeV}/\text{c}}$.

Track reconstruction

The Silicon Detector and the COT record a large number of hits in each event, viz. discrete positions from which ionizing particles seem to have passed. But the hits alone do not suffice. In each event there are tens of charged particles, as well as false hits. What is needed is an algorithm to reconstruct tracks out of the thousands of hits of each event.

Every track is a helix that can be parametrized in terms of the variables in Table 2.1. Essentially, tracking algorithms fit for those 5 parameters to best match the observed hits [32, 33].

Tracking in the COT using the Segment Linking algorithm involves first reconstructing linear segments of the track in each of the eight superlayers [33]. Then, the

43

θ the polar angle at minimum approach, which refers to the point of the track closest to the z axis.

C semi-curvature of the track (inverse of diameter), with the same sign as the particle's electric charge.

z_0 z coordinate at minimum approach.

D signed impact parameter: distance between helix and the z axis at minimum approach. The sign of D is given from its formal definition: $D = \text{sign}(q)(\sqrt{x_0^2 + y_0^2} - \rho)$, where q is the ionizing particle's charge, (x_0, y_0) is the center of the track's projection onto the $x - y$ plane, and ρ is the radius of the same projection. Fig. 2-10 demonstrates combinations of positive and negative D and C.

ϕ_0 Direction of track on $x - y$ plane at minimum approach, i.e. the polar angle of the particle's p_T at minimum approach.

<div align="center">Table 2.1: The 5 parameters of a helical track.</div>
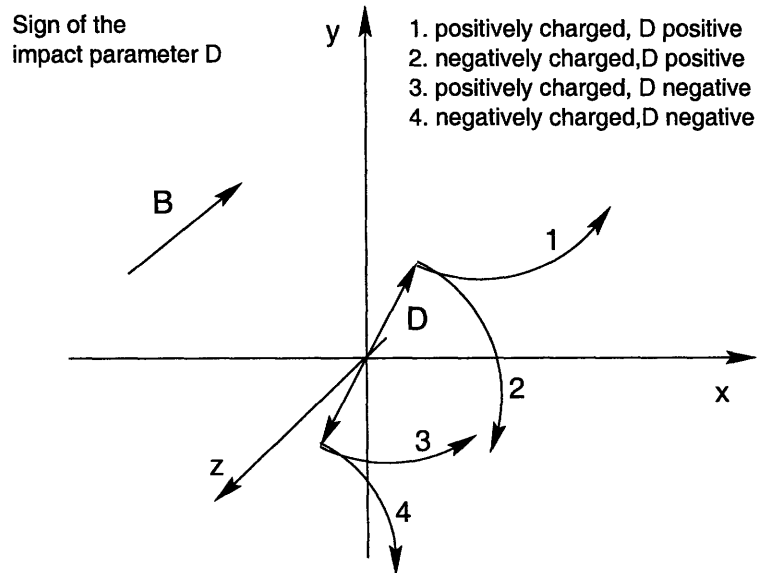
Figure 2-10: Combinations of positive and negative D and C (see Table 2.1).
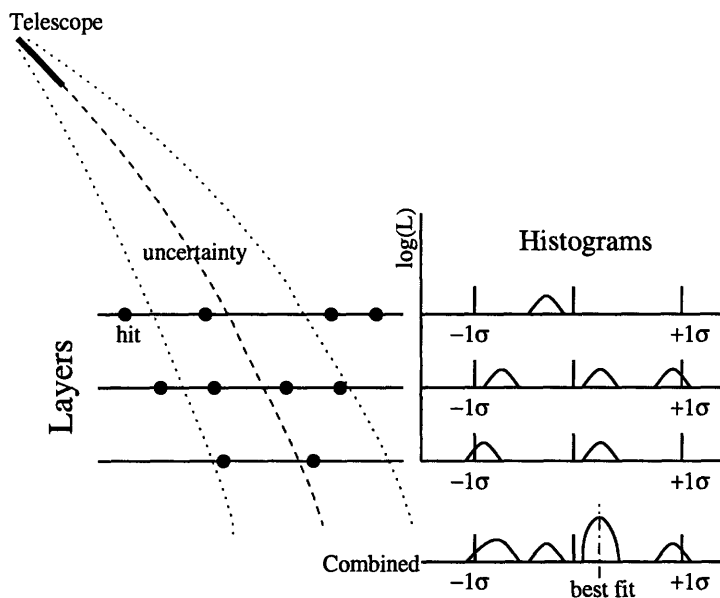
Figure 2-11: Schematic of the Histogram Tracking method.

linear segments from the axial layers are linked to form a 2D track on the $x - y$ plane, starting the extrapolation with the outmost segment as seed. The $r - z$ projection of the track is attained by linking the segments from the stereo superlayers. Eventually, the track is characterized by the χ^2 of the fit, and is only kept if that figure of merit is below threshold.

An alternative is the Histogram Tracking algorithm [33]. It starts with a coarse approximation of the final track, which is attained by extrapolating a segment of the track called "telescope", such as the outer superlayer segment. The extrapolated telescope corresponds to a helix whose parameters carry large uncertainty, therefore instead of a curve it can imagined as a tube, to visualize those uncertainties (Fig. 2-11). In each layer the tube crosses there may be hits that fall inside the tube. For those hits, the likelihood is calculated to belong to the track. Each crossed layer is translated into a histogram of those likelihoods. Those histograms coming from different layers are then combined into a final one, and the track is reconstructed as the helix which maximizes the combined likelihood. Compared to the Segment Linking algorithm, this alternative is slower but more efficient in cases of missing and accurate in cases of spurious hits.

The Histogram Tracking algorithm is also applied in Silicon tracking, where the

part of the track in the COT is used as the telescope.

In Silicon tracking [33], the information of the z of the primary vertex is used. That is known by combining hits from the stereo strips and extrapolating to the beam axis. This produces a variety of candidates, each of different likelihood, so in the end the primary vertex is at the most likely z.

The Stand-Alone algorithm for Silicon tracking uses information exclusively from silicon hits, therefore has the advantage of using the whole $|\eta| < 2$ acceptance of the Silicon Detector. It starts by finding hits in places where axial and stereo strips intersect. Then, triplets of aligned hits are identified. The information of the primary vertex is used to constrain the candidate helices. In the end the best fitting helix is kept.

The Outside-In algorithm [34] takes COT tracks and extrapolates them into the Silicon Detector, adding hits via a progressive fit. As each layer of silicon is encountered, a road size is established based on the error matrix of the track. Hits that are within the road are added to the track, and the track parameters and error matrix are refit with this new information. A new track candidate is generated for each hit in the road, and each of these new candidates are then extrapolated to the next layer in, where the process is repeated. As the extrapolation proceeds, the track error matrix is inflated to reflect the amount of scattering material encountered. At the end of this process, there may be many track candidates associated with the original COT track. The candidate that has hits in the largest number of silicon layers is chosen as the winner; if more than one candidate has the same number of hits, the χ^2 of the fit in the silicon is used to decide.

The Inside-Out algorithm [35] performs the reverse extrapolation: from the Silicon Detector to the COT. Its goal is to use the Stand-Alone silicon track to associate it with COT hits and improve the efficiency of reconstruction of tracks that do not cross more than 4 COT superlayers.

2.2.3 Calorimetry

CDF is equipped with sampling electromagnetic and hadronic calorimeters in the central and plug region, enhanced with shower maximum and preshower detectors for improved particle identification [26]. Central calorimeters cover 2π rads in ϕ (Fig. 2-2). The central electromagnetic calorimeter covers $|\eta| < 1.1$ and the hadronic $|\eta| < 1.3$. The plug calorimeters reach as far as $|\eta| = 3.6$. They are segmented in wedge-shaped towers pointing to the center of CDF. Each tower covers about 0.1 units of η and 15° in ϕ (Fig. 2-3). For increased acceptance, the hadronic calorimeter has the endwall calorimeter, spanning $30° < |90° - \theta| < 45°$ (Fig. 2-6).

Electromagnetic Calorimeter

CEM and PEM comprise lead absorber sheets alternating with scintillator layers. Light produced at the scintillator is transfered by WLS fibers to two PMTs that correspond to each tower[7].

The CEM has a total maximum thickness of about 19 X_0, in 20-30 (varying with $|\eta|$) layers of 3 mm lead and 5 mm scintillator. Its energy resolution, after *in situ* calibration, is found to be $13.5\%/\sqrt{E_T} \oplus 2\%$.

PEM contains 22 layers of lead, 4.5 mm each[8], and its scintillator layers are 4 mm thick. Its total thickness is 21 X_0. Its resolution is $16\%/\sqrt{E_T} \oplus 1\%$.

In both CEM and PEM, there is a shower maximum detector, 6 X_0 into the calorimeter, where an electromagnetic shower statistically contains the biggest number of particles [3]. CES is a multi-wire proportional chamber with strip readout in the z direction and wire along ϕ. PES has scintillator strips that cross to form a 2-dimensional grid in each plug. With resolution of about 2 mm in the central and 1 mm in the plug, the showermax detectors facilitate the matching of tracks with calorimeter hits, improving $e^\pm/\gamma$ identification. Also, sampling the profile of the electromagnetic showers at 6 X_0 allows for improved γ/π^0 identification.

---

[7]Having two PMTs per tower allows for cross-check of the validity of signals, using time information and comparing the difference in the signal intensity in the two.

[8]The first layer is an exception, being 1 cm thick and read out separately to be used as a preshower detector.

Finally, between the solenoid and the first layer of the CEM lies a set of multi-wire proportional chambers, the CPR, which samples the electromagnetic showers at 1.075 X_0, viz. the solenoid's thickness. This information greatly enhances γ and soft $e^{\pm}$ identification [26].

Hadronic Calorimeter

The hadronic calorimeter is similar to the electromagnetic, except that it uses iron for absorber instead of lead. The CHA is 4.7 λ_0 thick, consisting of 32 2.5 cm iron layers alternating with 1 cm scintillator layers. Its energy resolution is $75\%/\sqrt{E_T} \oplus 3\%$.

The WHA has similar energy resolution [36]; $75\%/\sqrt{E_T} \oplus 4\%$. It contains 15 layers of iron, 5 cm each, alternating with 1 cm layers of scintillator, adding up to 4.5 λ_0.

The PHA is thicker, containing 7 λ_0 in 23 layers of iron, 51 mm each, alternating with 6 mm layers of scintillator. Its energy resolution is $80\%/\sqrt{E_T} \oplus 5\%$.

2.2.4 Muon System

CDF is equipped with four muon detectors (Fig. 2-12), which will be described in this section.

Muons weigh 200 times more than electrons, therefore radiate about $200^2 = 40,000$ times less by bremsstrahlung. They do not deposit much energy in the calorimeter, but rather traverse the whole detector almost unimpeded. This makes them easier to identify by installing wire chambers around the detector, beyond the calorimeter and even beyond extra absorbing material; muons are virtually the only ionizing particles that can reach there.

Shielding the muon detectors behind absorber increases the detected muons' purity, but also enhances multiple scattering, which makes it harder to match the small track segment in the muon detector (called "stub") with the corresponding COT track. However this is not a very big problem, especially for high-p_T muons, since the displacement due to multiple scattering is about $\frac{15 \text{ cm}}{p_T}$, for the p_T is in GeV/c [26].
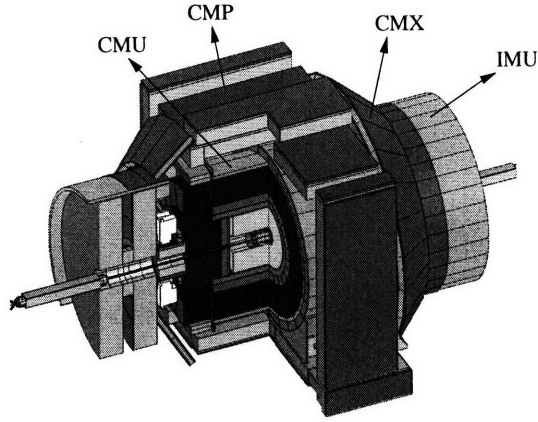
Figure 2-12: The muon detectors of CDF.

Furthermore, some low-p_T muons can not reach the muon detectors, but that is not a problem either, since the threshold is lower than 2.2 GeV/c [26], far lower than the p_T of the muons considered in this analysis.

Central Muon detector (CMU)

The CMU [26] surrounds the hadronic calorimeter, at radius 3.47 m, covering the $|\eta| <$ 0.6 region. It consists of argon-ethane wire chamber cells operating in proportional mode, organized in stacks of four. Each wire chamber is $2.7 \times 6.4 \times 226$ cm^3 with a resistive stainless steel wire along its biggest dimension, which is aligned parallel to the z axis. In ϕ it is segmented in 24 wedges, each containing 4 stacks side by side, therefore each wedge contains a chamber of $4 \times 4 = 16$ cells (Fig. 2-13).

The drift times (< 800 ns) are used to measure the $r - \phi$ projection of the track. The z coordinate of the track is extracted with about 10 cm precision, using the *charge division* method, whose principle is explained in Fig. 2-14. To apply this method, every couple of ϕ−adjacent cells have their wires ganged together at one end.

Central Muon Upgrade detector (CMP)

The CMP (Fig. 2-12) is shielded behind about 7.8 λ_0, comprising the calorimeter, the magnet return yoke and extra steel absorber. Compared to the CMU, which was

49
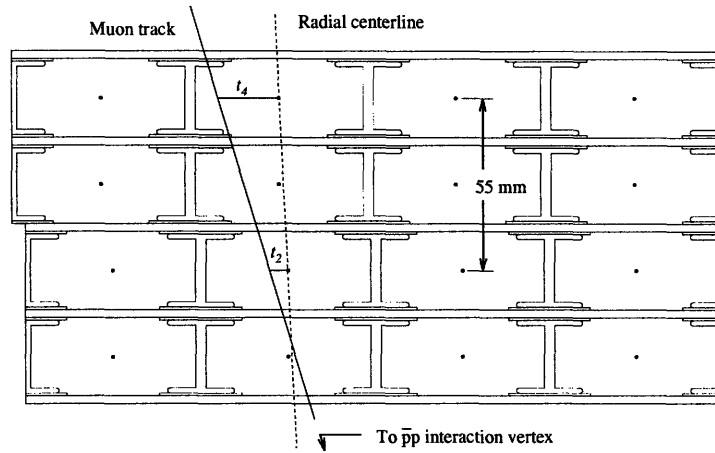
Figure 2-13: Cross section of a CMU chamber. Each vertical array is one stack.
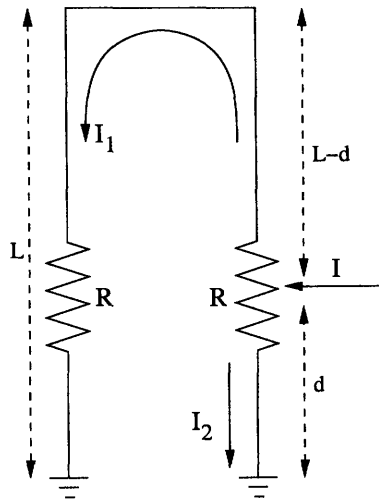
Figure 2-14: The principle of charge division method. The ionization charge is collected at some position d along the z axis, and splits into two currents: I_1 and I_2. From Ohm's law, $I_1 R(1 + \frac{L-d}{L}) = I_2 R \frac{d}{L} \Rightarrow I_1(2L - d) = I_2 d \Rightarrow d = \frac{2LI_1}{I_2+I_1}$. With the approximation that all currents last for the same amount of time Δt, we can write $Q_i = I_i \text{const}_{\Delta t}$. Therefore, by measuring Q_1 and Q_2 one can determine $d = \frac{2LQ_1}{Q_1+Q_2}$.

shielded behind only 5 hadronic interaction lengths, the CMP provides higher purity in muon identification [26]. Those reconstructed muons that have a stub in both the CMU and the CMP are called "CMUP muons".

The CMP is not azimuthally symmetric, but resembles a box surrounding the central region of the detector ($|\eta| < 0.6$). It is made of wire chambers similar to those used for the CMU, but just bigger: $2.5 \times 15 \times 640$ cm^3.

A bigger difference is that CMP contains scintillator counters in addition to wire chambers. The scintillator layers lie on the outer side of the chambers and provide timing information that is used to discard out-of-time muon candidates, which could not possibly be muons originating from the center of the detector. Furthermore, timing helps not have stubs from different bunch crossings piled up, given that the drift time in the CMP can be as large as 1.7 μs [26]. Eventually, the dimensions of the scintillator counters are $2.5 \times 30 \times 320$ cm^3, so two silicon counters are needed to cover the z dimension of the CMU, providing the very crude information of whether a muon stub has positive or the negative z coordinate.

CMX

CMX [26] is very similar to CMP; it consists of same type wire chambers and silicon counters. It differs significantly in geometry though. It covers the region $0.6 < |\eta| < 1$ and is shaped like a conic section on each side of the detector (Fig. 2-12). The wire chambers are grouped in wedges, each 15° in ϕ. Each wedge contains 48 chambers, arranged in 8 layers. The lower 90° of the CMX, which physically penetrate the floor supporting the detector, are called "miniskirt" for obvious reason (Fig. 2-12). This part was not instrumented until past 2003.

IMU

IMU [26] covers the region $1 < |\eta| < 1.5$ (Fig. 2-12). It comprises silicon counters and wire chambers of dimensions $2.5 \times 8.4 \times 363$ cm^3. In combination with ISL tracking, it provides muon reconstruction and momentum measurement in the $|\eta| > 1$ region.

2.2.5 Cerenkov Luminosity Counter

CDF is equipped with the CLC [37], a detector dedicated to measuring instantaneous luminosity ($\mathcal{L}$). It consists of 2×48 Cerenkov counters placed in the far forward and backward region ($3.75 < |\eta| < 4.75$). filled with isobutane at nearly atmospheric pressure.

The number of $p\bar{p}$ interactions (n) in a bunch crossing follows the Poisson distribution with mean $\mu = \sigma_{p\bar{p}}\mathcal{L}t_{BC}$, where $\sigma_{p\bar{p}}$ is the cross section of inelastic $p\bar{p}$ scattering and t_{BC} is the time interval between bunch crossings.

Bunch crossings with $n = 0$ occur with probability $P_0(\mu) = e^{-\mu}$. By measuring the fraction of empty crossings μ can be measured[9] and therefore $\mathcal{L}$.

An alternative method consists in measuring directly μ as N/N_1, where N is the number of CLC counts of some bunch crossing, and N_1 is the average number of CLC counts in the case of single-interaction bunch crossings. N_1 can be measured at low $\mathcal{L}$, when $\mu \ll 1$.

The first method, of measuring empty crossings, has the advantage of not needing any information such as N_1, but at high $\mathcal{L}$ empty crossings become rare, making this method inefficient. On the other hand, the second method depends on the N_1 information, and N/N_1 in reality does not scale linearly with $\mathcal{L}$, as the CLC occupancy grows and is eventually saturated due to the finite number of counters, therefore correction for this non-linearity are required.

The uncertainty in the integrated luminosity measured with the CLC is 6%, to which the biggest contribution comes from the uncertainty in $\sigma_{p\bar{p}}$ at 1.96 TeV.

2.2.6 Data Acquisition

CDF employs approximately 10^6 readout channels. A bunch crossing at $\mathcal{L} \sim 2 \times 10^{32}$ $\text{s}^{-1}\text{cm}^{-2}$ yields on average about 5 $p\bar{p}$ interactions. An event of such multiplicity takes about 200 kB of digitized information volume. It becomes then obvious that not every single bunch crossing can be read, as that would require the enormous bandwidth of

---

[9]Of course it is necessary to correct the measured μ by dividing with the CLC acceptance ϵ.
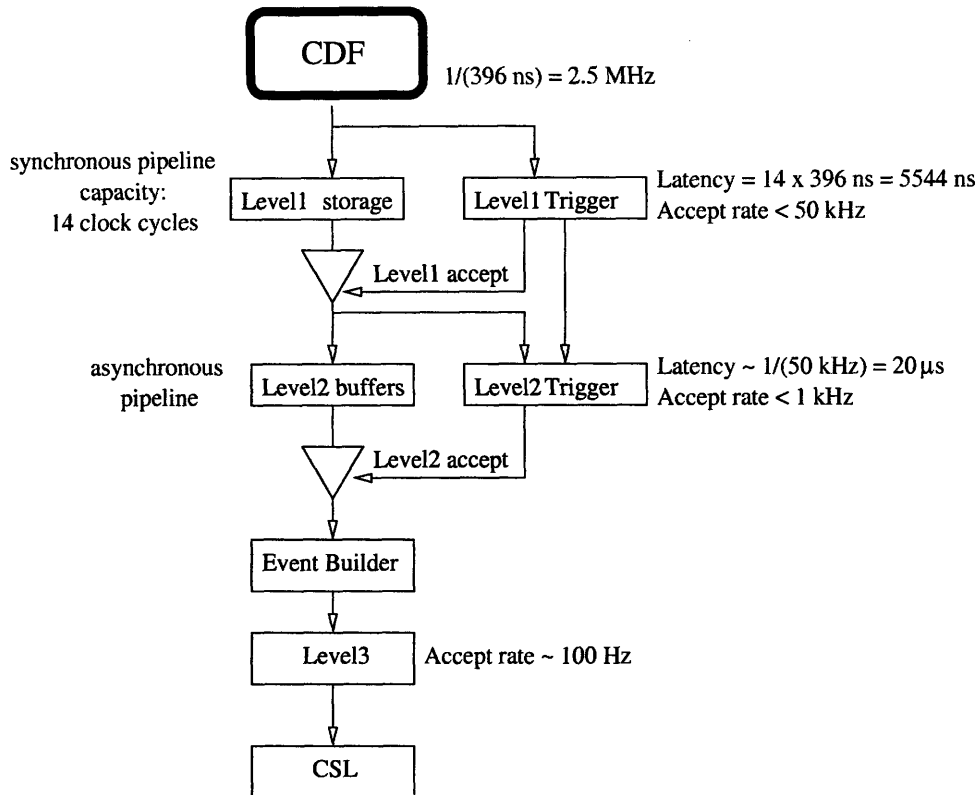
Figure 2-15: Diagram of the CDF DAQ system.

~630 GB/s.

Apart from technically inevitable, it is also sensible to record only those events that pass some quality selection and would be of some interest[10]. For example, an event with leptons should be retained, while for multi-jet events it is enough to keep only a fraction of them, since they are so abundant in $p\bar{p}$ collisions.

The DAQ system [26] is responsible for selecting the best events as they occur. Fig. 2-15 provides an overview of the DAQ architecture.

Level-1

The frequency of 2.5 MHz at which bunches cross is too high to allow for full reconstruction of every event, so the first level of selection is based on fragments of information. This happens in Level-1; an accept/reject decision is made using "prim-

---

[10]In an experiment of the broad scope of CDF it is not trivial to decide which events could be of some interest, since different analyses may see interest in different kinds of events. Furthermore, nobody is certain what the signature of physics beyond the Standard Model will be.

itives", namely coarse information on COT tracks and stubs in the CMU, CMP and CMX [26]. Systems providing primitives are depicted in Fig. 2-16. The XFT crudely reconstructs COT tracks on the $x - y$ plane. The XTRP extrapolates XFT tracks through the calorimeter and the muon system finding matching hits/towers.

Based on the primitives, several algorithms — also called "individual triggers" — contribute to the Level-1 decision. For example, effort is made to keep events with high-p_T tracks, or leptons, or large missing transverse energy ($\not{E}_T$) etc.

The latency of Level-1 is 5.5 μs, in which 14 bunch crossings occur. Therefore, all front-end electronics are equipped with buffers of enough capacity to contain information from 14 bunch crossings. Level-1 then works as a synchronous pipeline; by the time 14 events are pushed back into the buffer, at least one event has been examined and pulled from it, freeing a slot for the current event to be buffered.

Less than 2% of the events pass Level-1, making its accept rate less than 50 kHz.

Level-2

Level-2 functions as an asynchronous pipeline, where events are processed in FIFO mode [26]. With no more than 50 kHz input rate, it can afford up to 1/50 kHz = 20 μs to decide on each event[11].

In its decision, Level-2 takes into account the primitives of Level-1, in addition to showermax information, as shown in Fig. 2-16.

The acceptance rate of Level-2 is less than 1 kHz. Effort is made to maintain this rate as close to 1 kHz as possible, by readjusting the trigger requirements as $\mathcal{L}$ changes, making them stricter at high $\mathcal{L}$ and looser at low $\mathcal{L}$.

Event Builder

In the case of a Level-2 accept, the whole detector is eventually read out. The EVB collects the fragments of the event and passes them to Level-3. Reading out the front-end electronics of the whole detector takes about 1 ms, which is why this step

---

[11]Actually, since up to 4 events can be kept in the Level-2 buffer, the latency can be even greater, without causing dead-time, provided that this is not the case for too many events.
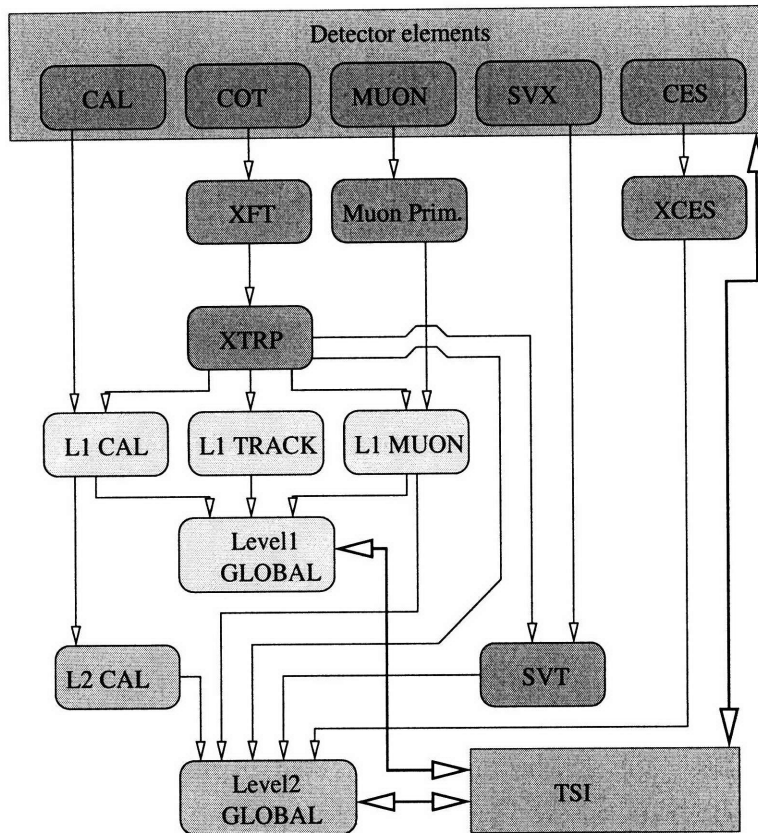
Figure 2-16: Information flow within Level-1 and Level-2. XCES is a system that generates the stimulated showermax bitmap and finds matching tracks extrapolated by XTRP to define electron candidates. The SVT extrapolates XFT tracks into the SVX, providing the D and ϕ_0 information (Table 2.1). The TSI coordinates the flow of information and interfaces to the CDF clock, which is used to know when a bunch crossing is occurring.
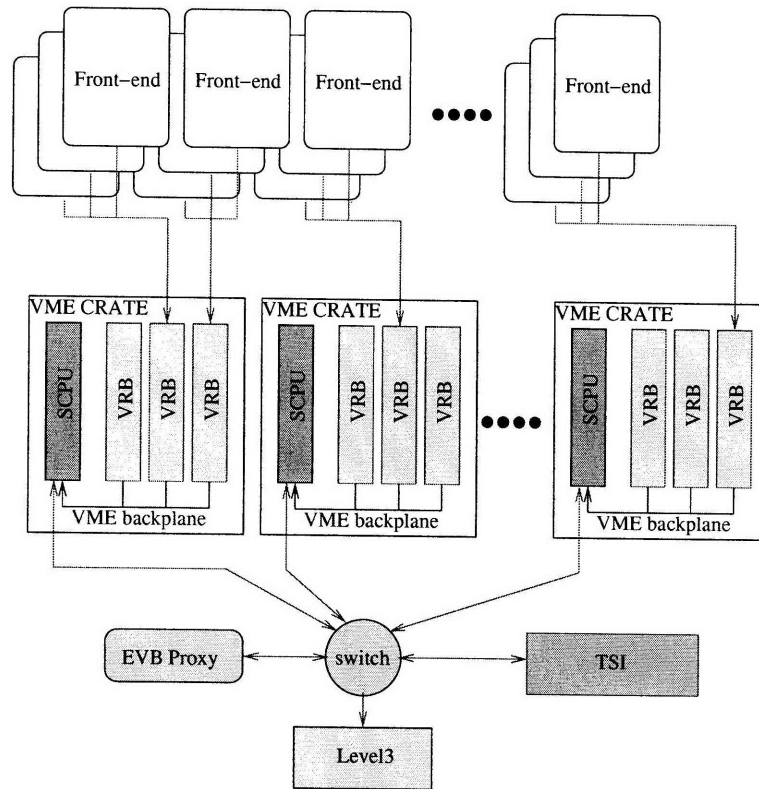
Figure 2-17: Diagram of the Event Builder.

is only possible after having discarded over 99.96% of the events.

EVB (Fig. 2-17) lies in 21 VME crates, each containing one Linux computer, referred to as SCPU [38]. Each crate is dedicated to reading a different part of the detector. Apart from the SCPU, each crate contains a series of memory buffers, the VRBs. When the front-end crates are read, the information of the event is first stored in the VRBs. Each SCPU reads the VRBs of it own crate through the VME backplane of the crate, which in combination with the GigaBit Ethernet networking allows for the desired system speed. On reading the VRBs, a byte-count check is performed, as well as checks of the size of each buffer entry [39]. Though in principle EVB should not be discarding any events, it does so if information is missing or corrupted.

The function of the EVB is coordinated by the EVB Proxy, a process running on a dedicated Linux machine. All acknowledgement messages within the EVB are circulated through the EVB Proxy, and so does any information exchanged with the TSI and Level-3.

56

Level-3

Level-3 is the last stage of trigger selection [38]. Receiving events from the EVB at
< 1 kHz, it is purely software implemented, performing three basic functions:

1. Concatenates same-event fragments coming from the EVB into an event entry.

2. Imposes the final selection, taking into account the reconstructed objects information.

3. Submits passing events to the CSL for storage.

There is a whole cluster of 411 Linux computers counting 2.4 THz of CPU dedicated to Level-3. Though all computers are nearly identical, they are separated in three categories, depending on their task:

- 18 Converter nodes: They receive event fragments from the EVB and combine them to form self-contained event records which they pass to available Processor nodes.

- 384 Processor nodes: Upon reception of events from a Converter, they apply the Level-3 filter to either discard or pass them to an Output node, after some reformatting that reshapes the passing entries to their final format.

- 9 Output nodes: They receive the passing events from Processor nodes and propagate them to the CSL for storage.

The Level-3 cluster is separated in 18 identical subsets, called "subfarms"[12]. This way, data handling proceeds in 18 independent, parallel streams which share the load of incoming events. Each subfarm contains 1 Converter, 21 or 22 Processors, and shares an Output with another subfarm. On every Processor, 5 Level-3 filters run simultaneously, on hyper-threaded dual-core Intel CPUs. The Converter of each subfarm is allowed to only submit events to Processors of its own subfarm, and the Processors of each subfarm can only send events to the Output node serving it.

---

[12]A term appropriate for a subdivision of the whole Level-3 cluster, which is called "farm" in CDF jargon.

The operation of Level-3 is coordinated by the Level-3 Proxy application, running on a dedicated computer. The Proxy collects and sends acknowledgements from and to the computers of the cluster, and communicates with the EVB Proxy to indicate among other things which Converter is available to receive the next event.

Filtering is done by a program written in C++, the Level-3 filter executable, which applies criteria stored in a centralized database implemented in Oracle. In the database is stored the *trigger table*, which is a list of "triggers". Each trigger is structured to contain the following information:

1. The prerequisite Level-1 and Level-2 triggers.

2. The C++ reconstruction modules that should be used and in what order.

3. The specific selection criteria decided having some physics goal, for example a cut in some invariant mass in the event.

4. The name of the dataset in which to store the event if it passes the trigger selection.

The output rate of Level-3 is about 100 Hz. The events passing Level-3 are sent to the CSL for immediate storage. From there, they are shortly sent to the FCC for permanent storage on magnetic tape.

2.2.7 Off-line production

Data analysis is not performed on the raw data. Before the data on tape are usable, the off-line production process has to take place.

At production [26], the raw data banks are unpacked and physics objects are reconstructed in full detail. This is similar to what is done at Level-3, but the off-line reconstruction is much more elaborate, applying the latest calibrations, since those reconstructed objects will be the final ones to be used for analysis.

Since passing Level-3, each event contains the information of the dataset(s) it belongs to. At the production, even further partitioning is made; datasets are collections of filesets, which are collections of files containing events.

For the needs of each analysis, the raw data are taken from the appropriate dataset and are converted to a convenient format. Since ROOT [40] is the adopted analysis framework, the format varies between different architectures of ROOT Trees. For example, one is the "topNtuple", used mostly by collaborators doing t-quark analyses, but a more common format, used also in the present analysis, is the "Standard Ntuple" (Stntuple).

Chapter 3

Data Analysis

The analysis going into this thesis was conducted in two rounds: first with 1 fb^{-1} of data, and then with 2 fb^{-1}. The first round has been documented in [41, 42, 43, 44]. An updated publication is currently being prepared for the second one. This chapter is an adaptation of [41], while chapter 4 presents material that will be in the publication of the second round.

3.1 Strategy

Sec. 1.3 motivates the goal of this analysis, viz. the model-independent search for new physics. The method is to obtain a satisfactory description of the Standard Model expectation in channels where high-p_T data are observed, and employ an array of probes to seek for statistically significant discrepancies between data and Standard Model background.

Crucial for model-independence is to not focus on channels sensitive to particular models, but examine data in as many channels as possible. That introduces to this analysis over two million events (in 1 fb^{-1}), ranging from abundant QCD to rare electroweak ones. Studying this large volume of qualitatively diverse data requires reducing the information content of each event to bare bones and characterizing each event in terms of physics objects that maintain the same meaning universally in any kind of event. In each event, the 4-momenta of any reconstructed physics objects in

61

its final state are recorded. These objects can be leptons, photons, hadronic jets or missing energy.

Another ingredient of model-independence is to not segregate the data into "control" and "signal" regions *a priori*, namely into regions where new physics is assumed to not exist or to exist respectively. In most analyses control regions are predefined, to adjust correction factors, under the assumption that there is no new physics in those regions and that the extrapolation of correction factors from the control to the signal region is valid. However, what is considered control region in one analysis is often signal region in some other, so, to be as generic as possible, one needs to treat all data as signal and control regions simultaneously, to address the question "how well does the Standard Model implementation describe the data?" If there is indeed detectable new physics, then it will be impossible to achieve good agreement between data and Standard Model simultaneously in all regions. More in Sec. C.

The Standard Model prediction is implemented in three steps:

1. Monte Carlo generation and matching [45] of samples simulating the Standard Model processes.

2. CDF detector simulation, which models the detector response to the MC generated events. For that, the GEANT-based package CDFSIM is used.

3. Fine-tuning of the outcome of CDFSIM to account for theoretical and experimental correction factors.

Structurally, the analysis contains four parts:

1. The VISTA global fit, which adjusts and applies the correction model, providing the Standard Model background of the best possible global agreement with the data, exploiting the flexibility granted by the correction model.

2. The VISTA comparison, which examines the statistical significance of features in the bulk of all distributions and sorts the information in a comprehensive way.

3. The SLEUTH search, which focuses on the high-$\sum p_T$ tails searching for excesses of data.

4. The Bump Hunter search (present only for the second round of the analysis), which scans all mass variables for local excesses of data, potentially indicating a new resonance.

The above statistical probes are employed simultaneously, rather than sequentially. So, an effect highlighted by SLEUTH prompts additional investigation of the discrepancy, usually resulting in a specific hypothesis explaining the discrepancy in terms of a detector effect or adjustment to the Standard Model prediction that is then fed back and tested for global consistency.

Statistical significance is a necessary but insufficient condition for discovery. A statistically significant discrepancy could be attributed to inaccuracy in the Standard Model implementation, or in modeling the detector response. These possibilities would need to be considered on a case-by-case basis. In the event of a significant discrepancy, the breadth of view of this analysis can be exploited to evaluate the plausibility of it being a detector effect or a problem in the Standard Model implementation.

Forming hypotheses for the cause of specific discrepancies, implementing those hypotheses to assess their wider consequences, and testing global agreement after the implementation are emphasized as the crucial activities for the investigator throughout the process of data analysis. This process is constrained by the requirement that all adjustments be physically motivated. The investigation and resolution of discrepancies highlighted by the algorithms is *the* defining characteristic of this global analysis [1].

This search for new physics terminates when either a compelling case for new physics is made, or there remain no statistically significant discrepancies on which a new physics case can be made. In the former case, to quantitatively assess the sig-

---

[1]It is not possible to systematically simulate the process of constructing, implementing, and testing hypotheses motivated by particular discrepancies, since this process is carried out by individuals. The statistical interpretation of this analysis is made bearing this process in mind.

nificance of the potential discovery, a full treatment of systematic uncertainties must be implemented. In the latter case, it is sufficient to demonstrate that all observed effects are not in significant disagreement with an appropriate global Standard Model description.

3.2 VISTA

This section describes VISTA: object identification, event selection, estimation of Standard Model backgrounds, simulation of the CDF detector response, development of a correction model, and results.

3.2.1 Object identification

Energetic and isolated electrons, muons, taus, photons, jets, and b-tagged jets with $|\eta_{\text{det}}| < 2.5$ and $p_T > 17$ GeV are identified according to CDF standard criteria. The same criteria are used for all events. The isolation criteria employed vary according to object, but roughly require less than 2 GeV of extra energy flow within a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ in η–ϕ space around each object.

Standard CDF criteria [46] are used to identify electrons ($e^{\pm}$) in the central and plug regions of the CDF detector. Electrons are characterized by a narrow shower in the central or plug electromagnetic calorimeter and a matching isolated track in the central gas tracking chamber or a matching plug track in the silicon detector.

Standard CDF muons ($\mu^{\pm}$) are identified using three separate subdetectors in the regions $|\eta_{\text{det}}| < 0.6$, $0.6 < |\eta_{\text{det}}| < 1.0$, and $1.0 < |\eta_{\text{det}}| < 1.5$ [46]. Muons are characterized by a track in the central tracking chamber matched to a track segment in the central muon detectors, with energy consistent with minimum ionizing deposition in the electromagnetic and hadronic calorimeters along the muon trajectory.

Narrow central jets with a single charged track are identified as tau leptons ($\tau^{\pm}$) that have decayed hadronically [47]. Taus are distinguished from electrons by requiring a substantial fraction of their energy to be deposited in the hadron calorimeter; taus are distinguished from muons by requiring no track segment in the muon detec-

tor coinciding with the extrapolated track of the tau. Track and calorimeter isolation requirements are imposed.

Standard CDF criteria requiring the presence of a narrow electromagnetic cluster with no associated tracks are used to identify photons (γ) in the central and plug regions of the CDF detector [48].

Jets (j) are reconstructed using the JetClu [49] clustering algorithm with a cone of size $\Delta R = 0.4$, unless the event contains one or more jets with $p_T > 200$ GeV and no leptons or photons, in which case cones of $\Delta R = 0.7$ are used. Jet energies are appropriately corrected to the parton level [50]. Since uncertainties in the Standard Model prediction grow with increasing jet multiplicity, up to the four largest p_T jets are used to characterize the event; any reconstructed jets with p_T-ordered ranking of five or greater are neglected and their energy is treated as unclustered, except in final states with small summed scalar transverse momentum containing only jets.

A secondary vertex b-tagging algorithm is used to identify jets likely resulting from the fragmentation of a bottom quark (b) produced in the hard scattering [51].

Momentum visible in the detector but not clustered into an electron, muon, tau, photon, jet, or b-tagged jet is referred to as unclustered momentum (uncl).

Missing momentum ($\not{p}$) is calculated as the negative vector sum of the 4-vectors of all identified objects and unclustered momentum. An event is said to contain a $\not{p}$ object if the transverse momentum of this object exceeds 17 GeV, and if additional quality criteria discriminating against fake missing momentum due to jet mismeasurement are satisfied [2].

---

[2]An additional quality criterion is applied to the significance of the missing transverse momentum $\vec{\not{p}}_T$ in an event, requiring that the energies of hadronic objects can not be adjusted within resolution to reduce the missing transverse momentum to less than 10 GeV. The transverse components of all hadronic energy clusters $\vec{p}_{Ti}$ in the event are projected onto the unit missing transverse momentum vector $\hat{\not{p}}_T = \vec{\not{p}}_T / |\vec{\not{p}}_T|$, and a "conservative" missing transverse momentum $\not{p}_T' = \not{p}_T - 2.5\sqrt{\sum_i |\vec{p}_{Ti} \cdot \hat{\not{p}}_T|}$ is defined, where the sum is over hadronic energy clusters in the event, and the hadronic energy resolution of the CDF detector has been approximated as $100\%\sqrt{p_{Ti}}$, expressed in GeV. An event is said to contain missing transverse momentum if $\not{p}_T > 17$ GeV and $\not{p}_T' > 10$ GeV.

3.2.2 Event selection

Events containing an energetic and isolated electron, muon, tau, photon, or jet are selected. A set of three level online triggers requires:

- a central electron candidate with $p_T > 18$ GeV passing level 3, with an associated track having $p_T > 8$ GeV and an electromagnetic energy cluster with $p_T > 16$ GeV at levels 1 and 2; or

- a central muon candidate with $p_T > 18$ GeV passing level 3, with an associated track having $p_T > 15$ GeV and muon chamber track segments at levels 1 and 2; or

- a central or plug photon candidate with $p_T > 25$ GeV passing level 3, with hadronic to electromagnetic energy less than 1:8 and with energy surrounding the photon to the photon's energy less than 1:7 at levels 1 and 2; or

- a central or plug jet with $p_T > 20$ GeV passing level 3, with 15 GeV of transverse momentum required at levels 1 and 2, with corresponding prescales of 50 and 25, respectively; or

- a central or plug jet with $p_T > 100$ GeV passing level 3, with energy clusters of 20 GeV and 90 GeV required at levels 1 and 2; or

- a central electron candidate with $p_T > 4$ GeV and a central muon candidate with $p_T > 4$ GeV passing level 3, with a muon segment, electromagnetic cluster, and two tracks with $p_T > 4$ GeV required at levels 1 and 2; or

- a central electron or muon candidate with $p_T > 4$ GeV and a plug electron candidate with $p_T > 8$ GeV, requiring a central muon segment and track or central electromagnetic energy cluster and track at levels 1 and 2, together with an isolated plug electromagnetic energy cluster; or

- two central or plug electromagnetic clusters with $p_T > 18$ GeV passing level 3, with hadronic to electromagnetic energy less than 1:8 at levels 1 and 2; or

66

- two central tau candidates with $p_T > 10$ GeV passing level 3, each with an associated track having $p_T > 10$ GeV and a calorimeter cluster with $p_T > 5$ GeV at levels 1 and 2.

Events satisfying one or more of these online triggers are recorded for further study. Offline event selection for this analysis uses a variety of further filters. Single object requirements keep events containing:

- a central electron with $p_T > 25$ GeV, or

- a plug electron with $p_T > 40$ GeV, or

- a central muon with $p_T > 25$ GeV, or

- a central photon with $p_T > 60$ GeV, or

- a central jet or b-tagged jet with $p_T > 200$ GeV, or

- a central jet or b-tagged jet with $p_T > 40$ GeV (prescaled by a factor of roughly 10^4),

possibly with other objects present. Multiple object criteria select events containing:

- two electromagnetic objects (electron or photon) with $|\eta| < 2.5$ and $p_T > 25$ GeV, or

- two taus with $|\eta| < 1.0$ and $p_T > 17$ GeV, or

- a central electron or muon with $p_T > 17$ GeV and a central or plug electron, central muon, or central tau with $p_T > 17$ GeV, or

- a central photon with $p_T > 40$ GeV and a central electron or muon with $p_T > 17$ GeV, or

- a central or plug photon with $p_T > 40$ GeV and a central tau with $p_T > 40$ GeV, or

- a central photon with $p_T > 40$ GeV and a central b-jet with $p_T > 25$ GeV, or

- a central jet or b-tagged jet with $p_T > 40$ GeV and a central tau with $p_T >$ 17 GeV (prescaled by a factor of roughly 10^3), or

- a central or plug photon with $p_T > 40$ GeV and two central taus with $p_T >$ 17 GeV, or

- a central or plug photon with $p_T > 40$ GeV and two central b-tagged jets with $p_T > 25$ GeV, or

- a central or plug photon with $p_T > 40$ GeV, a central tau with $p_T > 25$ GeV, and a central b-tagged jet with $p_T > 25$ GeV,

possibly with other objects present. Explicit online triggers feeding this offline selection are required. The p_T thresholds for these criteria are chosen to be sufficiently above the online trigger turn-on curves that trigger efficiencies can be treated as roughly independent of object p_T.

Good run criteria are imposed, requiring the operation of all major subdetectors. To reduce contributions from cosmic rays and events from beam halo, standard CDF cosmic ray and beam halo filters are applied [52].

These selections result in a sample of roughly two million high-p_T data events in an integrated luminosity of 927 pb^{-1}.

3.2.3 Event generation

Standard Model backgrounds are estimated by generating a large sample of Monte Carlo events, using the PYTHIA [53], MADEVENT [54], and HERWIG [55] generators. MADEVENT performs a leading order matrix element calculation, and provides 4-vector information corresponding to the outgoing legs of the underlying Feynman diagrams, together with color flow information. PYTHIA 6.218 is used to handle showering and fragmentation. The CTEQ5L [56] parton distribution functions are used.

QCD jets. QCD dijet and multijet production are estimated using PYTHIA. Samples are generated with Tune A [57] with lower cuts on $\hat{p}_T$, the transverse momentum of the scattered partons in the center of momentum frame of the incoming partons, of 0, 10, 18, 40, 60, 90, 120, 150, 200, 300, and 400 GeV. These samples are combined to provide a complete estimation of QCD jet production, using the sample with greatest statistics in each range of $\hat{p}_T$.

γ+jets. The estimation of QCD single prompt photon production comes from PYTHIA. Five samples are generated with Tune A corresponding to lower cuts on $\hat{p}_T$ of 8, 12, 22, 45, and 80 GeV. These samples are combined to provide a complete estimation of single prompt photon production in association with one or more jets, placing cuts on $\hat{p}_T$ to avoid double counting.

γγ+jets. QCD diphoton production is estimated using PYTHIA.

V+jets. The estimation of V+jets processes (with V denoting W or Z), where the W or Z decays to first or second generation leptons, comes from MADEVENT, with PYTHIA employed for showering. Tune AW [57] is used within PYTHIA for these samples. The CKKW matching prescription [45] with a matching scale of 15 GeV is used to combine these samples and avoid double counting. Additional statistics are generated on the high-p_T tails using the MLM matching prescription [58]. The factorization scale is set to the vector boson mass; the renormalization scale for each vertex is set to the p_T of the jet. W+4 jets are generated inclusively in the number of jets; Z+3 jets are generated inclusively in the number of jets.

VV+jets. The estimation of WW, WZ, and ZZ production with zero or more jets comes from PYTHIA.

$V\gamma$+jets. The estimation of $W\gamma$ and $Z\gamma$ production comes from MADEVENT, with showering provided by PYTHIA. These samples are inclusive in the number of jets.

$W(\rightarrow \tau\nu)$+jets. Estimation of $W \rightarrow \tau\nu$ with zero or more jets comes from PYTHIA.

69

$Z(\to \tau\tau)$+jets. Estimation of $Z \to \tau\tau$ with zero or more jets comes from PYTHIA.

$t\bar{t}$. Top quark pair production is estimated using HERWIG assuming a top quark mass of 175 GeV and NNLO cross section of 6.77 ± 0.42 pb [59].

Remaining processes, including for example $Z(\to \nu\bar{\nu})\gamma$ and $Z(\to \ell^+\ell^-)b\bar{b}$, are generated by systematically looping over possible final state partons, using MAD-GRAPH [60] to determine all relevant diagrams, and using MADEVENT to perform a Monte Carlo integration over the final state phase space and to generate events. The MLM matching prescription is employed to combine samples with different numbers of final state jets.

A higher statistics estimate of the high-p_T tails is obtained by computing the thresholds in $\sum p_T$ corresponding to the top 10% and 1% of each process, where $\sum p_T$ denotes the scalar summed transverse momentum of all identified objects in an event. Roughly ten times as many events are generated for the top 10%, and roughly one hundred times as many events are generated for the top 1%.

Cosmic rays. Backgrounds from cosmic ray or beam halo muons that interact with the hadronic or electromagnetic calorimeters, producing objects that look like a photon or jet, are estimated using a sample of data events containing fewer than three reconstructed tracks. This procedure is described in more detail in Appendix A.2.1.

Minimum bias. Minimum bias events are overlaid according to run-dependent instantaneous luminosity in some of the Monte Carlo samples, including those used for inclusive W and Z production. In all samples not containing overlaid minimum bias events, including those used to estimate QCD dijet production, additional unclustered momentum is added to events to mimic the effect of the majority of multiple interactions, in which a soft dijet event accompanies the rare hard scattering of interest. A random number is drawn from a Gaussian centered at 0 with width 1.5 GeV for each of the x and y components of the added unclustered momentum. Backgrounds due to two rare hard scatterings occurring in the same bunch crossing are estimated by forming overlaps of events, as described in Appendix A.2.2.

Each generated Standard Model event is assigned a weight, calculated as the cross section for the process (in units of picobarns) divided by the number of events generated for that process, representing the number of such events expected in a data sample corresponding to an integrated luminosity of 1 pb^{-1}. When multiplied by the integrated luminosity of the data sample used in this analysis, the weight gives the predicted number of such events in this analysis.

3.2.4 Detector simulation

The response of the CDF detector is simulated using a GEANT-based detector simulation (CDFSIM) [61], with GFLASH [62] used to simulate shower development in the calorimeter.

In $p\bar{p}$ collisions there is an ordering of frequency with which objects of different types are produced: many more jets (j) are produced than b-jets (b) or photons (γ), and many more of these are produced than charged leptons (e, μ, τ). The CDF detectors and reconstruction algorithms have been designed so that the probability of misreconstructing a frequently produced object as an infrequently produced object is small. The fraction of central jets that CDFSIM misreconstructs as photons, electrons, and muons is $\sim 10^{-3}$, $\sim 10^{-4}$, and $\sim 10^{-5}$, respectively. Due to these small numbers, the use of CDFSIM to model these fake processes would require generating samples with prohibitively large statistics. Instead, the modeling of a frequently produced object faking a less frequently produced object (specifically: j faking b, γ, e, μ, or τ; or b or γ faking e, μ, or τ) is obtained by the application of a misidentification probability, a particular type of correction factor in the VISTA correction model, described in the next section.

In Monte Carlo samples passed through CDFSIM, reconstructed leptons and photons are required to match to a corresponding generator level object. This procedure removes reconstructed leptons or photons that arise from a misreconstructed quark or gluon jet.

3.2.5   Correction model

Unfortunately some numbers that can not be determined from first principles enter the comparison between data and the Standard Model prediction. These numbers are referred to as "correction factors". This correction model is applied to generated Monte Carlo events to obtain the Standard Model prediction across all final states.

Correction factors must be obtained from the data themselves. These factors may be thought of as Bayesian nuisance parameters. The actual values of the correction factors are not directly of interest. Of interest is the comparison of data to Standard Model prediction, with correction factors adjusted to whatever they need to be, consistent with external constraints, to bring the Standard Model into closest agreement with the data.

The traditional prescription for determining these correction factors is to "measure" them in a "control region" in which no signal is expected. This procedure encounters difficulty when the entire high-p_T data sample is considered to be a signal region. The approach adopted instead is to ask whether a consistent set of correction factors can be chosen so that the Standard Model prediction is in agreement with the CDF high-p_T data.

The correction model is obtained by an iterative procedure informed by observed inadequacies in modeling. The process of correction model improvement, motivated by observed discrepancies, may allow a real signal to be artificially suppressed. If adjusting correction factor values within allowed bounds removes a signal, then the case for the signal disappears, since it can be explained in terms of known physics. This is true in any analysis. The stronger the constraints on the correction model, the more difficult it is to artificially suppress a real signal. By requiring a consistent interpretation of hundreds of final states, VISTA is less likely to mistakenly explain away new physics than analyses of more limited scope.

The 44 correction factors currently included in the correction model are shown in Table 4.2. These factors can be classified into two categories: theoretical and experimental. A more detailed description of each individual correction factor is

Code	Category	Explanation	Value	Error	Error(%)
0001	luminosity	CDF integrated luminosity	927	20	2.2
0002	k-factor	cosmic γ	0.69	0.05	7.3
0003	k-factor	cosmic j	0.446	0.014	3.1
0004	k-factor	$1\gamma1j$ photon+jet(s)	0.95	0.04	4.2
0005	k-factor	$1\gamma2j$	1.2	0.05	4.1
0006	k-factor	$1\gamma3j$	1.48	0.07	4.7
0007	k-factor	$1\gamma4j+$	1.97	0.16	8.1
0008	k-factor	$2\gamma0j$ diphoton(+jets)	1.81	0.08	4.4
0009	k-factor	$2\gamma1j$	3.42	0.24	7.0
0010	k-factor	$2\gamma2j+$	1.3	0.16	12.3
0011	k-factor	$W0j$ W (+jets)	1.453	0.027	1.9
0012	k-factor	$W1j$	1.06	0.03	2.8
0013	k-factor	$W2j$	1.02	0.03	2.9
0014	k-factor	$W3j+$	0.76	0.05	6.6
0015	k-factor	$Z0j$ Z (+jets)	1.419	0.024	1.7
0016	k-factor	$Z1j$	1.18	0.04	3.4
0017	k-factor	$Z2j+$	1.03	0.05	4.8
0018	k-factor	$2j$ $\hat{p}_T < 150$	0.96	0.022	2.3
0019	k-factor	$2j$ $150 < \hat{p}_T$	1.256	0.028	2.2
0020	k-factor	$3j$ $\hat{p}_T < 150$	0.921	0.021	2.3
0021	k-factor	$3j$ $150 < \hat{p}_T$	1.36	0.03	2.4
0022	k-factor	$4j$ $\hat{p}_T < 150$	0.989	0.025	2.5
0023	k-factor	$4j$ $150 < \hat{p}_T$	1.7	0.04	2.3
0024	k-factor	$5j+$	1.25	0.05	4.0
0025	ID eff	$p(e \rightarrow e)$ central	0.986	0.006	0.6
0026	ID eff	$p(e \rightarrow e)$ plug	0.933	0.009	1.0
0027	ID eff	$p(\mu \rightarrow \mu)$ $|\eta| < 0.6$	0.845	0.008	0.9
0028	ID eff	$p(\mu \rightarrow \mu)$ $0.6 < |\eta|$	0.915	0.011	1.2
0029	ID eff	$p(\gamma \rightarrow \gamma)$ central	0.974	0.018	1.8
0030	ID eff	$p(\gamma \rightarrow \gamma)$ plug	0.913	0.018	2.0
0031	ID eff	$p(b \rightarrow b)$ central	1	0.04	4.0
0032	fake rate	$p(e \rightarrow \gamma)$ plug	0.045	0.012	27.0
0033	fake rate	$p(q \rightarrow e)$ central	9.71×10^{-5}	1.9×10^{-6}	2.0
0034	fake rate	$p(q \rightarrow e)$ plug	0.000876	1.8×10^{-5}	2.1
0035	fake rate	$p(q \rightarrow \mu)$	1.157×10^{-5}	2.7×10^{-7}	2.3
0036	fake rate	$p(j \rightarrow b)$	0.01684	0.00027	1.6
0037	fake rate	$p(q \rightarrow \tau)$ $p_T < 60$	0.00341	0.00012	3.5
0038	fake rate	$p(q \rightarrow \tau)$ $60 < p_T$	0.00038	4×10^{-5}	10.5
0039	fake rate	$p(q \rightarrow \gamma)$ central	0.000265	1.5×10^{-5}	5.7
0040	fake rate	$p(q \rightarrow \gamma)$ plug	0.00159	0.00013	8.2
0041	trigger	$p(e \rightarrow \text{trig})$ central, $p_T > 25$	0.976	0.007	0.7
0042	trigger	$p(e \rightarrow \text{trig})$ plug, $p_T > 25$	0.835	0.015	1.8
0043	trigger	$p(\mu \rightarrow \text{trig})$ $|\eta| < 0.6$, $p_T > 25$	0.917	0.007	0.8
0044	trigger	$p(\mu \rightarrow \text{trig})$ $0.6 < |\eta| < 1.0$, $p_T > 25$	0.96	0.01	1.0

Table 3.1: The 44 factors introduced in the correction model. All values are dimensionless with the exception of code 0001 (luminosity), which has units of pb^{-1}. The values and uncertainties of these correction factors are valid within the context of this correction model.

provided in Appendix A.4.

Theoretical correction factors reflect the practical difficulty of calculating accurately within the framework of the Standard Model. These factors take the form of k-factors, so-called "knowledge factors," representing the ratio of the unavailable all order cross section to the calculable leading order cross section. Twenty-three k-factors are used for Standard Model processes including QCD multijet production, W+jets, Z+jets, and (di)photon+jets production.

Experimental correction factors include the integrated luminosity of the data, efficiencies associated with triggering on electrons and muons, efficiencies associated with the correct identification of physics objects, and fake rates associated with the mistaken identification of physics objects. Obtaining an adequate description of object misidentification has required an understanding of the underlying physical mechanisms by which objects are misreconstructed, as described in Appendix A.1.

In the interest of simplicity, correction factors representing k-factors, efficiencies, and fake rates are generally taken to be constants, independent of kinematic quantities such as object p_T, with only five exceptions. The p_T dependence of three fake rates is too large to be treated as approximately constant: the jet faking electron rate $p(j \rightarrow e)$ in the plug region of the CDF detector; the jet faking b-tagged jet rate $p(j \rightarrow b)$, which increases steadily with increasing p_T; and the jet faking tau rate $p(j \rightarrow \tau)$, which decreases steadily with increasing p_T. Two other fake rates possess geometrical features in η–ϕ due to the construction of the CDF detector: the jet faking electron rate $p(j \rightarrow e)$ in the central region, because of the fiducial tower geometry of the electromagnetic calorimeter; and the jet faking muon rate $p(j \rightarrow \mu)$, due to the non-trivial fiducial geometry of the muon chambers. After determining appropriate functional forms, a single overall multiplicative correction factor, determined by the global fit, is used

Correction factor values are obtained from a global fit to the data. The procedure is outlined here, with further details relegated to Appendix A.3.

Events are first partitioned into final states according to the number and types of objects present. Each final state is then subdivided into bins according to each

object's detector pseudorapidity (η_{det}) and transverse momentum (p_T), as described in Appendix A.3.1.

Generated Monte Carlo events, adjusted by the correction model, provide the Standard Model prediction for each bin. The Standard Model prediction in each bin is therefore a function of the correction factor values. A figure of merit is defined to quantify global agreement between the data and the Standard Model prediction, and correction factor values are chosen to maximize this agreement, consistent with external experimental constraints.

Letting $\vec{s}$ represent a vector of correction factors, for the kth bin

$$\chi_k^2(\vec{s}) = \frac{(\text{Data}[k] - \text{SM}[k])^2}{\sqrt{\text{SM}[k]}^2 + \delta\text{SM}[k]^2}, \tag{3.1}$$

where Data[k] is the number of data events observed in the kth bin, SM[k] is the number of events predicted by the Standard Model in the kth bin, δSM[k] is the Monte Carlo statistical uncertainty on the Standard Model prediction in the kth bin [3], and $\sqrt{\text{SM}[k]}$ is the statistical uncertainty on the expected data in the kth bin. The Standard Model prediction SM[k] in the kth bin is a function of $\vec{s}$.

Relevant information external to the VISTA high-p_T data sample provides additional constraints in this global fit. The CDF luminosity counters measure the integrated luminosity of the sample described in this article to be 902 pb^{-1} ± 6% by measuring the fraction of bunch crossings in which zero inelastic collisions occur [63]. The integrated luminosity of the sample measured by the luminosity counters enters in the form of a Gaussian constraint on the luminosity correction factor. Higher order theoretical calculations exist for some Standard Model processes, providing constraints on corresponding k-factors, and some CDF experimental correction factors are also constrained from external information. In total, 26 of the 44 correction factors

---

[3]Given a set of Monte Carlo events with individual weights w_j, so that the total Standard Model prediction from these Monte Carlo events is SM = $\sum_j w_j$ events, the "effective weight" w_{eff} of these events can be taken to be the weighted average of the weights: $w_{\text{eff}} = \frac{\sum_j w_j w_j}{\sum_j w_j}$. The "effective number of Monte Carlo events" is $N_{\text{eff}} = \text{SM}/w_{\text{eff}}$, and the error on the Standard Model prediction is $\delta\text{SM} = \text{SM}/\sqrt{N_{\text{eff}}}$.

are constrained. The specific constraints employed are provided in Appendix A.3.2.

The overall function to be minimized takes the form

$$\chi^2(\vec{s}) = \left(\sum_{k \in \text{bins}} \chi_k^2(\vec{s}) \right) + \chi_{\text{constraints}}^2(\vec{s}), \qquad (3.2)$$

where the sum in the first term is over bins in the CDF high-p_T data sample with $\chi_k^2(\vec{s})$ defined in Eq. 3.1, and the second term is the contribution from explicit constraints.

Minimization of $\chi^2(\vec{s})$ in Eq. 3.2 as a function of the vector of correction factors $\vec{s}$ results in a set of correction factor values $\vec{s}_0$ providing the best global agreement between the data and the Standard Model prediction. The best fit correction factor values are shown in Table 4.2, together with absolute and fractional uncertainties. The determined uncertainties are not used explicitly in the subsequent analysis, but rather provide information used implicitly to assist in appropriate adjustment to the correction model in light of observed discrepancies. The uncertainties are verified by subdividing the data into thirds, performing separate fits on each third, and noting that the correction factor values obtained with each subset are consistent within quoted uncertainties. Further details on the correlation matrix and other technical aspects of this global fit can be found in Appendix A.3.3.

Although the correction factors are determined from a global fit, in practice the determination of many correction factors' values are dominated by one recognizable subsample. The rate $p(j \to e)$ for a jet to fake an electron is determined largely by the number of events in the ej final state, since the largest contribution to this final state is from dijet events with one jet misreconstructed as an electron. Similarly, the rates $p(j \to b)$ and $p(j \to \tau)$ for a jet to fake a b-tagged jet and tau lepton are determined largely by the number of events in the bj and τj final states, respectively. The determination of the fake rate $p(j \to \gamma)$, photon efficiency $p(\gamma \to \gamma)$, and k-factors for prompt photon production and prompt diphoton production are dominated by the γj, $\gamma j j$, and $\gamma \gamma$ final states. Additional knowledge incorporated in the determination of fake rates is described in Appendix A.1.

The global fit χ^2 per number of bins is 288.1 / 133 + 27.9, where the last term is the
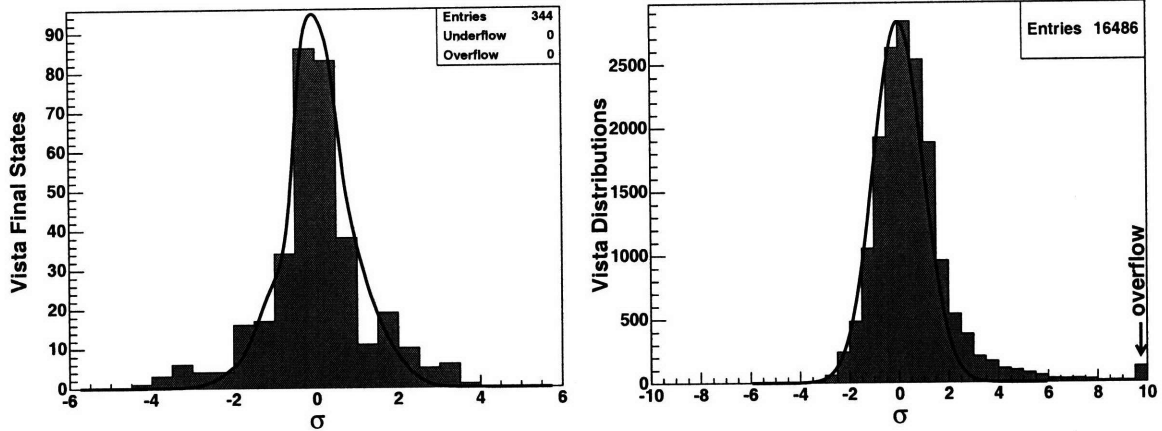
Figure 3-1: Distribution of observed discrepancy between data and the Standard Model prediction, measured in units of standard deviation (σ), shown as the solid (green) histogram, before accounting for the trials factor. The left pane shows the distribution of discrepancies between the total number of events observed and predicted in the 344 populated final states considered. Negative values on the horizontal axis correspond to a deficit of data compared to Standard Model prediction; positive values indicate an excess of data compared to Standard Model prediction. The right pane shows the distribution of discrepancies between the observed and predicted shapes in 16,486 kinematic distributions. Distributions in which the shapes of data and Standard Model prediction are in relative disagreement correspond to large positive σ. The solid (black) curves indicate expected distributions, if the data were truly drawn from the Standard Model background. Interest is focused on the entries in the tails of the left distribution and the high tail of the right distribution.

contribution to the χ^2 from the imposed constraints. A χ^2 per degree of freedom larger than unity is expected, since the limited set of correction factors in this correction model is not expected to provide a complete description of all features of the data. Emphasis is placed on individual outlying discrepancies that may motivate a new physics claim, rather than overall goodness of fit.

Corrections to object identification efficiencies are typically less than 10%; fake rates are consistent with an understanding of the underlying physical mechanisms responsible; k-factors range from slightly less than unity to greater than two for some processes with multiple jets. All values obtained are physically reasonable. Further analysis is provided in Appendix A.4.

With the details of the correction model in place, the complete Standard Model prediction can be obtained. For each Monte Carlo event after detector simulation,

Final State	Data	Background	σ
$3j\tau^{\pm}$	71	113.7 ± 3.6	-2.3
$5j$	1661	1902.9 ± 50.8	-1.7
$2j\tau^{\pm}$	233	296.5 ± 5.6	-1.6
$2j2\tau^{\pm}$	6	27 ± 4.6	-1.4
$be^{\pm}j$	2207	2015.4 ± 28.7	$+1.4$
$3j$, high $\sum p_T$	35436	37294.6 ± 524.3	-1.1
$e^{\pm}3j\not p$	1954	1751.6 ± 42	$+1.1$
$be^{\pm}2j$	798	695.3 ± 13.3	$+1.1$
$3j\not p$, low $\sum p_T$	811	967.5 ± 38.4	-0.8
$e^{\pm}\mu^{\pm}$	26	11.6 ± 1.5	$+0.8$
$e^{\pm}\gamma$	636	551.2 ± 11.2	$+0.7$
$e^{\pm}3j$	28656	27281.5 ± 405.2	$+0.6$
$b5j$	131	95 ± 4.7	$+0.5$
$j2\tau^{\pm}$	50	85.6 ± 8.2	-0.4
$j\tau^{\pm}\tau^{\mp}$	74	125 ± 13.6	-0.4
$b\not p$, low $\sum p_T$	10	29.5 ± 4.6	-0.4
$e^{\pm}j\gamma$	286	369.4 ± 21.1	-0.3
$e^{\pm}j\not p\tau^{\mp}$	29	14.2 ± 1.8	$+0.2$
$2j$, high $\sum p_T$	96502	92437.3 ± 1354.5	$+0.1$
$be^{\pm}3j$	356	298.6 ± 7.7	$+0.1$
$8j$	11	6.1 ± 2.5	
$7j$	57	35.6 ± 4.9	
$6j$	335	298.4 ± 14.7	
$4j$, low $\sum p_T$	39665	40898.8 ± 649.2	
$4j$, high $\sum p_T$	8241	8403.7 ± 144.7	
$4j2\gamma$	38	57.5 ± 11	
$4j\tau^{\pm}$	20	36.9 ± 2.4	
$4j\not p$, low $\sum p_T$	516	525.2 ± 34.5	
$4j\gamma\not p$	28	53.8 ± 11	
$4j\gamma$	3693	3827.2 ± 112.1	
$4j\mu^{\pm}$	576	568.2 ± 26.1	
$4j\mu^{\pm}\not p$	232	224.7 ± 8.5	
$4j\mu^{\pm}\mu^{\mp}$	17	20.1 ± 2.5	
3γ	13	24.2 ± 3	
$3j$, low $\sum p_T$	75894	75939.2 ± 1043.9	
$3j2\gamma$	145	178.1 ± 7.4	
$3j\not p$, high $\sum p_T$	20	30.9 ± 14.4	
$3j\gamma\tau^{\pm}$	13	11 ± 2	
$3j\gamma\not p$	83	102.9 ± 11.1	
$3j\gamma$	11424	11506.4 ± 190.6	
$3j\mu^{\pm}\not p$	1114	1118.7 ± 27.1	
$3j\mu^{\pm}\mu^{\mp}$	61	84.5 ± 9.2	
$3j\mu^{\pm}$	2132	2168.7 ± 64.2	
$3bj$, low $\sum p_T$	14	9.3 ± 1.9	
$2\tau^{\pm}$	316	290.8 ± 24.2	
$2\gamma\not p$	161	176 ± 9.1	
2γ	8482	8349.1 ± 84.1	
$2j$, low $\sum p_T$	93408	92789.5 ± 1138.2	
$2j2\gamma$	645	612.6 ± 18.8	
$2j\tau^{\pm}\tau^{\mp}$	15	25 ± 3.5	
$2j\not p$, low $\sum p_T$	74	106 ± 7.8	
$2j\not p$, high $\sum p_T$	43	37.7 ± 100.2	
$2j\gamma$	33684	33259.9 ± 397.6	
$2j\gamma\tau^{\pm}$	48	41.4 ± 3.4	
$2j\gamma\not p$	403	425.2 ± 29.7	
$2j\mu^{\pm}\not p$	7287	7320.5 ± 118.9	
$2j\mu^{\pm}\gamma\not p$	13	12.6 ± 2.7	
$2j\mu^{\pm}\gamma$	41	35.7 ± 6.1	
$2j\mu^{\pm}\mu^{\mp}$	374	394.2 ± 24.8	

Final State	Data	Background
$2j\mu^{\pm}$	9513	9362.3 ± 166.8
$2e^{\pm}j$	13	9.8 ± 2.2
$2e^{\pm}e^{\mp}$	12	4.8 ± 1.2
$2e^{\pm}$	23	36.1 ± 3.8
$2b$, low $\sum p_T$	327	335.8 ± 7
$2b$, high $\sum p_T$	187	173.1 ± 7.1
$2b3j$, high $\sum p_T$	28	33.5 ± 5.5
$2b2j$, low $\sum p_T$	355	326.3 ± 8.4
$2b2j$, high $\sum p_T$	56	80.2 ± 5
$2b2j\gamma$	16	15.4 ± 3.6
$2b\gamma$	37	31.7 ± 4.8
$2bj$, low $\sum p_T$	415	393.8 ± 9.1
$2bj$, high $\sum p_T$	161	195.8 ± 8.3
$2bj\not p$, low $\sum p_T$	28	23.2 ± 2.6
$2bj\gamma$	25	24.7 ± 4.3
$2be^{\pm}2j\not p$	15	12.3 ± 1.6
$2be^{\pm}2j$	30	30.5 ± 2.5
$2be^{\pm}j$	28	29.1 ± 2.8
$2be^{\pm}$	48	45.2 ± 3.7
$\tau^{\pm}\tau^{\mp}$	498	428.5 ± 22.7
$\gamma\tau^{\pm}$	177	204.4 ± 5.4
$\gamma\not p$	1952	1945.8 ± 77.1
$\mu^{\pm}\tau^{\pm}$	18	19.8 ± 2.3
$\mu^{\pm}\tau^{\mp}$	151	179.1 ± 4.7
$\mu^{\pm}\not p$	321351	320500 ± 3475.5
$\mu^{\pm}\not p\tau^{\mp}$	22	25.8 ± 2.7
$\mu^{\pm}\gamma$	269	285.5 ± 5.9
$\mu^{\pm}\gamma\not p$	269	282.2 ± 6.6
$\mu^{\pm}\mu^{\mp}\not p$	49	61.4 ± 3.5
$\mu^{\pm}\mu^{\mp}\gamma$	32	29.9 ± 2.6
$\mu^{\pm}\mu^{\mp}$	10648	10845.6 ± 96
$j2\gamma$	2196	2200.3 ± 35.2
$j2\gamma\not p$	38	27.3 ± 3.2
$j\tau^{\pm}$	563	585.7 ± 10.2
$j\not p$, low $\sum p_T$	4183	4209.1 ± 56.1
$j\gamma$	49052	48743 ± 546.3
$j\gamma\tau^{\pm}$	106	104 ± 4.1
$j\gamma\not p$	913	965.2 ± 41.5
$j\mu^{\pm}$	33462	34026.7 ± 510.1
$j\mu^{\pm}\tau^{\mp}$	29	37.5 ± 4.5
$j\mu^{\pm}\not p\tau^{\mp}$	10	9.6 ± 2.1
$j\mu^{\pm}\not p$	45728	46316.4 ± 568.2
$j\mu^{\pm}\gamma\not p$	78	69.8 ± 9.9
$j\mu^{\pm}\gamma$	70	98.4 ± 12.1
$j\mu^{\pm}\mu^{\mp}$	1977	2093.3 ± 74.7
$e^{\pm}4j$	7144	6661.9 ± 147.2
$e^{\pm}4j\not p$	403	363 ± 9.9
$e^{\pm}3j\tau^{\mp}$	11	7.6 ± 1.6
$e^{\pm}3j\gamma$	27	21.7 ± 3.4
$e^{\pm}2\gamma$	47	74.5 ± 5
$e^{\pm}2j$	126665	122457 ± 1672.6
$e^{\pm}2j\tau^{\mp}$	53	37.3 ± 3.9
$e^{\pm}2j\tau^{\pm}$	20	24.7 ± 2.3
$e^{\pm}2j\not p$	12451	12130.1 ± 159.4
$e^{\pm}2j\gamma$	101	88.9 ± 6.1
$e^{\pm}\tau^{\mp}$	609	555.9 ± 10.2
$e^{\pm}\tau^{\pm}$	225	211.2 ± 4.7
$e^{\pm}\not p$	476424	479572 ± 5361.2
$e^{\pm}\not p\tau^{\mp}$	48	35 ± 2.7

Final State	Data	Background
$e^{\pm}\not p\tau^{\pm}$	20	18.7 ± 1.9
$e^{\pm}\gamma\not p$	141	144.2 ± 6
$e^{\pm}\mu^{\mp}\not p$	54	42.6 ± 2.7
$e^{\pm}\mu^{\pm}\not p$	13	10.9 ± 1.3
$e^{\pm}\mu^{\mp}$	153	127.6 ± 4.2
$e^{\pm}j$	386880	392614 ± 5031.8
$e^{\pm}j2\gamma$	14	15.9 ± 2.9
$e^{\pm}j\tau^{\pm}$	79	79.3 ± 2.9
$e^{\pm}j\tau^{\mp}$	162	148.8 ± 7.6
$e^{\pm}j\not p$	58648	57391.7 ± 661.6
$e^{\pm}j\gamma\not p$	52	76.2 ± 9
$e^{\pm}j\mu^{\mp}\not p$	22	13.1 ± 1.7
$e^{\pm}j\mu^{\mp}$	28	26.8 ± 2.3
$e^{\pm}e^{\mp}4j$	103	113.5 ± 5.9
$e^{\pm}e^{\mp}3j$	456	473 ± 14.6
$e^{\pm}e^{\mp}2j\not p$	30	39 ± 4.6
$e^{\pm}e^{\mp}2j$	2149	2152 ± 40.1
$e^{\pm}e^{\mp}\tau^{\pm}$	14	11.1 ± 2
$e^{\pm}e^{\mp}\not p$	491	487.9 ± 12
$e^{\pm}e^{\mp}\gamma$	127	132.3 ± 4.2
$e^{\pm}e^{\mp}j$	10726	10669.3 ± 123.5
$e^{\pm}e^{\mp}j\not p$	157	144 ± 11.2
$e^{\pm}e^{\mp}j\gamma$	26	45.6 ± 4.7
$e^{\pm}e^{\mp}$	58344	58575.6 ± 603.9
$b6j$	24	15.5 ± 2.3
$b4j$, low $\sum p_T$	13	9.2 ± 1.8
$b4j$, high $\sum p_T$	464	499.2 ± 12.4
$b3j$, low $\sum p_T$	5354	5285 ± 72.4
$b3j$, high $\sum p_T$	1639	1558.9 ± 24.1
$b3j\not p$, low $\sum p_T$	111	116.8 ± 11.2
$b3j\gamma$	182	194.1 ± 8.8
$b3j\mu^{\pm}\not p$	37	34.1 ± 2
$b3j\mu^{\pm}$	47	52.2 ± 3
$b2\gamma$	15	14.6 ± 2.1
$b2j$, low $\sum p_T$	8812	8576.2 ± 97.9
$b2j$, high $\sum p_T$	4691	4646.2 ± 57.7
$b2j\not p$, low $\sum p_T$	198	209.2 ± 8.3
$b2j\gamma$	429	425.1 ± 13.1
$b2j\mu^{\pm}\not p$	46	40.1 ± 2.7
$b2j\mu^{\pm}$	56	60.6 ± 3.4
$b\tau^{\pm}$	19	19.9 ± 2.2
$b\gamma$	976	1034.8 ± 15.6
$b\gamma\not p$	18	16.7 ± 3.1
$b\mu^{\pm}$	303	263.5 ± 7.9
$b\mu^{\pm}\not p$	204	218.1 ± 6.4
bj, low $\sum p_T$	9060	9275.7 ± 87.8
bj, high $\sum p_T$	7236	7030.8 ± 74
$bj2\gamma$	13	17.6 ± 3.3
$bj\tau^{\pm}$	13	12.9 ± 1.8
$bj\not p$, low $\sum p_T$	53	60.4 ± 19.9
$bj\gamma$	937	989.4 ± 20.6
$bj\gamma\not p$	34	30.5 ± 4
$bj\mu^{\pm}\not p$	104	112.6 ± 4.4
$bj\mu^{\pm}$	173	141.4 ± 4.8
$be^{\pm}3j\not p$	68	52.2 ± 2.2
$be^{\pm}2j\not p$	87	65 ± 3.3
$be^{\pm}\not p$	330	347.2 ± 6.9
$be^{\pm}j\not p$	211	176.6 ± 5
$be^{\pm}e^{\mp}j$	22	34.6 ± 2.6

Table 3.2: A subset of the comparison between data and Standard Model prediction, showing the most discrepant final states and all final states populated with ten or more data events. Final states are labeled according to the number and types of objects present, and whether (high $\sum p_T$) or not (low $\sum p_T$) the summed scalar transverse momentum of all objects in the events exceeds 400 GeV. Final states are ordered according to decreasing discrepancy between the total number of events expected, taking into account the error from Monte Carlo statistics and the total number observed in the data. Final states exhibiting mild discrepancies are shown together with the significance of the discrepancy in units of standard deviations (σ) after accounting for a trials factor corresponding to the number of final states considered. Final states that do not exhibit even mild discrepancies are listed below the horizontal line in inverted alphabetical order. Only Monte Carlo statistical uncertainties on the background prediction are included.
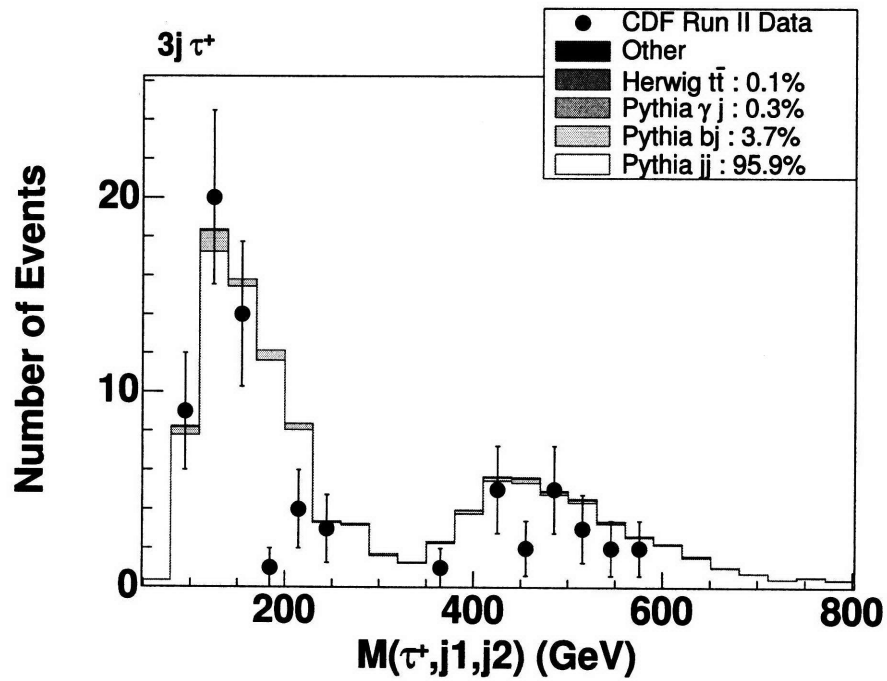
Figure 3-2: The invariant mass of the tau lepton and two leading jets in the final state consisting of three jets and one positively or negatively charged tau. (The VISTA final state naming convention gives the tau lepton a positive charge.) Data are shown as filled circles, with the Standard Model prediction shown as the shaded histogram. This is the most discrepant kinematic distribution in the final state exhibiting the largest population discrepancy.

the event weight is multiplied by the value of the luminosity correction factor and the k-factor for the relevant Standard Model process. The single Monte Carlo event can be misreconstructed in a number of ways, producing a set of Monte Carlo events derived from the original, with weights multiplied by the probability of each misreconstruction. The weight of each resulting event is multiplied by the probability the event satisfies trigger criteria. The resulting Standard Model prediction, corrected as just described, is referred to as "the Standard Model prediction" throughout the rest of this document, with "corrected" implied in all cases.

3.2.6 Results

Data and Standard Model events are partitioned into exclusive final states, depending on the combinations of reconstructed final objects. This partitioning is orthogonal, with each event ending up in one and only one final state, as shown schematically in Fig. 3-3. Data are compared to Standard Model prediction in each final state, considering the total number of events observed and predicted, and the shapes of relevant kinematic distributions.

In a data driven search, it is crucial to explicitly account for the *trials factor*, quantifying the number of places where we checked for an interesting signal. Purely statistical fluctuations at the level of three or more standard deviations are expected to appear, simply because a large number of regions are considered. A reasonably rigorous accounting of this trials factor is possible as long as the measures of interest and the regions to which these measures are applied are specified *a priori*, as is done here. In this analysis a discrepancy at the level of 3σ or greater after accounting for the trials factor (typically corresponding to a discrepancy at the level of 5σ or greater before accounting for the trials factor) is considered "significant." It is worth noting that dedicated searches, checking only a small number of signal regions, typically do not account for any trials factor, simply because it is very difficult to quantify the effect of many people looking for new physics in different ways within the same experiment. For that reason, instead of a mild 3σ, a strong 5σ significance is considered necessary to discover something new in our field. The assumption made silently is
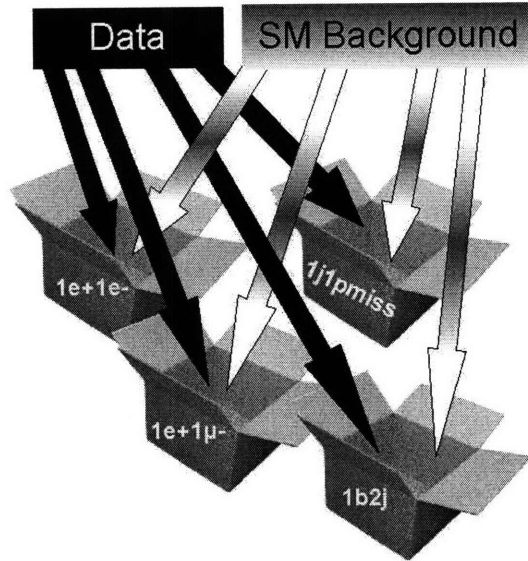
Figure 3-3: VISTA partitioning in final states. Final states can be viewed as boxes, each containing events of one specific final configuration of objects. Final states have not been prescribed, but are created automatically as new types of events appear. In this way, every event, no matter how exotic, stays within the analysis, in the appropriate final state.

that if one observes a 5σ effect in just one attempt, then if one could include somehow the trials factor, the actual significance of the observation would turn out to be still greater than 3σ, therefore convincing. However, in cases where the "new physics" is well-expected (like $t\bar{t}$ or dibosons, which are processes within the Standard Model) "discovery" is claimed even with just 3σ *without* considering the trials factor. Certainly, for physics beyond the Standard Model, a 3σ sans trials factor should not be considered convincing proof of existence.

Discrepancy in the total number of events in a final state (fs) is measured by the Poisson probability p_{fs} that the number of predicted events would fluctuate up to or above (or down to or below) the number of events observed. Since the expected population is known with some uncertainty, its probability density function is convoluted to obtain p_{fs}. To account for the trials factor due to the 344 VISTA final states examined, the quantity $p = 1 - (1 - p_{fs})^{344}$ is calculated for each final state. The result is the probability p of observing a discrepancy corresponding to a probability less than p_{fs} in the total sample studied. This probability p can then be converted

into units of standard deviations by solving for σ such that $\int_\sigma^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = p$ [4]. A final state exhibiting a population discrepancy greater than 3σ after the trials factor is thus accounted for is considered significant.

Many kinematic distributions are considered in each final state, including the transverse momentum, pseudorapidity, detector pseudorapidity, and azimuthal angle of all objects, masses of individual jets and b-jets, invariant masses of all object combinations, transverse masses of object combinations including $p\!\!\!/$, angular separation $\Delta\phi$ and ΔR of all object pairs, and several other more specialized variables. A Kolmogorov-Smirnov (KS) test is used to quantify the difference in shape of each kinematic distribution between data and Standard Model prediction. As with populations, a trials factor is assessed to account for the 16,486 distributions examined, and the resulting probability is converted into units of standard deviations. A distribution with KS statistic greater than 0.02 and probability corresponding to greater than 3σ after assessing the trials factor is considered significant.

Table 3.2 shows a subset of the VISTA comparison of data to Standard Model prediction. Shown are all final states containing ten or more data events, with the most discrepant final states in population heading the list. After accounting for the trials factor, no final state has a statistically significant ($> 3\sigma$) population discrepancy. The most discrepant final state ($3j\,\tau^\pm$) contains 71 data events and 113.7 ± 3.6 events expected from the Standard Model. The Poisson probability for 113.7 ± 3.6 expected events to result in 71 or fewer events observed in this final state is 2.8×10^{-5}, corresponding to an entry at -4.03σ in Fig. 3-1. The probability for one or more of the 344 populated final states considered to display disagreement in population corresponding to a probability less than 2.8×10^{-5} is 1%. The $3j\,\tau^\pm$ population discrepancy is thus not statistically significant. The most discrepant kinematic distribution in this final state is the invariant mass of the tau lepton and the two highest transverse momentum jets, shown in Fig. 3-2.

---

[4]Final states for which $p > 0.5$ after accounting for the trials factor are not even mildly interesting, and the corresponding σ after accounting for the trials factor is not quoted. For the mildly interesting final states with $p < 0.5$ after accounting for the trials factor, σ is quoted as positive if the number of observed data events exceeds the Standard Model prediction, and negative if the number of observed data events is less than the Standard Model prediction.

The six final states with largest population discrepancy are $3j\,\tau$, $5j$, $2j\,\tau$, $2j\,2\tau$, $b\,e\,j$, and the low-p_T $3j$ final state, with $b\,e\,j$ being the only one of these six to exhibit an excess of data. The $3j\,\tau$, $2j\,\tau$, and $2j\,2\tau$ final states appear to reflect an incomplete understanding of the rate of jets faking taus ($p(j \to \tau)$) as a function of the number of jets in the event, at the level of $\sim 30\%$ difference between the total number of observed and predicted events in the most populated of these final states. The value of $p(j \to \tau)$ is primarily determined by the $j\,\tau$ final state. Interestingly, although the underlying physical mechanism for $p(j \to e)$ is very similar to that for $p(j \to \tau)$, as discussed in Appendix A.1, a significant dependence on the presence of additional jets is not observed for $p(j \to e)$.

The $5j$ discrepancy results from a tension with the $e\,4j$ final state, whose dominant contribution comes from $5j$ production convoluted with $p(j \to e)$. The low-p_T $3j$ discrepancy results from a tension with the $e\,2j$ final state, whose dominant contribution comes from $3j$ production convoluted with $p(j \to e)$. The $b\,e\,j$ final state is predominantly $3j$ production convoluted with $p(j \to b)$ and $p(j \to e)$; this discrepancy also arises from a tension with the low-p_T $3j$ and $e\,2j$ final states. The $b\,e\,j$ final state is the VISTA final state in which the largest excess of data over Standard Model prediction is seen. The fraction of hypothetical similar CDF experiments that would produce a VISTA normalization excess as significant as the excess observed in this final state is 8%. The $5j$, $b\,e\,j$, and low-p_T $3j$ discrepancies correspond to a difference of $\sim 10\%$ between the total number of observed and predicted events in these final states.

Figure 3-1 summarizes in a histogram the measured discrepancies between data and the Standard Model prediction for CDF high-p_T final state populations and kinematic distributions. Values in this figure represent individual discrepancies, and do not account for the trials factor associated with examining many possibilities.

Of the 16,484 kinematic distributions considered, 384 distributions are found to correspond to a discrepancy greater than 3σ after accounting for the trials factor, entering with a KS probability of roughly 5σ or greater in Fig. 3-1. Of these 384 discrepant distributions, 312 are attributed to modeling parton radiation, deriving
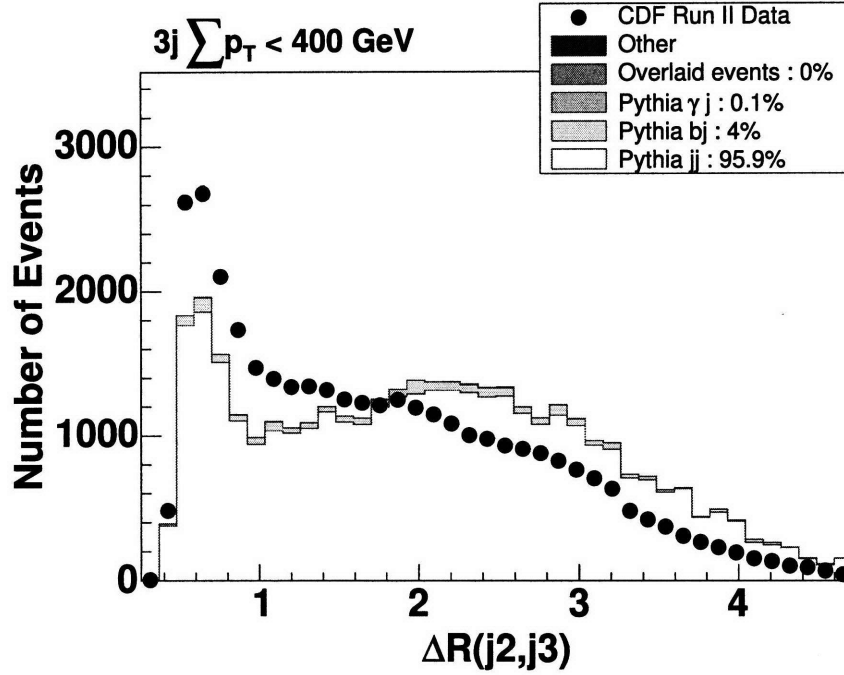
Figure 3-4: A shape discrepancy highlighted by VISTA in the final state consisting of exactly three reconstructed jets with $|\eta| < 2.5$ and $p_T > 17$ GeV, and with one of the jets satisfying $|\eta| < 1$ and $p_T > 40$ GeV. This distribution illustrates the effect underlying most of the VISTA shape discrepancies. Filled circles show CDF data, with the shaded histogram showing the prediction of PYTHIA. The discrepancy is clearly statistically significant, with statistical error bars smaller than the size of the data points. The vertical axis shows the number of events per bin, with the horizontal axis showing the angular separation ($\Delta R = \sqrt{\Delta\eta^2 + \delta\phi^2}$) between the second and third jets, where the jets are ordered according to decreasing transverse momentum. In the region $\Delta R(j_2, j_3) \gtrsim 2$, populated primarily by initial state radiation, the Standard Model prediction can to some extent be adjusted. The region $\Delta R(j_2, j_3) \lesssim 2$ is dominated by final state radiation, the description of which is constrained by data from LEP 1.
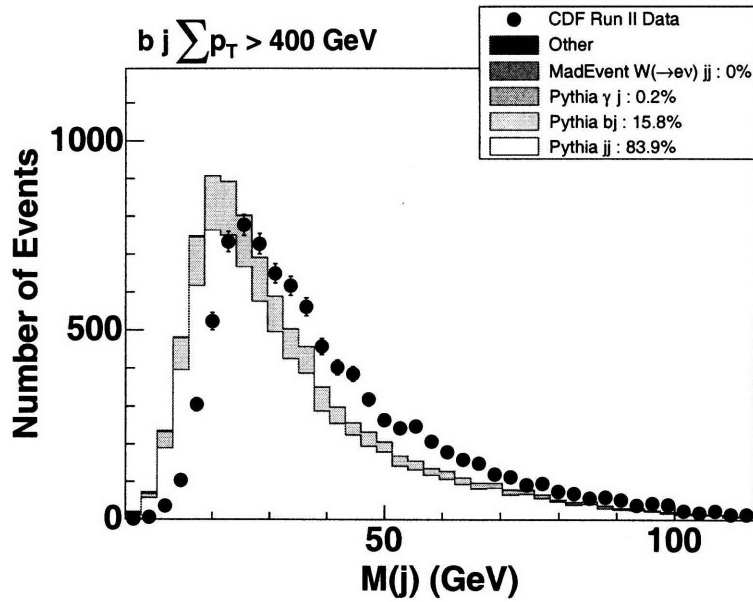
Figure 3-5: The jet mass distribution in the bj final state with $\sum p_T > 400$ GeV. The $3j$ $\Delta R(j_2, j_3)$ discrepancy illustrated in Fig. 3-4 manifests itself also by producing jets more massive in data than predicted by PYTHIA's showering algorithm. The mass of a jet is determined by treating energy deposited in each calorimeter tower as a massless 4-vector, summing the 4-vectors of all towers within the jet, and computing the mass of the resulting (massive) 4-vector.
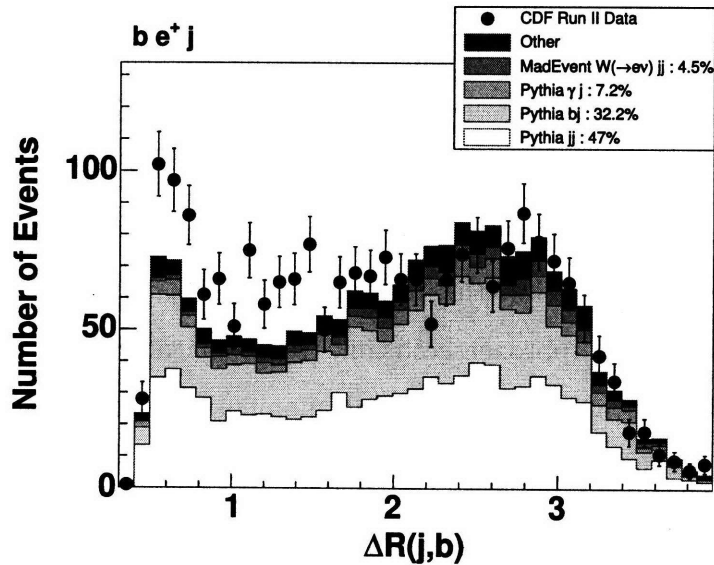
Figure 3-6: The distribution of ΔR between the jet and b-tagged jet in the final state $b\,e\,j$. The primary Standard Model contribution to this final state is QCD three jet production with one jet misreconstructed as an electron. The similarity to the $3j$ $\Delta R(j_2, j_3)$ discrepancy illustrated in Fig. 3-4 in the region $\Delta R(j, b) < 2$ is clear. Less clear is the underlying explanation for the difference with respect to Fig. 3-4 in the region $\Delta R(j, b) > 2$.

from the $3j$ $\Delta R(j_2, j_3)$ discrepancy shown in Fig. 3-4, with 186 of these 312 shape discrepancies pointing out that individual jet masses are larger in data than in the prediction, as shown in Fig. 3-5. In the literature, that the same effect was observed (but not emphasized) by both CDF [64, 65] and DØ [66] in Tevatron Run I. The $3j$ $\Delta R(j_2, j_3)$ and jet mass discrepancies appear to be two different views of a single underlying discrepancy, noting that two sufficiently nearby distinct jets correspond to a pattern of calorimetric energy deposits similar to a single massive jet. The underlying $3j$ $\Delta R(j_2, j_3)$ discrepancy is manifest in many other final states. The final state bej, arising primarily from QCD production of three jets with one misreconstructed as an electron, shows a similar discrepancy in $\Delta R(j, b)$ in Fig. 3-6.

While these discrepancies are clearly statistically significant, basing a new physics claim on them would be premature. In the kinematic regime of the discrepancy, different algorithms to match exact leading order calculations with a parton shower lead to different predictions [67]. Newer predictions have not been systematically compared to LEP 1 data, which provide constraints on parton showering reflected in PYTHIA's tuning. Further investigation into obtaining an adequate QCD-based description of this discrepancy continues.

An additional 59 discrepant distributions reflect an inadequate modeling of the overall transverse boost of the system. The overall transverse boost of the primary physics objects in the event is attributed to two sources: the intrinsic Fermi motion of the colliding partons within the proton, and soft or collinear radiation of the colliding partons as they approach collision. Together these effects are here referred to as "intrinsic k_T," representing an overall momentum kick to the hard scattering. Further discussion appears in Appendix A.2.3.

The remaining 13 discrepant distributions are seen to be due to the coarseness of the VISTA correction model. Most of these discrepancies, which are at the level of 10% or less when expressed as $(\text{data} - \text{theory})/\text{theory}$, arise from modeling most fake rates as independent of transverse momentum.

In summary, this global analysis of the bulk features of the high-p_T data has not yielded a discrepancy motivating a new physics claim. There are no statistically sig-

nificant population discrepancies in the 344 populated final states considered, and although there are several statistically significant discrepancies among the 16,486 kinematic distributions investigated, the nature of these discrepancies makes it difficult to use them to support a new physics claim.

This global analysis of course can not conclude with certainty that there is no new physics hiding in the CDF data. The VISTA population and shape statistics may be insensitive to a small excess of events appearing at large $\sum p_T$ in a highly populated final state. For such signals, different probes are required. SLEUTH, and the Bump Hunter, which was added in the second round of this analysis, serve this purpose.

3.3 SLEUTH

Taking a broad view of proposed models that might extend the Standard Model, something common is noted: nearly all predict an excess of events at high p_T, concentrated in a particular final state. This feature is exploited by SLEUTH [68]. SLEUTH is quasi model independent, where "quasi" refers to the assumption that the first sign of new physics will appear as an excess of events in some final state at large summed scalar transverse momentum ($\sum p_T$).

The first version of SLEUTH was essentially developed by DØ in Tevatron Run I [69, 70, 71], and subsequently improved by H1 in HERA Run I [72], with small modifications.

SLEUTH relies on the following assumptions for new physics:

1. The data can be categorized into exclusive final states in such a way that any signature of new physics is apt to appear predominantly in one of these final states.

2. New physics will appear with objects at high summed transverse momentum ($\sum p_T$) relative to Standard Model and instrumental background.

3. New physics will appear as an excess of data over Standard Model and instrumental background.

To the extent that the above are true, SLEUTH would be more sensitive to a new physics signal.

3.3.1 Algorithm

The SLEUTH algorithm consists of three steps, following the above three assumptions.

Final states

In the first step of the algorithm, all events are placed into exclusive final states as in VISTA, with the following modifications.

- Jets are identified as pairs, rather than individually, to reduce the total number of final states and to keep signal events with one additional radiated gluon within the same final state. Final state names include "$n\ jj$" if n jet pairs are identified, with possibly one unpaired jet assumed to have originated from a radiated gluon.

- The present understanding of quark flavor suggests that b quarks should be produced in pairs. Bottom quarks are identified as pairs, rather than individually, to increase the robustness of identification and to reduce the total number of final states. Final state names include "$n\ b\bar{b}$" if n b pairs are identified.

- Final states related through global charge conjugation are considered to be equivalent. Thus $e^+e^-\gamma$ is a different final state than $e^+e^+\gamma$, but $e^+e^+\gamma$ and $e^-e^-\gamma$ together make up a single SLEUTH final state.

- Final states related through global interchange of the first and second generation are considered to be equivalent. Thus $e^+\not{p}\gamma$ and $\mu^+\not{p}\gamma$ together make up a single SLEUTH final state. The decision to treat third generation objects (b quarks and τ leptons) differently from first and second generation objects reflects theoretical prejudice that the third generation may be special, and the experimental ability (in the case of b quarks) and experimental challenge (in the case of τ leptons) in the identification of third generation objects.

The symbol ℓ is used to denote electron or muon. The symbol W is used in naming final states containing one electron or muon, significant missing momentum, and perhaps other non-leptonic objects. Thus the final states $e^+\not{p}\gamma$, $e^-\not{p}\gamma$, $\mu^+\not{p}\gamma$, and $\mu^-\not{p}\gamma$ are combined into the SLEUTH final state $W\gamma$. A table showing the relationship between VISTA and SLEUTH final states is provided in Appendix A.5.1.

Summed Transverse Momentum Variable

The second step of the algorithm considers a single variable in each exclusive final state: the summed scalar transverse momentum of all objects in the event $(\sum p_T)$. Assuming momentum conservation in the plane transverse to the axis of the colliding beams,

$$\sum_i \vec{p}_i + \overrightarrow{\mathrm{uncl}} + \vec{\not{p}} = \vec{0}, \tag{3.3}$$

where the sum over i represents a sum over all identified objects in the event, the ith object has momentum $\vec{p}_i$, $\overrightarrow{\mathrm{uncl}}$ denotes the vector sum of all momentum visible in the detector but not clustered into an identified object, $\vec{\not{p}}$ denotes the missing momentum, and the equation is a two-component vector equality for the components of the momentum along the two spatial directions transverse to the axis of the colliding beams. The SLEUTH variable $\sum p_T$ is then defined by

$$\sum p_T \equiv \sum_i |\vec{p}_i| + \left|\overrightarrow{\mathrm{uncl}}\right| + \left|\vec{\not{p}}\right|, \tag{3.4}$$

where only the momentum components transverse to the axis of the colliding beams are considered when computing magnitudes.

Regions

The algorithm's third step involves searching for regions in which more events are seen in the data than expected from Standard Model and instrumental background. This search is performed in the variable $\sum p_T$ defined in the second step of the algorithm, for each of the exclusive final states defined in the first step.

The steps of the search can be sketched as follows.

- In each final state, the regions considered are the one dimensional intervals in $\sum p_T$ extending from each data point up to infinity. A region is required to contain at least three data events, as described in Appendix A.5.

- In a particular final state, the data point with the dth largest value of $\sum p_T$ defines an interval in the variable $\sum p_T$ extending from this data point up to infinity. This semi-infinite interval contains d data events. The Standard Model prediction in this interval, estimated from the VISTA comparison, integrates to b predicted events. In this final state, the interest of the dth region is defined as the Poisson probability $p\text{-val} = \sum_{i=d}^{\infty} \frac{b^i}{i!} e^{-b}$ that the Standard Model background b would fluctuate up to or above the observed number of data events d in this region. The most interesting region in this final state is the one with smallest Poisson probability ($p\text{-val}_{\min}$).

- For this final state, pseudo experiments are generated, with pseudo data pulled from the Standard Model background. For each pseudo experiment, the interest of the most interesting region is calculated. An ensemble of pseudo experiments determines the fraction $\mathcal{P}$ of pseudo experiments in this final state in which the most interesting region is more interesting than the most interesting region in this final state observed in the data. Namely, for each final state, $\mathcal{P}$ is the fraction of pseudo-data distributions, pulled from the Standard Model expectation, where $p\text{-val}_{\min}$ was smaller than the $p\text{-val}_{\min}$ observed in the actual data distribution. If there is no new physics in this final state, $\mathcal{P}$ is expected to be a random number pulled from a uniform distribution in the unit interval[5]. If

---

[5] There is a small caveat, for final states with small expected population: We require at least 3 data in a $\sum p_T$ tail. If $d < 3$, then $p\text{-val} = 1$ by convention, i.e. the tail is totally uninteresting by definition. Apart from $p\text{-val} = 1$, the most uninteresting a tail can possibly be is to have exactly $d = 3$ and as big a background b as possible. So, the largest $p\text{-val}$ attainable for a final state with total background b_{tot}, before we run into $p\text{-val} = 1$, is $p\text{-val}_{\max} = \sum_{i=3}^{\infty} \frac{b_{\text{tot}}^i}{i!} e^{-b_{\text{tot}}}$. I will show now that $\mathcal{P}$ can not assume values between $p\text{-val}_{\max}$ and 1, therefore its distribution is not exactly uniform, but has a gap: If the actual $p\text{-val}_{\min}$ were equal to $p\text{-val}_{\max}$, then the fraction of pseudo-data distributions which would have $p\text{-val}_{\min} > p\text{-val}_{\max}$ would be $\sum_{i=0}^{2} \frac{b_{\text{tot}}^i}{i!} e^{-b_{\text{tot}}}$, because they would be given $p\text{-val}_{\min} = 1$ by convention. The rest of the pseudo-data distributions

there is new physics in this final state, $\mathcal{P}$ is expected to be small.

- Looping over all final states, $\mathcal{P}$ is computed for each final state. The minimum of these values is denoted $\mathcal{P}_{\min}$. Let $\mathcal{R}$ be the most interesting region in the final state with the smallest $\mathcal{P}$.

- The interest of the most interesting region $\mathcal{R}$ in the most interesting final state is defined as $\tilde{\mathcal{P}} = 1 - \prod_a (1 - \hat{p}_a)$, where the product is over all SLEUTH final states a, and $\hat{p}_a$ is the lesser of $\mathcal{P}_{\min}$ and the probability for the total number of events predicted by the Standard Model in the final state a to fluctuate up to or above three data events. The quantity $\tilde{\mathcal{P}}$ is the fraction of hypothetical similar CDF experiments that would produce a final state with $\mathcal{P} < \mathcal{P}_{\min}$[6].The range of $\tilde{\mathcal{P}}$ is the unit interval. If the data are distributed according to our Standard Model prediction, $\tilde{\mathcal{P}}$ is expected to be a random number pulled from a uniform distribution in the unit interval, as was also demonstrated experimentally (see Fig. 3-7). If new physics is present, $\tilde{\mathcal{P}}$ is expected to be small.

An alternative statistic to $\tilde{\mathcal{P}}$ was first implemented in this analysis. The new measure of significance, $\tilde{p}$-val, is the probability that, in a pseudo-experiment, at least one $\sum p_T$ tail, in any final state, would have a p-val smaller than the smallest p-val found among all tails and all final states in the data. In other words, $\tilde{p}$-val is

---

would have p-val$_{\min}$ $\leq$ p-val$_{\max}$, therefore $\mathcal{P} = 1 - \sum_{i=0}^{2} \frac{b_{\text{tot}}^i}{i!} e^{-b_{\text{tot}}} = p$-val$_{\max}$. For any actual p-val$_{\min}$ $<$ p-val$_{\max}$, $\mathcal{P}$ will be even smaller than p-val$_{\max}$, as it will be more challenging for a pseudo-data distribution to exceed that p-val$_{\min}$. If p-val$_{\min} = 1$, which has probability $\sum_{i=0}^{2} \frac{b_{\text{tot}}^i}{i!} e^{-b_{\text{tot}}}$, then all pseudo-data distributions would be at least as interesting, therefore $\mathcal{P} = 1$. Therefore, the distribution of $\mathcal{P}$ has a Kronecker δ term at 1, multiplied by $\sum_{i=0}^{2} \frac{b_{\text{tot}}^i}{i!} e^{-b_{\text{tot}}}$, and the rest is spread at values $\mathcal{P} \leq p$-val$_{\max}$. This gap in possible $\mathcal{P}$ values shrinks as $b_{\text{tot}} \gg 3$, and practically vanishes for $b_{\text{tot}} \gtrsim 10$.

[6] This point deserves some explanation to become more obvious. We have N final states, and we want to find the probability that one or more of them would give a $\mathcal{P}$ smaller than the observed $\mathcal{P}_{\min}$. If the expected distribution of $\mathcal{P}$ were exactly uniform for all N final states, without the gap discussed in footnote 5, then each final state would have equal probability $\mathcal{P}_{\min}$ to give $\mathcal{P} \leq \mathcal{P}_{\min}$. In that simple case, we would just need to define $\tilde{\mathcal{P}} \equiv 1 - \prod_a (1 - \mathcal{P}_{\min}) = 1 - (1 - \mathcal{P}_{\min})^N$. However, depending on the total background b_{tot}, $\mathcal{P}$ is not distributed exactly uniformly for small final states, which gives rise to two possibilities: If for a final state the gap starts at a p-val$_{\max} \geq \mathcal{P}_{\min}$, then the probability that this final state would give $\mathcal{P} \leq \mathcal{P}_{\min}$ is simply $\mathcal{P}_{\min}$. If, however, b_{tot} is such that p-val$_{\max} \leq \mathcal{P}_{\min}$, then $\mathcal{P}_{\min}$ falls in the gap, and then that final state has probability $\sum_{i=3}^{\infty} \frac{b_{\text{tot}}^i}{i!} e^{-b_{\text{tot}}}$ to return $\mathcal{P} \leq p$-val$_{\max} < \mathcal{P}_{\min}$, as explained in footnote 5. This complication necessitates the introduction of $\hat{p}_a$ in $\tilde{\mathcal{P}}$, to treat appropriately the two possible cases.
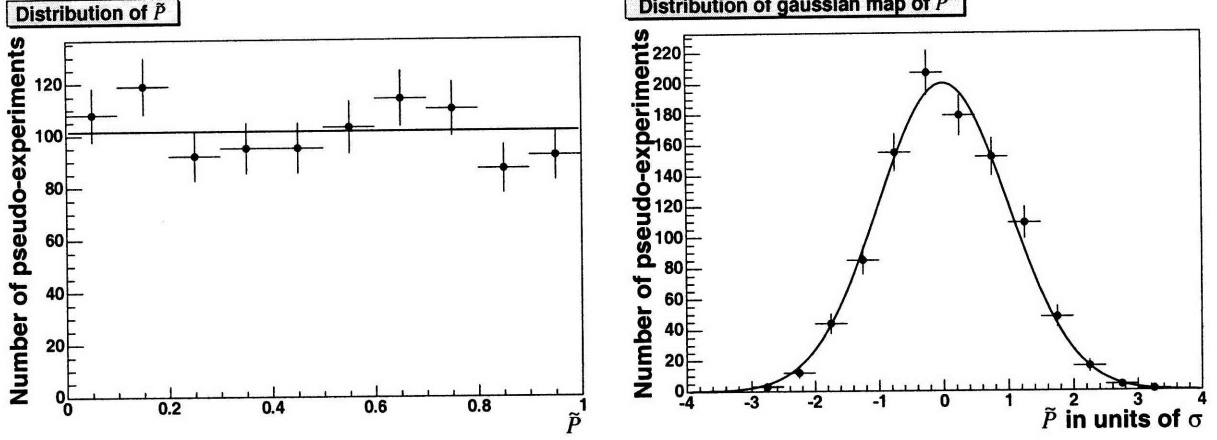
Figure 3-7: Distribution of expected values of $\tilde{\mathcal{P}}$ in ~ 1000 pseudo-experiments, where pseudo-data are pulled from the Standard Model $\sum p_T$ distributions. On the right is shown the distribution of the same values of $\tilde{\mathcal{P}}$ translated into standard deviations (σ) through the transformation: $\tilde{\mathcal{P}} = \int_\sigma^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. The expected distribution is consistent with a uniform distribution in the interval $[0, 1]$, represented by the black curve.

the probability that in a pseudo-experiment some $\sum p_T$ tail would be more significant than the globally most significant tail found in the data. The definition of $\tilde{p}$-val is

$$\tilde{p}\text{-val} \equiv 1 - \prod_a \left(1 - \mathcal{P}_{(a, p\text{-val}_{\min})}\right), \tag{3.5}$$

where a denotes a final state, $\mathcal{P}_{(a, p\text{-val}_{\min})}$ is the probability for final state a to have (in a pseudo-experiment) a $\sum p_T$ tail of p-val $\leq p$-val$_{\min}$, and p-val$_{\min}$ is the smallest p-val found among all tails in all final states using data. Note that, unlike when defining $\mathcal{P}$ for a final state a, where p-val$_{\min}$ was the smallest p-val within that final state, this p-val$_{\min}$ going into $\mathcal{P}_{(a, p\text{-val}_{\min})}$ is the global smallest p-val. Therefore, for a final state a, $\mathcal{P}_{(a, p\text{-val}_{\min})}$ is not the same as the $\mathcal{P}$ defined earlier for each final state, because there $\mathcal{P}$ was the probability for a final state to exceed in significance its own most interesting tail, while $\mathcal{P}_{(a, p\text{-val}_{\min})}$ is the probability for final state a to exceed in significance the *globally* most interesting tail, which may or may not be within a.

The qualitative difference between $\tilde{p}$-val and traditional $\tilde{\mathcal{P}}$ is that $\tilde{\mathcal{P}}$ focusses on fluctuations producing a smaller $\mathcal{P}$ than the $\mathcal{P}_{\min}$ observed in the data, while $\tilde{p}$-val focusses on fluctuations producing a smaller p-val. The $\mathcal{P}$ of a final state depends not

only on the significance (p-val$_{\mathrm{min}}$) of the most interesting tail therein, but also on the total expected population of the final state where that tail is: A $\sum p_T$ tail of numerically identical p-val$_{\mathrm{min}}$, but found in a final state with larger expected background, results into larger $\mathcal{P}$, because bigger population means more pseudo-data, hence more $\sum p_T$ tails, hence more chances to have p-val $\leq p$-val$_{\mathrm{min}}$. So, $\mathcal{P}$ is not a measure of the significance of a tail *per se*, but rather of a whole $\sum p_T$ distribution. Whether to use $\tilde{\mathcal{P}}$ or $\tilde{p}$-val is a matter of preference. $\tilde{p}$-val is more intuitive, because it quantifies the significance of $\sum p_T$ tails, which are fundamentally the features SLEUTH detects, while $\tilde{\mathcal{P}}$ quantifies the significance of whole $\sum p_T$ distributions from the view-point of their own $\sum p_T$ excesses. Since $\tilde{\mathcal{P}}$ was invented first and has been part of SLEUTH since its conception, its use was continued in this analysis.

Output

The output of the algorithm is the most interesting region $\mathcal{R}$ observed in the final state with the smallest $\mathcal{P}$, and a number $\tilde{\mathcal{P}}$ quantifying the interest of $\mathcal{R}$[7]. A reasonable threshold for discovery is $\tilde{\mathcal{P}} \lesssim 0.001$, which corresponds loosely to a local 5σ effect after the trials factor is accounted for[8].

Although no integration over systematic errors is performed in computing $\tilde{\mathcal{P}}$, systematic uncertainties do affect the final SLEUTH result. If SLEUTH highlights a discrepancy in a particular final state, explanations in terms of a correction to the background estimate are considered. This process necessarily requires physics judgement. A reasonable explanation of a SLEUTH discrepancy in terms of an inadequacy in the modeling of the detector response or Standard Model prediction that is consistent with external information is fed back into the VISTA correction model and tested for global consistency. In this way, plausible explanations for discrepancies observed by SLEUTH are incorporated into the VISTA correction model. This iteration con-

---

[7]If SLEUTH used $\tilde{p}$-val instead of $\tilde{\mathcal{P}}$, then the most interesting tail $\mathcal{R}$ would be the one with the globally smallest p-val. That region may happen to be the same with the most interesting region within the final state with of smallest $\mathcal{P}$, but it doesn't have to.

[8]That is empirically confirmed in sensitivity tests (Sec. 3.3.2), where it was observed that the $\tilde{\mathcal{P}}$ discovery threshold is met approximately at the same time when p-val$_{\mathrm{min}} \simeq \int_5^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

tinues until either all reasonable explanations for a significant SLEUTH discrepancy are exhausted, resulting in a possible new physics claim, or no significant SLEUTH discrepancy remains.

3.3.2 Sensitivity

Two important questions must be asked:

- Will SLEUTH find nothing if there is nothing to be found?

- Will SLEUTH find something if there is something to be found?

If there is nothing to be found, SLEUTH will find nothing 999 times out of 1000, given a uniform distribution of $\tilde{\mathcal{P}}$ and a discovery threshold of $\tilde{\mathcal{P}} \lesssim 0.001$. The uniform distribution of $\tilde{\mathcal{P}}$ in the absence of new physics is illustrated in Fig. 3-7. SLEUTH will of course return spurious signals if provided improperly modeled backgrounds. The algorithm directly addresses the issue of whether an observed hint is due to a statistical fluctuation. SLEUTH itself is unable to address systematic mismeasurement or incorrect modeling, but is useful in bringing these to attention.

The answer to the second question depends on the degree to which the new physics satisfies the three assumptions on which SLEUTH is based: new physics will appear predominantly in one final state, at high summed scalar transverse momentum, and as an excess of data over Standard Model prediction.

Known Standard Model processes

Consideration of specific Standard Model processes can provide intuition for SLEUTH's sensitivity to new physics. This section tests SLEUTH's sensitivity to the production of top quark pairs, W boson pairs, single top, and the Higgs boson.

Top quark pairs. Top quark pair production results in two b jets and two W bosons, each of which may decay leptonically or hadronically. The W branching ratios are such that this signal predominantly populates the SLEUTH final state $W b \bar{b} j j$, where "W" denotes an electron or muon and significant missing momentum. Although
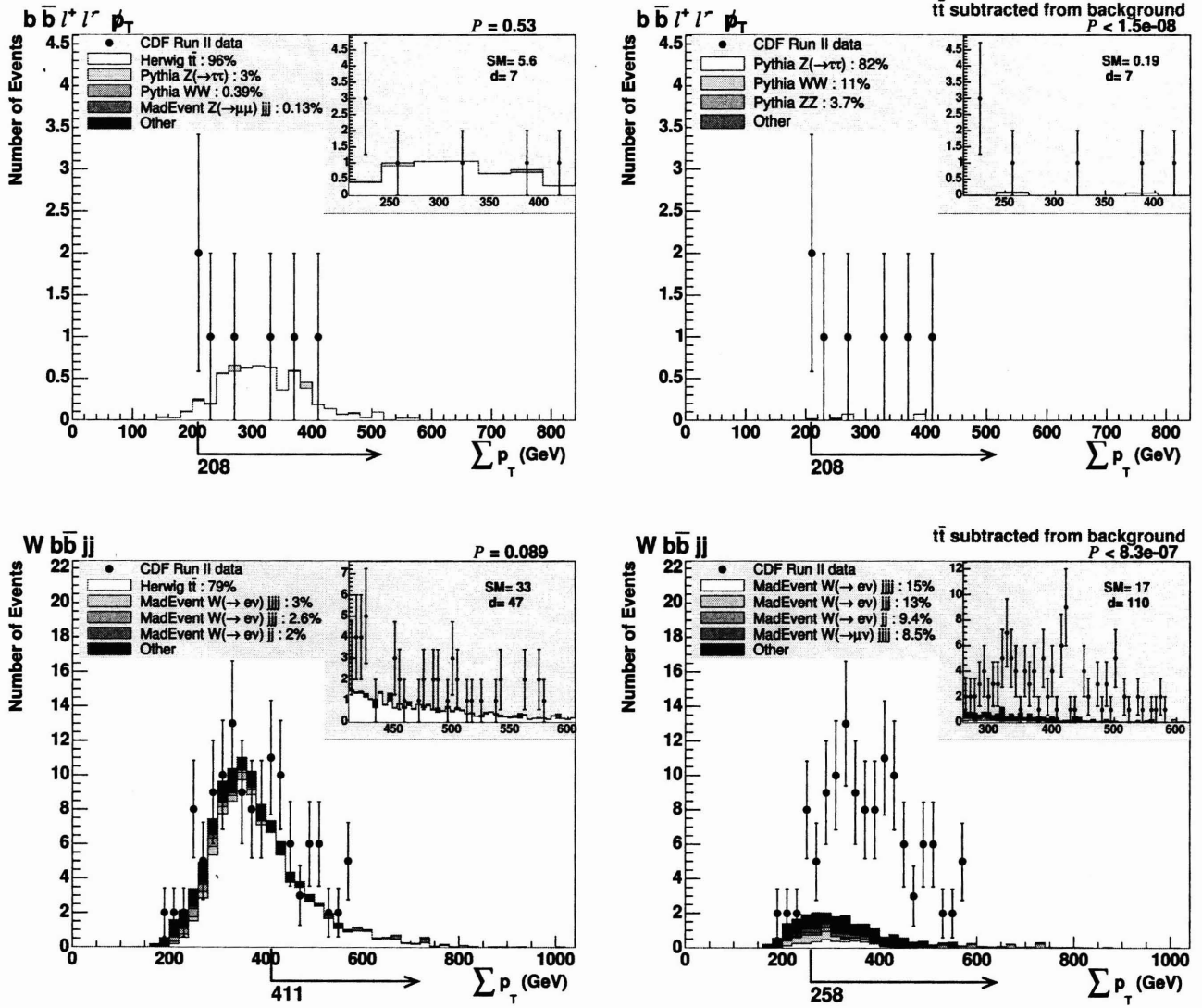
Figure 3-8: (Top left) The SLEUTH final state $b\bar{b}\ell^+\ell^-\displaystyle{\not}p$, consisting of events with one electron and one muon of opposite sign, missing momentum, and two or three jets, one or two of which are b-tagged. Data corresponding to 927 pb^{-1} are shown as filled circles; the Standard Model prediction is shown as the shaded histogram. (Top right) The same final state with $t\bar{t}$ subtracted from the Standard Model prediction. (Bottom row) The SLEUTH final state $Wb\bar{b}jj$, with the Standard Model $t\bar{t}$ contribution included (lower left) and removed (lower right). Significant discrepancies far surpassing SLEUTH's discovery threshold are observed in these final states with $t\bar{t}$ removed from the Standard Model background estimate.
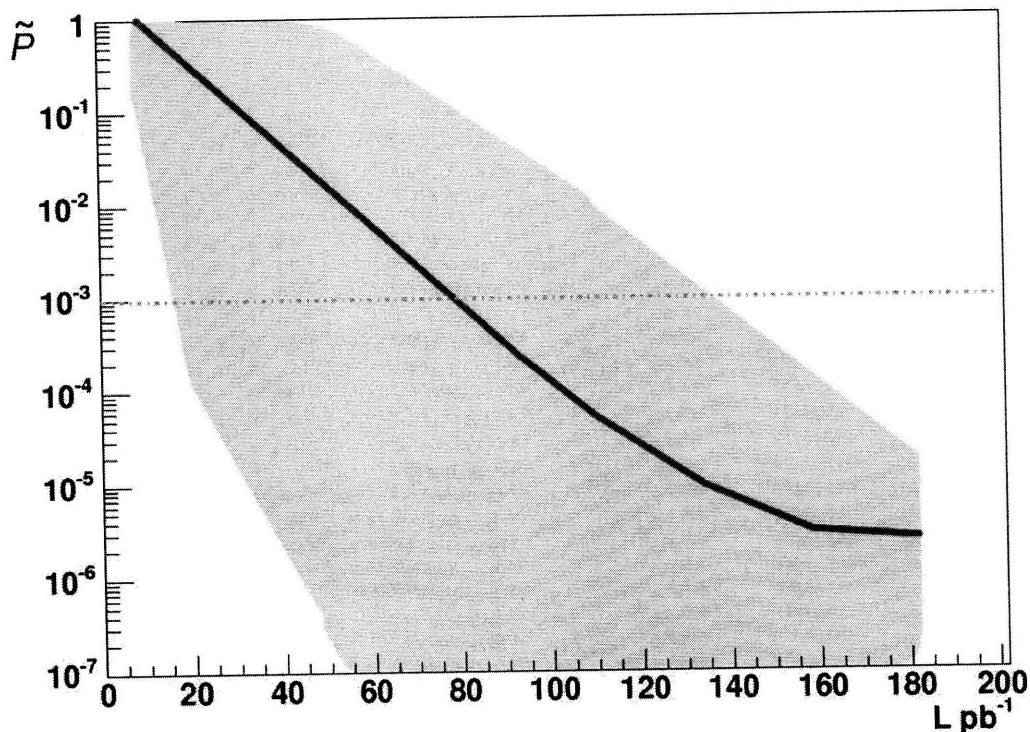
Figure 3-9: SLEUTH's $\tilde{\mathcal{P}}$ as a function of assumed integrated luminosity, with top quark pair production removed from the Standard Model background estimate. The horizontal axis shows integrated luminosity, in units of pb^{-1}. The vertical axis shows SLEUTH's $\tilde{\mathcal{P}}$. With Standard Model $t\bar{t}$ production omitted from the background estimate and actual data including $t\bar{t}$ production, SLEUTH's $\tilde{\mathcal{P}}$ decreases with increasing integrated luminosity, shown as the solid (green) line, crossing at roughly 80 pb^{-1} the discovery threshold of $\tilde{\mathcal{P}} < 0.001$, shown as the horizontal dashed (gray) line. The shaded (yellow) band shows the range of values of $\tilde{\mathcal{P}}$ obtained in a number of trials, with the width of the band resulting from the statistical fluctuations of individual top quark events.

Model	Description	Sensitivity
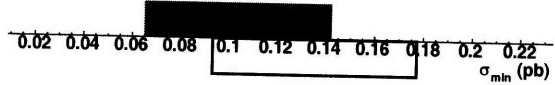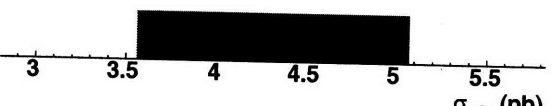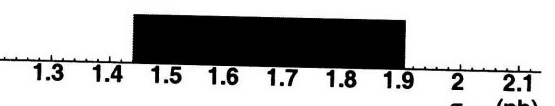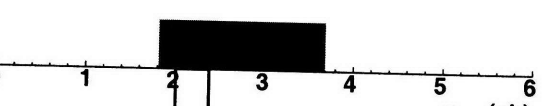
Model	Description	Sensitivity
1	GMSB, $\Lambda = 82.6$ GeV, $\tan\beta = 15$, $\mu > 0$, with one messenger of $M = 2\Lambda$.	
2	$Z' \to \ell^+\ell^-$, $m_{Z'} = 250$ GeV, with standard model couplings to leptons.	
3	$Z' \to q\bar{q}$, $m_{Z'} = 700$ GeV, with standard model couplings to quarks.	
4	$Z' \to q\bar{q}$, $m_{Z'} = 1$ TeV, with standard model couplings to quarks.	
5	$Z' \to t\bar{t}$, $m_{Z'} = 500$ GeV, with standard model couplings to $t\bar{t}$.	

Table 3.3: Summary of SLEUTH's sensitivity to several new physics models, expressed in terms of the minimum production cross section needed for discovery with 927 pb^{-1}. Where available, a comparison is made to the sensitivity of a dedicated search for this model. The solid (red) box represents SLEUTH's sensitivity, and the open (white) box represents the sensitivity of the dedicated analysis. Systematic uncertainties are not included in the sensitivity calculation. The width of each box shows typical variation under fluctuation of data statistics. In Models 3 and 4, there is no targeted analysis available for comparison.
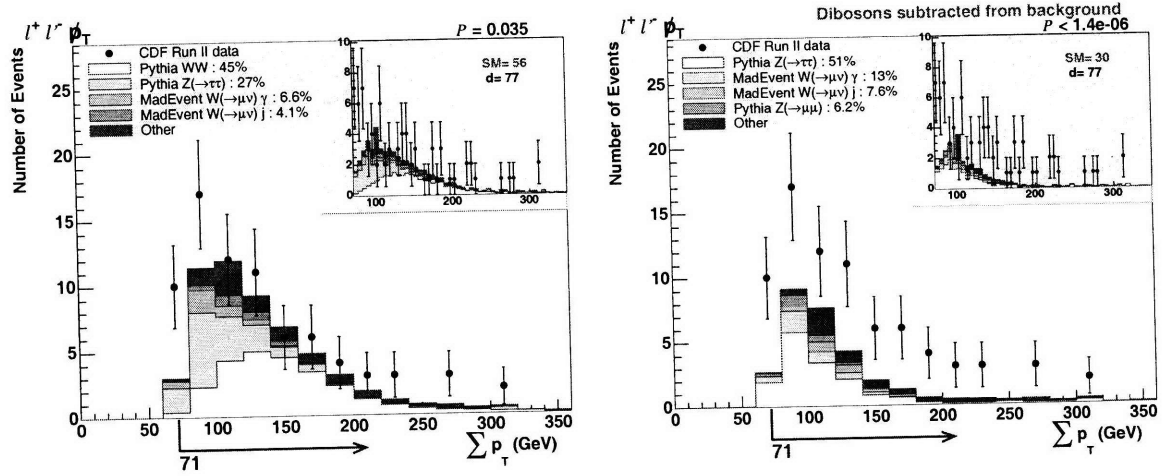
Figure 3-10: (Left) The final state $\ell^+\ell'^-\not{p}$, consisting of events with an electron and muon of opposite sign and missing transverse momentum, in 927 pb^{-1} of CDF data. (Right) The same final state with Standard Model WW, WZ, and ZZ contributions subtracted, and with the correction factors re-fit in the absence of these contributions. SLEUTH finds the final state $\ell^+\ell'^-\not{p}$ to contain a discrepancy surpassing the discovery threshold of $\tilde{\mathcal{P}} < 0.001$ with the processes WW, WZ, and ZZ removed from the Standard Model background.

the final states $\ell^+\ell^-\not{p}b\bar{b}$ were important in verifying the top quark pair production hypothesis in the initial observation by CDF [73] and DØ [74] in 1995, most of the statistical power came from the final state $Wb\bar{b}jj$. The fully hadronic decay into $b\bar{b}\,4j$ has only convincingly been seen after integrating substantial Run II luminosity [75]. SLEUTH's first assumption that new physics will appear predominantly in one final state is thus reasonably well satisfied. Since the top quark has a mass of 170.9 ± 1.8 GeV [76], the production of two such objects leads to a signal at large $\sum p_T$ relative to the Standard Model background of W bosons produced in association with jets, satisfying SLEUTH's second and third assumptions. SLEUTH is expected to perform reasonably well on this example.

To quantitatively test SLEUTH's sensitivity to top quark pair production, this process is removed from the Standard Model prediction, and the correction factors are re-obtained from a global fit assuming ignorance of $t\bar{t}$ production. SLEUTH easily discovers $t\bar{t}$ production in 927 pb^{-1} in the final states $b\bar{b}\ell^+\ell'^-\not{p}$ and $Wb\bar{b}jj$, shown in Fig. 3-8. SLEUTH finds $\mathcal{P}_{b\bar{b}\ell^+\ell'^-\not{p}} < 1.5 \times 10^{-8}$ and $\mathcal{P}_{Wb\bar{b}jj} < 8.3 \times 10^{-7}$, far surpassing the discovery threshold of $\tilde{\mathcal{P}} \lesssim 0.001$.

98

The test is repeated as a function of assumed integrated luminosity (Fig. 3-9), and SLEUTH is found to highlight the top quark signal at an integrated luminosity of roughly 80 ± 60 pb^{-1}, where the large variation arises from statistical fluctuations in the $t\bar{t}$ signal events. Weaker constraints on the VISTA correction factors at lower integrated luminosity marginally increase the integrated luminosity required to claim a discovery.

W boson pairs. The sensitivity to Standard Model WW production is tested by removing this process from the Standard Model background prediction and allowing the VISTA correction factors to be re-fit. In 927 pb^{-1} of Tevatron Run II data, SLEUTH identifies an excess in the final state $\ell^{+}\ell'^{-}\not{p}$, consisting of an electron and muon of opposite sign and missing momentum. This excess corresponds to $\tilde{\mathcal{P}} < 2 \times 10^{-4}$, sufficient for the discovery of WW, as shown in Fig. 3-10.

Single top. Single top quarks are produced weakly, either through a t-channel process like $bu \to td \to Wb + jet$, or through a s-channel, such as $u\bar{d} \to W^{+} \to t\bar{b} \to Wb\bar{b}$. Both of these final states are merged into SLEUTH's $Wb\bar{b}$ final state, satisfying SLEUTH's first assumption. Single top production will appear as an excess of events, satisfying SLEUTH's third assumption. SLEUTH's second assumption is not well satisfied for this example, since single top production does not lie at large $\sum p_T$ relative to other Standard Model processes. SLEUTH is thus expected to be outperformed by a targeted search in this example.

Higgs boson. Assuming a Standard Model Higgs boson of mass $m_h = 115$ GeV, the dominant observable production mechanism is $p\bar{p} \to Wh$ and $p\bar{p} \to Zh$, populating the final states $Wb\bar{b}$, $\ell^{+}\ell^{-}b\bar{b}$, and $\not{p}b\bar{b}$. The signal is thus spread over three SLEUTH final states. Events in the last of these ($\not{p}b\bar{b}$) do not pass the VISTA event selection, which does not use $\not{p}$ as a trigger object. SLEUTH's first assumption is thus poorly satisfied for this example. The Standard Model Higgs boson signal will appear as an excess, but as in the case of single top production it does not appear at particularly large $\sum p_T$ relative to other Standard Model processes. Since the Standard Model

Higgs boson poorly satisfies SLEUTH's first and second assumptions, a targeted search for this specific signal is expected to outperform SLEUTH.

Specific models of new physics

To build intuition for SLEUTH's sensitivity to new physics signals, several sensitivity tests are conducted for a variety of new physics possibilities. Some of the new physics models chosen have already been considered by more specialized analyses within CDF, making possible a comparison between SLEUTH's sensitivity and the sensitivity of these previous analyses.

SLEUTH's sensitivity can be compared to that of a dedicated search by determining the minimum new physics cross section σ_{min} required for a discovery by each. The discovery for SLEUTH occurs when $\tilde{\mathcal{P}} < 0.001$. In most SLEUTH regions satisfying the discovery threshold of $\tilde{\mathcal{P}} < 0.001$, the probability for the predicted number of events to fluctuate up to or above the number of events observed corresponds to greater than 5σ. The discovery for the dedicated search occurs when the observed excess of data corresponds to a 5σ effect. Smaller σ_{min} corresponds to greater sensitivity.

The sensitivity tests are performed by first generating pseudo data from the Standard Model background prediction. Signal events for the new physics model are generated, passed through the chain of CDF detector simulation and event reconstruction, and consecutively added to the pseudo data until SLEUTH finds $\tilde{\mathcal{P}} < 0.001$. The number of signal events needed to trigger discovery is used to calculate σ_{min}.

For each dedicated analysis to which SLEUTH is compared, the number of Standard Model events expected in 927 pb^{-1} within the region targeted is used to calculate the number of signal events required in that region to produce a discrepancy corresponding to 5σ. Using the signal efficiency determined in the dedicated analysis, σ_{min} is calculated. The effect of systematic uncertainties is not included in SLEUTH, so it is also removed from the dedicated analyses.

The results of five such sensitivity tests are summarized in Table 3.3. SLEUTH is seen to perform comparably to targeted analyses on models satisfying the assumptions on which SLEUTH is based. For models in which SLEUTH's simple use of $\sum p_T$ can be
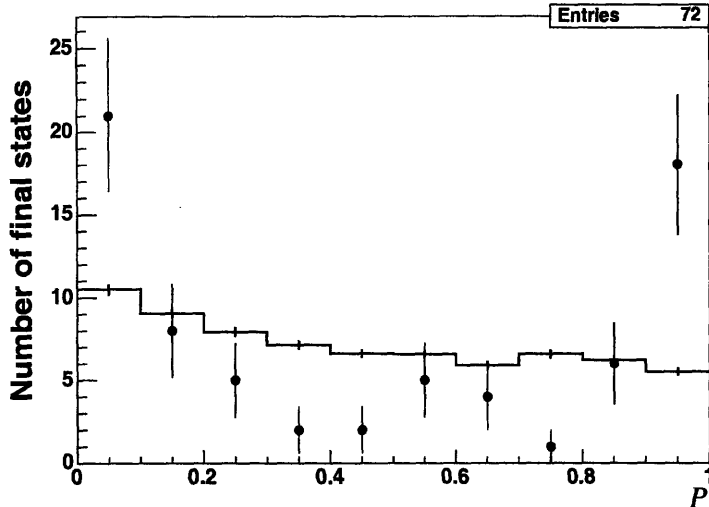
Figure 3-11: *Blue points:* The $\mathcal{P}$ distribution observed in 927 pb^{-1}, with one entry for each of the 72 SLEUTH final states with at least 3 data. There are 131 SLEUTH final states with non-zero background and less than 3 data, which are assigned $\mathcal{P} = 1$. *Black histogram:* The expected $\mathcal{P}$ distribution from all 203 SLEUTH final states with non-zero background, if instead of actual data we use pseudo-data pulled from the expected $\sum p_T$ distribution of each final state, and omit the final states where pseudo-data are less than 3 and therefore have $\mathcal{P} = 1$. As explained in Sec. 3.3.1, footnote 5, the $\mathcal{P}$ of final states with expected population $\lesssim 10$ is not uniformly distributed. Of the 203 final states SLEUTH considers in 927 pb^{-1}, 150 have Standard Model background of less than 10 events, which causes the expected $\mathcal{P}$ distribution to slightly favor smaller values.

improved upon by optimizing for a specific feature, a targeted search may be expected to achieve greater sensitivity. One of the important features of SLEUTH is that it not only performs reasonably well, but that it does so broadly. In Model 1, a search for a particular model point in a gauge mediated supersymmetry breaking (GMSB) scenario, SLEUTH gains an advantage by exploiting a final state not considered in the targeted analysis [77]. In Model 2, a search for a Z' decaying to lepton pairs, the targeted analysis [78] exploits the narrow resonance in the e^+e^- invariant mass. In Models 3 and 4, which are searches for a hadronically decaying Z' of different masses, there is no targeted analysis against which to compare. In Model 5, a search for a $Z' \to t\bar{t}$ resonance, the signal appears at large summed scalar transverse momentum in a particular final state, resulting in comparable sensitivity between SLEUTH and the targeted analysis [79].

101
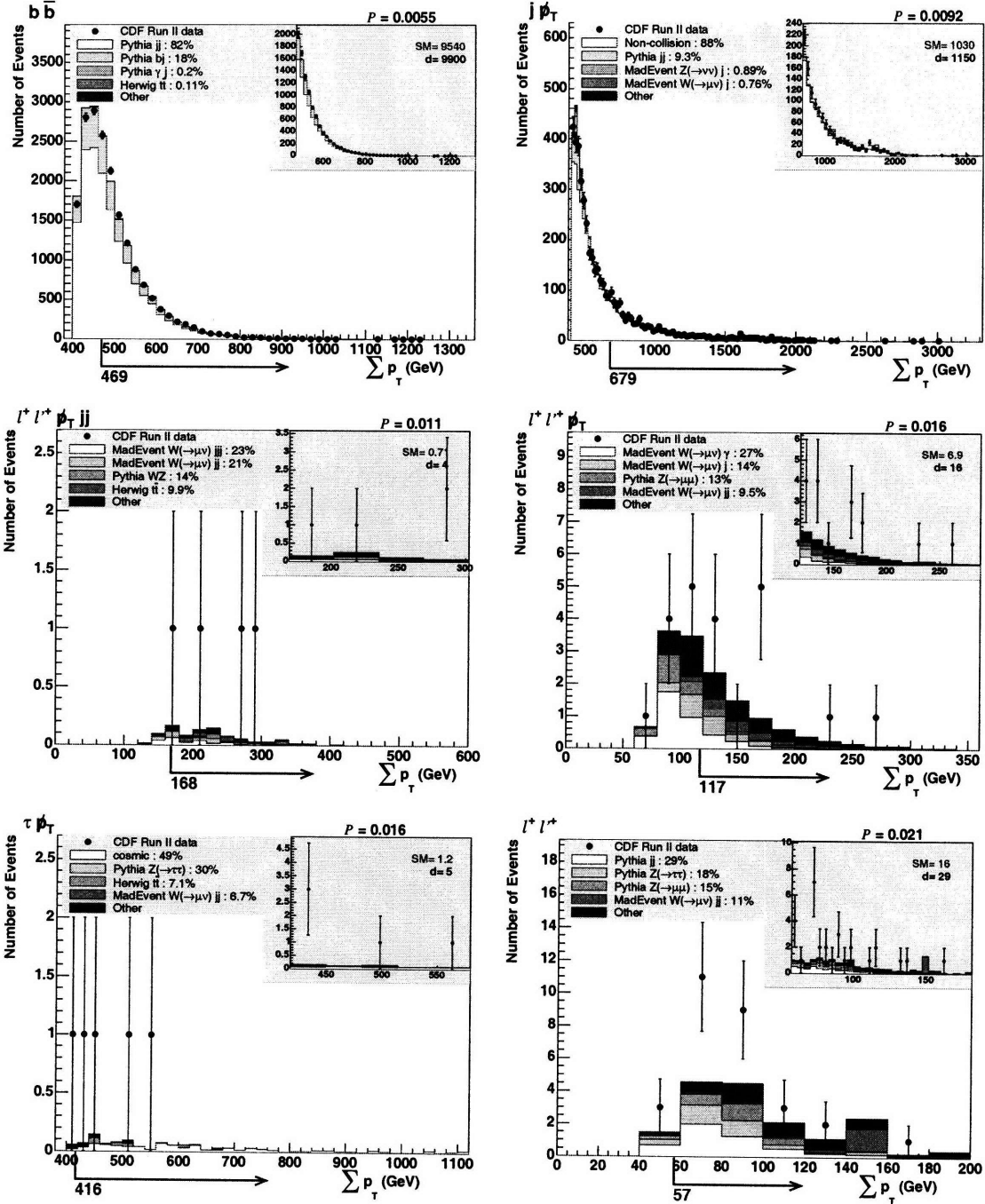
Figure 3-12: The most interesting final states identified by SLEUTH. The region chosen by SLEUTH, extending up to infinity, is shown by the (blue) arrow just below the horizontal axis. Data are shown as filled circles, and the Standard Model prediction is shown as the shaded histogram. The SLEUTH final state is labeled in the upper left corner of each panel, with ℓ denoting e or μ, and $\ell^+\ell'^+$ denoting an electron and muon with the same electric charge. The number at upper right in each panel shows $\mathcal{P}$, defined in Sec. 3.3.1. The inset in each panel shows an enlargement of the region selected by SLEUTH, together with the number of events (SM) predicted by the Standard Model in this region, and the number of data events (d) observed in this region.

102

3.3.3 Results

The distribution of $\mathcal{P}$ for the final states considered by SLEUTH in the data is shown in Fig. 3-11. The concavity of this distribution reflects the degree to which the correction model described in Sec. 3.2.5 has been tuned. A crude correction model tends to produce a distribution that is concave upwards, as seen in this figure, while an overly tuned correction model produces a distribution that is concave downwards, with more final states than expected having $\mathcal{P}$ near the midpoint of the unit interval.

The most interesting final states identified by SLEUTH are shown in Fig. 3-12, together with a quantitative measure $(\mathcal{P})$ of the interest of the most interesting region in each final state, determined as described in Sec. 3.3.1. The legends of Fig. 3-12 show the primary contributing Standard Model processes in each of these final states, together with the fractional contribution of each. The top six final states, which correspond to entries in the leftmost bin in Fig. 3-11. span a range of populations, relevant physics objects, and important background contributions. This picture is suggestive of statistical fluctuations, spread among unrelated final states.

The final state $b\bar{b}$, consisting of two or three reconstructed jets, one or two of which are b-tagged, heads the list. These events enter the analysis by satisfying the VISTA offline selection requiring one or more jets or b-jets with $p_T > 200$ GeV. The definition of SLEUTH's $\sum p_T$ variable is such that all events in this final state consequently have $\sum p_T > 400$ GeV. SLEUTH chooses the region $\sum p_T > 469$ GeV, which includes nearly 10^4 data events. The Standard Model prediction in this region is sensitive to the b-tagging efficiency $p(b \to b)$ and the fake rate $p(j \to b)$, which have few strong constraints on their values for jets with $p_T > 200$ GeV other than those imposed by other VISTA kinematic distributions within this and a few other related final states. For this region SLEUTH finds $\mathcal{P}_{b\bar{b}} = 0.0055$, which is unfortunately not statistically significant after accounting for the trials factor associated with looking in many different final states, as discussed below.

The final state $j\not{p}$, consisting of events with one reconstructed jet and significant missing transverse momentum, is the second final state identified by SLEUTH.

The primary background is due to non-collision processes, including cosmic rays and beam halo backgrounds, whose estimation is discussed in Appendix A.2.1. Since the hadronic energy is not required to be deposited in time with the beam crossing, SLEUTH's analysis of this final state is sensitive to particles with a lifetime between 1 ns and 1 μs that lodge temporarily in the hadronic calorimeter, complementing Ref. [80].

The final states $\ell^+\ell'^+\not{p}jj$, $\ell^+\ell'^+\not{p}$, and $\ell^+\ell'^+$ all contain an electron (ℓ) and muon (ℓ') with identical reconstructed charge (either both positive or both negative). The final states with and without missing transverse momentum are qualitatively different in terms of the Standard Model processes contributing to the background estimate, with the final state $\ell^+\ell'^-$ composed mostly of dijets where one jet is misreconstructed as an electron and a second jet is misreconstructed as a muon; $Z \to \tau^+\tau^-$, where one tau decays to a muon and the other to a leading π^0, one of the two photons from which converts while traveling through the silicon support structure to result in an electron reconstructed with the same sign as the muon, as described in Appendix A.1; and $Z \to \mu^+\mu^-$, in which a photon is produced, converts, and is misreconstructed as an electron. The final states containing missing transverse momentum are dominated by the production of $W(\to \mu\nu)$ in association with one or more jets, with one of the jets misreconstructed as an electron. The muon is significantly more likely than the electron to have been produced in the hard interaction, since the fake rate $p(j \to \mu)$ is roughly an order of magnitude smaller than the fake rate $p(j \to e)$, as observed in Table 4.2. The final state $\ell^+\ell'^-\not{p}jj$, which contains two or three reconstructed jets in addition to the electron, muon, and missing transverse momentum, also has some contribution from WZ and top quark pair production.

The final state $\tau\not{p}$ contains one reconstructed tau, significant missing transverse momentum, and one reconstructed jet with $p_T > 200$ GeV. This final state in principle also contains events with one reconstructed tau, significant missing transverse momentum, and zero reconstructed jets, but such events do not satisfy the offline selection criteria described in Sec. 3.2.2. Roughly half of the background is non-collision, in which two different cosmic ray muons (presumably from the same cosmic

ray shower) leave two distinct energy deposits in the CDF hadronic calorimeter, one with $p_T > 200$ GeV, and one with a single associated track from a $p\bar{p}$ collision occurring during the same bunch crossing. Less than a single event is predicted from this non-collision source (using techniques described in Appendix A.2.1) over the past five years of Tevatron running.

In these CDF data, SLEUTH finds $\tilde{\mathcal{P}} = 0.46$. The fraction of hypothetical similar CDF experiments (assuming a fixed Standard Model prediction, detector simulation, and correction model) that would exhibit a final state with $\mathcal{P}$ smaller than the smallest $\mathcal{P}$ observed in the CDF Run II data is approximately 46%. The actual value obtained for $\tilde{\mathcal{P}}$ is not of particular interest, except to note that this value is significantly greater than the threshold of $\lesssim 0.001$ required to claim an effect of statistical significance. SLEUTH has not revealed a discrepancy of sufficient statistical significance to justify a new physics claim.[9]

Systematics are incorporated into SLEUTH in the form of the flexibility in the VISTA correction model, as described previously. This flexibility is significantly more important in practice than the uncertainties on particular correction factor values obtained from the fit. The inclusion of additional systematic uncertainties would not qualitatively change the conclusion that SLEUTH has not revealed a discrepancy of sufficient statistical significance to justify a new physics claim.

Starting from the current result of SLEUTH in 927 pb^{-1}, a projection (Fig. 3-13) shows that, if the dataset roughly doubles and nothing changes in the Standard Model implementation, then $\tilde{\mathcal{P}}$ will likely be smaller than discovery threshold. This implies that, either we are on the verge of a discovery that will happen with more data, or a doubling of data will likely enforce some more accurate modeling of Standard Model backgrounds, which will possibly increase $\tilde{\mathcal{P}}$ away from its predicted small value. This clue was the main motivation to repeat and improve this search with more data, as will be described in a later chapter.

---

[9] The alternative statistic, $\tilde{p}$-val, was found to be 22%. The region with the smallest p-val is in the final state $b\bar{b}$, which also has the smallest $\mathcal{P}$. Therefore, the most interesting region pointed by both statistics is the same: $\sum p_T \geq 469$ in $b\bar{b}$.
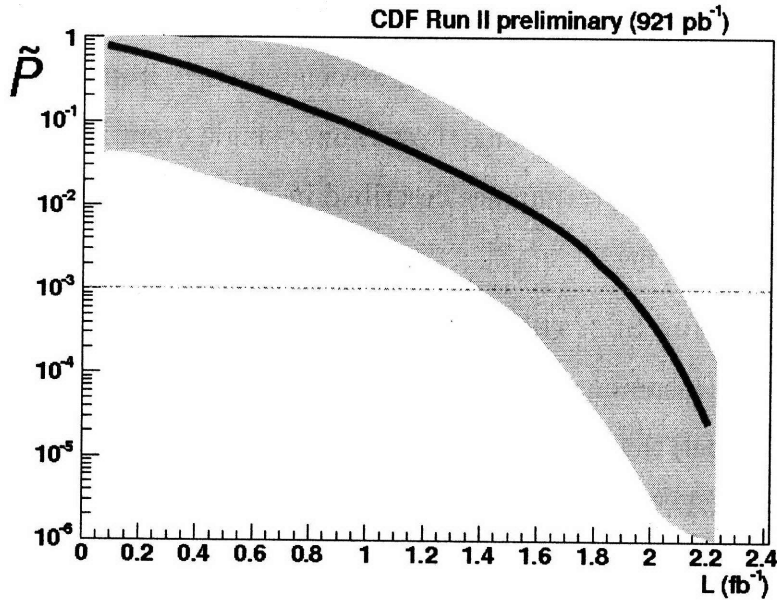
Figure 3-13: Projection of $\tilde{\mathcal{P}}$ towards lower and higher luminosities, starting from 927 pb^{-1}. Values were obtained by scaling down or up both data and backgrounds. The yellow band reflects uncertainty due to randomness in which of the present data events would have appeared in less data, or would recur in more. The Standard Model implementation is assumed invariant in all except total populations.

3.4   Summary of first round with 1 fb^{-1}

In the first round of this analysis, with 927 pb^{-1}, a complete Standard Model background estimate has been obtained and compared with data in 344 populated exclusive final states and 16,486 relevant kinematic distributions. Consideration of exclusive final state populations yields no statistically significant ($> 3\sigma$) discrepancy after the trials factor is accounted for. Quantifying the difference in shape of kinematic distributions using the Kolmogorov-Smirnov statistic, significant discrepancies are observed between data and Standard Model prediction. These discrepancies are believed to arise from mismodeling of the parton shower and intrinsic k_T, and represent observables for which a QCD-based understanding is highly motivated. None of the shape discrepancies highlighted motivates a new physics claim.

A further systematic search (SLEUTH) for regions of excess on the high-$\sum p_T$ tails of exclusive final states has been performed, representing a quasi-model-independent search for new electroweak scale physics. A measure of interest rigorously accounting

106

for the trials factor associated with looking in many regions with few events is defined, and used to quantify the most interesting region observed in the CDF Run II data. No region of excess on the high-$\sum p_T$ tail of any of the SLEUTH exclusive final states surpasses the discovery threshold.

Although this result of course can not prove that no new physics is hiding in the studied data, this search is the most encompassing test of the Standard Model at the energy frontier.

Chapter 4

Update with 2 fb^{-1}

This analysis was conducted in two rounds: first with 1 fb^{-1} of data, and then with 2 fb^{-1}. The first round was presented in Chapter 3. This chapter summarizes the second round.

4.1 Overview

Four separate statistics are employed to search for evidence of new physics. These statistics are

- a difference between the number of observed and predicted events in individual exclusive final states;

- a difference in distribution shape between data and Standard Model prediction in a variety of kinematic variables;

- an excess of data in the large $\sum p_T$ tail of exclusive final states; and

- a local excess (bump) in some invariant mass distribution, reflecting possibly a new resonance.

The next sections discus these statistics: Sec. 4.2 is about the normalization and shape statistics, Sec. 4.3 about the $\sum p_T$ statistic, and Sec. 4.4 about the mass bump statistic. Conclusions are provided in Sec. 4.5.

4.2 VISTA

Conceptually, VISTA in the second round of analysis is the same as in the first.

4.2.1 Object identification

The particle identification criteria used in this analysis are the same as in the first round, except for the following changes:

- Changed previously suboptimal conversion filter to the standard one. In the previous version, we required each lepton candidate to not have within $\Delta R < 0.4$ another track of opposite sign. The neighbor track was counted only if it had $p_T > 2$ GeV. In this version, we make no transverse momentum requirement on the candidate neighbor tracks. This change reduces significantly the rate for jets and photons to fake electrons, since both fakings involve conversions.

- For plug electrons we now require the presence of a good quality PES cluster[1], and that the PHX track matches to the electromagnetic cluster to within $\Delta R < 0.01$. This reduces the rate of jets faking electrons in the region $|\eta_{\mathrm{det}}| > 1$.

- For CMUP muons, we require CMU the distance between a stub and the track extrapolation (ΔX) to be less than 7 cm, instead of 3 cm. This follows a change in the standard muon identification criteria used by the experiment.

- For taus, the momentum is now taken from the calorimeter E_T rather than visible momentum (track momentum plus π^0s). The minimum seed track p_T requirement has been increased to 10.5 GeV, reflecting a change in online trigger criteria. We also added an additional muon veto cut requiring that the calorimter E_T over seed track p_T be greater than 0.5, inconsistent with a minimum ionizing particle.

- For plug photons, we apply the fiducial cut $|\eta_{\mathrm{det}}| > 1.2$.

Tables with identification criteria for all objects can be found in Appendix B.2.

---

[1]Variables $PES\ 5x9\ U$ and $PES\ 5x9\ V$ need to be defined and less than 0.65.

4.2.2 Event selection

The following criteria are used to keep events of interest. Single-object criteria accept events containing:

- a central electron with $p_T > 25$ GeV, or

- a plug electron with $p_T > 40$ GeV, or

- a central muon with $p_T > 25$ GeV, or

- a central photon with $p_T > 60$ GeV, or

- a central or plug photon with $p_T > 300$ GeV, or

- a central jet or b-tagged jet with $p_T > 200$ GeV, or

- a central b-tagged jet with $p_T > 60$ GeV (prescaled by the online jet20 trigger), or

- a central jet or b-tagged jet with $p_T > 40$ GeV (prescaled by 10 in addition to the online jet20 trigger prescale).

Di-object criteria keep events containing:

- one electron plus one electron or photon with $|\eta| < 2.5$ and $p_T > 25$ GeV, or

- a central or plug electron with $p_T > 40$ GeV and a central tau with $p_T > 17$ GeV, or

- a central muon with $p_T > 17$ GeV and a central or plug photon with $p_T > 25$ GeV, or

- a central muon with $p_T > 25$ GeV and a central b-tagged jet with $p_T > 17$ GeV, or

- two taus with $|\eta| < 1.0$ and $p_T > 25$ GeV, or

- a central or plug photon with $p_T > 40$ GeV and a central tau with $p_T > 40$ GeV, or

- one central photon with $p_T > 25$ GeV and one other central or plug photon with $p_T > 25$ GeV, or

- a central photon with $p_T > 40$ GeV and a central b-tagged jet with $p_T > 25$ GeV, or

- a central jet or b-tagged jet with $p_T > 40$ GeV and a central tau with $p_T > 17$ GeV (prescaled by the online jet20 trigger), or

- a central jet with $p_T > 60$ GeV and a central b-tagged jet with $p_T > 25$ GeV (prescaled by the online jet20 trigger), or

- two central muons with $p_T > 17$ GeV, or

- one central electron and one central muon with $p_T > 17$ GeV, or

- one central electron with $p_T > 20$ GeV and one central tau with $p_T > 17$ GeV, or

- one plug electron with $p_T > 25$ GeV and one central muon with $p_T > 17$ GeV, or

- one central muon with $p_T > 20$ GeV and one central tau with $p_T > 17$ GeV.

Tri-object criteria keep events containing:

- a central or plug photon with $p_T > 40$ GeV and two central taus with $p_T > 17$ GeV, or

- a central or plug photon with $p_T > 40$ GeV and two central b-tagged jets with $p_T > 25$ GeV, or

- a central or plug photon with $p_T > 40$ GeV, a central tau with $p_T > 25$ GeV, and a central b-tagged jet with $p_T > 25$ GeV, or

- one b-tagged jet with $p_T > 90$ GeV and two more b-tagged jets with $p_T > 60$ GeV, or

- one central muon with $p_T > 17$ GeV and two other central or plug muons with $p_T > 17$ GeV.

Additional special criteria accept events containing:

- one central or plug electron with $p_T > 40$ GeV, missing transverse momentum greater than 17 GeV, and two or more jets or central b-tagged jets with $p_T > 17$ GeV, or

- one central muon with $p_T > 25$ GeV, missing transverse momentum greater than 17 GeV, and two or more jets or central b-tagged jets with $p_T > 17$ GeV.

The above criteria are set by the requirements that the corresponding Standard Model prediction can be generated with enough Monte Carlo event to have weights $\lesssim 1$, and that trigger efficiencies can be treated as roughly independent of object p_T, while keeping as many potentially interesting events as possible.

Explicit online trigger paths are no longer required. CDF specific details are provided in Sec. B.1.

4.2.3   Event generation

Here are summarized changes made to our Monte Carlo event generation since the first round of analysis.

- A number of electroweak samples changed to use the newest (Gen6) CDFSIM version. They include (the Stntuple sample names are given in parentheses): Pythia $W \to e\nu$ (we0sfe, we0sge, we0she), Pythia $W \to \mu\nu$ (we0s8m, we0s9m), Pythia $W \to \tau\nu$ (we0s9t, we0sat), Pythia $Z \to ee$ (ze1s6d, ze1sad, ze0scd, ze0sdd, ze0sed, ze0see), Pythia $Z \to \mu\mu$ (ze1s9m, ze0sbm, ze0scm, ze0sdm, ze0sem), Pythia $Z \to \tau\tau$ (ze0s8t, ze0sat), Pythia WW (we0sbd, we0sgd), Pythia WZ (we0scd), Baur $W(\to e\nu) + \gamma$ (re0s28, re0s48), Baur $W(\to \mu\nu) + \gamma$ (re0s29, re0s49), Baur $W(\to \tau\nu) + \gamma$ (re0s1a, re0s4a).

- A low mass Drell-Yan sample was added with M_Z going down to 10 GeV (zx0sde, zx0sdm)

113

- We switched from using the Mrenna matched W+jets sample to the standard Top Group Alpgen W+jets samples: $W(\to e\nu)$+jets (ptopw0, ptopw1, ptop2w, ptop3w, ptop4w), $W(\to \mu\nu)$+jets (ptopw5, ptopw6, ptop7w,. ptop8w, ptop9w), $W(\to \tau\nu)$+jets (utopw0, utopw1, utop2w, utop3w, utop4w).

- We switched from using MadEvent W+bbar to the standard Top Group W+bbar sample: $W(\to e\nu)$+bb+jets (btop0w, btop1w, btop2w), $W(\to \mu\nu)$+bb+jets (btop5w, btop6w, btop7w).

Table 4.1 summarizes the contributions from each Monte Carlo sample. Specific modifications to the correction model implemented since the first round are described here.

- The integrated luminosity of the data sample considered has increased from 927 to 1990 pb^{-1}. The integrated luminosity correction factor has been adjusted accordingly.

- Events from more recent data have been included in the high-p_T jet and photon non-collision backgrounds. For events with $\sum p_T > 400$ GeV and at least two jets of $p_T > 10$ GeV and no objects of other kinds, we require the p_T of the jet with the second largest p_T to be greater than 75 GeV. This cut is to clean multijet samples of events where the second jet comes from the underlying event but the first jet is due to a cosmic ray. Such events are not modeled well by our cosmic background, which comprises events required to have less than three tracks; this requirement reduces the fraction of such cosmic + jet(s) events relative to the data sample, where more than three tracks are required. As a result of these changes, the `cosmic_ph` and `cosmic_j` correction factors have been readjusted.

- It was recognized that in the previous version of the analysis we had been using a suboptimal filter for conversion electrons. This filter has been updated and now yields a substantially reduced rate for jets faking electrons via fragmentation to a leading π^0.

Dataset	Process	Weights	* Number	= Total weight	Dataset	Process	Weights	* Number	= Total weight
pyth_jj_000	Pythia jj 0<pT<10	1700	1	1720.13	alpgen_muvmj	Alpgen W(-> mu v) j	0.3	281072	83565.2
pyth_jj_010	Pythia jj 10<pT<18	330	74	24368.4	utOs2w	Alpgen W(-> tau v)+jets	0.29	5230	1537.91
pyth_pj_008	Pythia j gamma 8<pT<12	86	5	430.59	mad_vtvt-a	MadEvent Z(->vv) gamma	0.27	136	37.22
mrenna_mu+mu-	MadEvent Z(-> mu mu)	29	220	6478.07	mad_veve-a	MadEvent Z(->vv) gamma	0.27	135	36.91
pyth_jj_090	Pythia jj 90<pT<120	22	2064	45825.1	weOs9t	Pythia W(-> tau v)	0.26	66024	17105.3
pyth_pj_012	Pythia j gamma 12<pT<22	21	1970	41809.2	utOsw1	Alpgen W(-> tau v)+jets	0.24	27785	6634.6
pyth_jj_018	Pythia jj 18<pT<40	18	24807	444683	pyth_pp	Pythia gamma gamma	0.23	25552	5783.53
alpgen_eve	Alpgen W(->e v)	12	5807	68133.2	zels6d	Pythia Z(->ee)	0.22	484676	106397
mad_vtvt-j	MadEvent Z(->vv) j	11	3	32.25	mad_e+e-b-b	MadEvent Z(->ee) bb	0.22	1028	224
mad_veve-j	MadEvent Z(->vv) j	11	3	32.01	alpgen_evejj	Alpgen W(->e v) jj	0.21	175665	37470.9
mrenna_e+e-	MadEvent Z(->ee)	10	5965	60080.4	reOs28	Baur W(->ev) gamma	0.21	22074	4698.99
alpgen_muvm	Alpgen W(-> mu v)	9.9	4483	44217.8	alpgen_muvmjj	Alpgen W(-> mu v) jj	0.2	112546	22201.2
pyth_jj_120	Pythia jj 120<pT<150	8.2	3291	27109.7	ztopcz	Pythia ZZ	0.19	588	110.12
pyth_bj_010	Pythia bj 10<pT<18	7.7	96	739.91	stelzer_Zaj	stelzer_Zaj	0.18	1592	288.68
pyth_jj_060	Pythia jj 60<pT<90	6.7	25300	170628	mad_aajj	MadEvent jj gamma gamma	0.18	7825	1406.23
mrenna_mu+mu-j	MadEvent Z(-> mu mu) j	6.6	3209	21131	mad_mu+mu-b-b	MadEvent Z(-> mu mu) bb	0.17	624	109.01
pyth_jj_040	Pythia jj 40<pT<60	5	87760	440765	mad_e+e-jj	MadEvent Z(->ee) jj	0.17	775	134.36
mad_veve-a_f	MadEvent Z(->vv) gamma	3.4	13	44.3	reOs29	Baur W(-> mu v) gamma	0.17	19972	3454.5
utOsw0	Alpgen W(-> tau v)+jets	3.2	643	2065.45	reOs1a	Baur W(-> tau v) gamma	0.17	2823	467.2
pyth_pj_022	Pythia j gamma 22<pT<45	3	31039	93761.3	pyth_jj_300	Pythia jj 300<pT<400	0.14	103800	14883.7
pyth_jj_150	Pythia jj 150<pT<200	2.7	59183	162311	mad_aaa_f	MadEvent gamma gamma gamma	0.14	52	7.4
weOsfe	Pythia W(->e v)	2.4	381263	921806	cosmic_j_hi	Cosmic (jet100)	0.12	36674	4487.59
cosmic_j_lo	Cosmic (jet20)	2.3	122	277.24	pyth_bj_040	Pythia bj 40<pT<60	0.12	160713	18850
cosmic_ph	Cosmic (photon_25_iso)	1.9	2694	4989.53	mrenna_e+e-jjj	MadEvent Z(->ee) jjj	0.11	23995	2667.48
pyth_pj_080	Pythia j gamma 80<pT	1.5	18378	28063.9	zeOs8t	Pythia Z(-> tau tau)	0.093	15030	1400.09
mrenna_e+e-j	MadEvent Z(->ee) j	1.5	28104	40784.4	pyth_bj_200	Pythia bj 200<pT<300	0.081	254807	20679.8
pyth_pj_045	Pythia j gamma 45<pT<80	1.4	82466	117398	hewk03	MadEvent Z(->ee) gamma	0.081	70476	5709.75
mrenna_mu+mu-jj	MadEvent Z(-> mu mu) jj	1.3	4146	5503.42	mad_aaa	MadEvent gamma gamma gamma	0.08	73	5.82
mad_veve-j_f	MadEvent Z(->vv) j	1.2	6	7.23	wenubb0p	Alpgen W(->e v) bb	0.075	41459	3105.72
pyth_bj_018	Pythia bj 18<pT<40	1.2	15659	18163.1	wmnubb0p	Alpgen W(-> mu v) bb	0.075	26067	1952.18
mad_e+e-	MadEvent Z(->ee)	1	520	540.85	zx0sm	Pythia Z(-> mu mu) (m_Z<20)	0.075	45	3.37
stelzer_l+l-j	stelzer_l+l-j	0.92	668	615.62	zx0se	Pythia Z(->ee) (m_Z<20)	0.074	73	5.37
mrenna_e+e-jj	MadEvent Z(->ee) jj	0.92	11258	10307.9	wenubb1p	Alpgen W(->e v) bb j	0.072	14111	1021.01
mad_mu+mu-	MadEvent Z(-> mu mu)	0.89	82	72.79	wmnubb1p	Alpgen W(-> mu v) bb j	0.072	8426	609.28
pyth_bj_060	Pythia bj 60<pT<90	0.87	10723	9348.26	hewk04	MadEvent Z(-> mu mu) gamma	0.072	2032	145.67
mad_vtvt-a_f	MadEvent Z(->vv) gamma	0.84	38	32	overlay	Overlaid events	0.071	11118	794.91
pyth_bj_090	Pythia bj 90<pT<120	0.84	2374	1990.9	pyth_jj_400	Pythia jj 400<pT	0.068	13153	894.13
mad_vtvt-j_f	MadEvent Z(->vv) j	0.81	5	4.07	alpgen_evejjj	Alpgen W(->e v) jjj	0.068	92857	6284.23
stelzer_Waj	MadEvent W(->l v)j gamma	0.69	1637	1124.5	alpgen_muvmjjj	Alpgen W(-> mu v) jjj	0.066	55704	3693.97
pyth_bj_120	Pythia bj 120<pT<150	0.67	2848	1903.72	ttop0z	Herwig ttbar	0.065	30518	1983.29
mad_aaj	MadEvent j gamma gamma	0.52	559	289.03	utOs3w	Alpgen W(-> tau v)+jets	0.063	4458	281.47
weOs8m	Pythia W(-> mu v)	0.49	1.2908e+06	630955	zeOsat	Pythia Z(-> tau tau)	0.063	22882	1438.37
pyth_bj_150	Pythia bj 150<pT<200	0.45	28272	12593.5	wmnubb2p	Alpgen W(-> mu v) bb jj	0.054	3529	189.63
mrenna_mu+mu-jjj	MadEvent Z(-> mu mu) jjj	0.44	3435	1498.33	wenubb2p	Alpgen W(->e v) bb jj	0.054	6075	325.2
mad_e+e-j	MadEvent Z(->ee) j	0.39	735	286.71	weOscd	Pythia WZ	0.053	2890	154.28
alpgen_evej	Alpgen W(->e v) j	0.35	398558	140544	weOsbd	Pythia WW	0.048	2839	136.4
weOsat	Pythia W(-> tau v)	0.35	49543	17155.3	weOsgd	Pythia WW	0.048	2567	122.92
mad_mu+mu-j	MadEvent Z(-> mu mu) j	0.34	494	166.35	alpgen_evejjjj	Alpgen W(->e v) jjjj	0.027	41696	1123.15
mad_mu+mu-jj	MadEvent Z(-> mu mu) jj	0.32	1681	532	alpgen_muvmjjjj	Alpgen W(-> mu v) jjjj	0.024	27099	662.36
zels9m	Pythia Z(-> mu mu)	0.3	370995	110522	utOs4w	Alpgen W(-> tau v)+jets	0.022	2551	56.84
					Total:				4.36864e+06

Table 4.1: The number of events contributing from each Standard Model process, ordered according to decreasing effective weight of individual Monte Carlo events. The data set names are shown in the leftmost column, with the corresponding process shown in the second column. The typical weight of individual events from each process is shown in the third column, and the "effective" number of events from each process contributing to the background estimate is shown in the fourth column. The weight from each process is totaled in the rightmost column, and the total weight is provided at bottom. The total weight is equal to the roughly four million events included in this analysis.

Code	Category	Explanation	Value	Error	Error(%)
0001	luminosity	CDF integrated luminosity	1990	50	2.6
0002	k-factor	cosmic_ph	0.83	0.05	6.0
0003	k-factor	cosmic_j	0.192	0.006	3.1
0004	k-factor	$1\gamma1j$ photon+jet(s)	0.92	0.04	4.4
0005	k-factor	$1\gamma2j$	1.26	0.05	4.0
0006	k-factor	$1\gamma3j$	1.61	0.08	5.0
0007	k-factor	$1\gamma4j+$	1.94	0.16	8.3
0008	k-factor	$2\gamma0j$ diphoton(+jets)	1.6	0.08	5.0
0009	k-factor	$2\gamma1j$	2.99	0.17	5.7
0010	k-factor	$2\gamma2j+$	1.2	0.09	7.5
0011	k-factor	W0j W (+jets)	1.38	0.03	2.2
0012	k-factor	W1j	1.33	0.03	2.3
0013	k-factor	W2j	1.99	0.05	2.5
0014	k-factor	W3j+	2.11	0.09	4.3
0015	k-factor	Z0j Z (+jets)	1.39	0.028	2.0
0016	k-factor	Z1j	1.23	0.04	3.2
0017	k-factor	Z2j+	1.02	0.04	3.9
0018	k-factor	2j $\hat{p}_T<150$ dijet	1.003	0.027	2.7
0019	k-factor	2j $150<\hat{p}_T$	1.34	0.03	2.2
0020	k-factor	3j $\hat{p}_T<150$ multijet	0.941	0.025	2.7
0021	k-factor	3j $150<\hat{p}_T$	1.48	0.04	2.7
0022	k-factor	4j $\hat{p}_T<150$	1.06	0.03	2.8
0023	k-factor	4j $150<\hat{p}_T$	1.93	0.06	3.1
0024	k-factor	5j low	1.33	0.05	3.8
0025	k-factor	1b2j $150<\hat{p}_T$	2.22	0.11	5.0
0026	k-factor	1b3j $150<\hat{p}_T$	2.98	0.15	5.0
0027	misId	p(e→e) central	0.978	0.006	0.6
0028	misId	p(e→e) plug	0.966	0.007	0.7
0029	misId	p(μ→μ) CMUP+CMX	0.888	0.007	0.8
0030	misId	p(γ→γ) central	0.949	0.018	1.9
0031	misId	p(γ→γ) plug	0.859	0.016	1.9
0032	misId	p(b→b) central	0.978	0.021	2.1
0033	misId	p(γ→e) plug	0.06	0.003	5.0
0034	misId	p(q→e) central	7.09×10^{-5}	1.9×10^{-6}	2.7
0035	misId	p(q→e) plug	0.000766	1.2×10^{-5}	1.6
0036	misId	p(q→μ)	1.14×10^{-5}	6×10^{-7}	5.2
0037	misId	p(b→μ)	3.3×10^{-5}	1.1×10^{-5}	33.0
0038	misId	p(j→b) 25$<p_T$	0.0183	0.0002	1.1
0039	misId	p(q→τ)	0.0052	0.0001	1.9
0040	misId	p(q→γ) central	0.000266	1.4×10^{-5}	5.3
0041	misId	p(q→γ) plug	0.00048	6×10^{-5}	12.6
0042	trigger	p(e→trig) plug, $p_T>25$	0.86	0.007	0.8
0043	trigger	p(μ→trig) CMUP+CMX, $p_T>25$	0.916	0.004	0.4

Table 4.2: The correction factors of VISTA correction model. The best fit values (Value) are given in the 4th column. Correction factor errors (Error) resulting from the fit are shown in the 5th column. The fractional error (Error(%)) is listed in the 6th column. All values are dimensionless except for the first one, which represents integrated luminosity and has units of pb^{-1}. These values and uncertainties are valid within the context of this correction model.

- In order to accommodate the ditau trigger, which in recent data requires a seed track with $p_T > 10$ GeV, and recognizing our concentration on the identification of single-prong taus, the track requirement for taus has been increased to 10.5 GeV. The fake rate $p(j \to \tau)$ and its dependence on p_T have been adjusted accordingly.

- In order to address questions regarding the fake rate $p(j \to \tau)$ and its consistent simultaneous application to many final states, the measurement of tau p_T is now based on the energy deposited in the calorimeter.

- In order to address questions regarding the fake rate $p(j \to \tau)$ and its consistent simultaneous application to multijet final states with large and small $\sum p_T$, a monotonically decreasing dependence of the fake rate $p(j \to \tau)$ on the generated summed scalar transverse momentum has been imposed.

- In the implementation of the fake rates $p(j \to e)$, $p(j \to \mu)$, and $p(j \to \tau)$, jets from a parent u or $\bar{d}$ quark now only fake positively charged μ and τ leptons (and positrons rather than electrons at a ratio of 2:1), and jets from a parent $\bar{u}$ or d quark now only fake negatively charged μ and τ leptons (and electrons rather than positrons at a ratio of 2:1).

- The ditau trigger, which turned on roughly 300 pb^{-1} into Run II, has now been live for a greater fraction of the total integrated dataset. The effective ditau trigger effeciency has been adjusted accordingly.

- A fake rate $p(b \to \mu)$ has been introduced.

- The p_T dependence of the fake rate $p(j \to b)$ and $p(j \to tau)$ has been adjusted.

- The η_{det} and ϕ dependence of the fake rate $p(j \to e)$ and $p(j \to ph)$ has been adjusted to take into account more geometric features of the detector including the calorimeter cracks at η_{det} of 0 and 1.1.

- The efficiency for reconstructing a jet as a non-b-tagged jet has been reduced from 1 to 1-$p(j \to b)$.
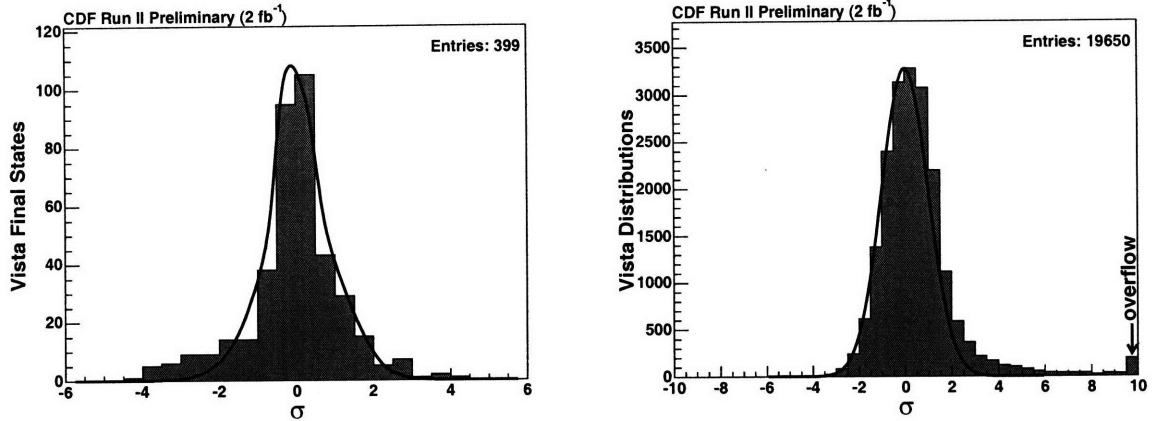
Figure 4-1: Distribution of discrepancy (before accounting trials factor) between data and Standard Model prediction, measured in units of standard deviation (σ). The left pane shows the distribution of discrepancies between the total number of events observed and predicted in the final states considered. Final states with data excess populate the right tail, while those with data deficit populate the left tail. The right pane shows the distribution of discrepancies between the observed and predicted shapes of roughly 2×10^4 kinematic distributions. Distributions in agreement correspond to small or negative σ, and distributions in disagreement correspond to large positive σ. Interest is in the entries in both tails of the distribution on the left, and in the right tail of the distribution on the right.

- Separate k-factors have been introduced for heavy flavor multijet production for the high-p_T sample. Specifically, a new k-factor has been introduced for events with at least one heavy flavor jet and three jets in total, with $\hat{p_T} > 150 \ GeV$. Another k-factor has been introduced for events with at least one heavy flavor jet and four or more jets in total, with $\hat{p_T} > 150 \ GeV$. They are listed in the table of correction factors 4.2 as 1b2j and 1b3j.

4.2.4 Results

The global fit χ^2, described in Sec. 3.2.5, was in the second round 784.43, from 335 bins, plus a 28.4 from external constraints. It is obviously a very large χ^2, even more unlikely than it was in the first round of the analysis, indicating that deviations from the fit are clearly non-statistical, but due to systematic imperfections in our Standard Model implementation. Higher statistics exacerbate systematic imperfections.

Table 4.3 shows the comparison of CDF Run II 2 fb^{-1} data to Standard Model

Final State	Data	Background	σ
$be^{\pm}\not{p}$	690	817.7 ± 9.2	−2.7
$\gamma\tau^{\pm}$	1371	1217.6 ± 13.3	+2.2
$\mu^{\pm}\tau^{\pm}$	63	35.2 ± 2.8	+1.7
$b2j\not{p}$ high-Σp_T	255	327.2 ± 8.9	−1.7
$2j\tau^{\pm}$ low-Σp_T	574	670.3 ± 8.6	−1.5
$3j\tau^{\pm}$ low-Σp_T	148	199.8 ± 5.2	−1.4
$e^{\pm}\not{p}\tau^{\pm}$	36	17.2 ± 1.7	+1.4
$2j\tau^{\pm}\tau^{\mp}$	33	62.1 ± 4.3	−1.3
$e^{\pm}j$	741710	764832 ± 6447.2	−1.3
$j2\tau^{\pm}$	105	150.8 ± 6.3	−1.2
$e^{\pm}2j$	256946	249148 ± 2201.5	+1.2
$2bj$ low-Σp_T	279	352.5 ± 11.9	−1.1
$j\tau^{\pm}$ low-Σp_T	1385	1525.8 ± 15	−1.1
$2b2j$ low-Σp_T	108	153.5 ± 6.8	−1
$b\mu^{\pm}\not{p}$	528	613.5 ± 8.7	−0.9
$\mu^{\pm}\gamma\not{p}$	523	611 ± 12.1	−0.8
$2b\gamma$	108	70.5 ± 7.9	+0.1
$8j$	14	13.1 ± 4.4	0
$7j$	103	97.8 ± 12.2	0
$6j$	653	659.7 ± 37.3	0
$5j$	3157	3178.7 ± 67.1	0
$4j$ high-Σp_T	88546	89096.6 ± 935.2	0
$4j$ low-Σp_T	14872	14809.6 ± 186.3	0
$4j2\gamma$	46	46.4 ± 3.9	0
$4j\tau^{\pm}$ high-Σp_T	29	26.6 ± 1.7	0
$4j\tau^{\pm}$ low-Σp_T	43	63.1 ± 3.3	0
$4j\not{p}$ high-Σp_T	1064	1012 ± 62.9	0
$4j\gamma\tau^{\pm}$	19	10.8 ± 2	0
$4j\gamma\not{p}$	62	104.2 ± 22.4	0
$4j\gamma$	7962	8271.2 ± 245.1	0
$4j\mu^{\pm}\not{p}$	574	590.5 ± 13.6	0
$4j\mu^{\pm}\mu^{\mp}$	38	48.4 ± 6.2	0
$4j\mu^{\pm}$	1363	1350.1 ± 37.7	0
$3j$ high-Σp_T	159926	159143 ± 1061.9	0
$3j$ low-Σp_T	62681	64213.1 ± 496	0
$3j2\gamma$	151	177.5 ± 7.1	0
$3j\tau^{\pm}$ high-Σp_T	68	76.9 ± 3	0
$3j\not{p}$ high-Σp_T	1706	1899.4 ± 77.6	0
$3j\not{p}$ low-Σp_T	42	36.2 ± 5.7	0
$3j\gamma\tau^{\pm}$	39	37.8 ± 3.6	0
$3j\gamma\not{p}$	204	249.8 ± 24.4	0
$3j\gamma$	24639	24899.4 ± 372.4	0
$3j\mu^{\pm}\not{p}$	2884	2971.5 ± 52.1	0
$3j\mu^{\pm}\gamma\not{p}$	10	3.6 ± 1.9	0
$3j\mu^{\pm}\gamma$	15	7.9 ± 2.9	0
$3j\mu^{\pm}\mu^{\mp}$	175	177.8 ± 16.2	0
$3j\mu^{\pm}$	5032	4989.5 ± 108.9	0
$3b2j$	23	28.9 ± 4.7	0
$3bj$	82	82.6 ± 5.7	0
$3b$	67	85.6 ± 7.7	0
$2\tau^{\pm}$	498	512.7 ± 14.2	0
$2\gamma\not{p}$	128	107.2 ± 6.9	0
2γ	5548	5562.8 ± 40.5	0
$2j$ high-Σp_T	190773	190842 ± 781.2	0
$2j$ low-Σp_T	165984	162530 ± 1581	0
$2j2\tau^{\pm}$	22	40.6 ± 3.2	0
$2j2\gamma\not{p}$	11	8 ± 2.4	0
$2j2\gamma$	580	581 ± 13.7	0
$2j\tau^{\pm}$ high-Σp_T	96	114.6 ± 3.3	0
$2j\not{p}$ high-Σp_T	87	80.9 ± 6.8	0
$2j\not{p}$ low-Σp_T	114	79.5 ± 100.8	0
$2j\not{p}\tau^{\pm}$	18	13.2 ± 2.2	0
$2j\gamma\tau^{\pm}$	142	144.6 ± 5.7	0
$2j\gamma\not{p}$	908	980.3 ± 63.7	0
$2j\gamma$	71364	73021.4 ± 595.9	0
$2j\mu^{\pm}\tau^{\mp}$	16	19.3 ± 2.2	0
$2j\mu^{\pm}\not{p}$	17927	18340.6 ± 201.9	0
$2j\mu^{\pm}\gamma\not{p}$	31	27.7 ± 7.7	0
$2j\mu^{\pm}\gamma$	57	58.2 ± 13	0
$2j\mu^{\pm}\mu^{\mp}\not{p}$	11	7.8 ± 2.7	0
$2j\mu^{\pm}\mu^{\mp}$	956	924.9 ± 61.2	0
$2j\mu^{\pm}$	22461	23111.4 ± 366.6	0
$2e^{\pm}j$	14	13.8 ± 2.3	0
$2e^{\pm}e^{\mp}$	20	17.5 ± 1.7	0
$2e^{\pm}$	32	49.2 ± 3.4	0
$2b$ high-Σp_T	666	689 ± 9.4	0
$2b$ low-Σp_T	323	313.2 ± 10.3	0
$2b3j$ low-Σp_T	53	57.4 ± 6.5	0
$2b2j$ high-Σp_T	718	803.3 ± 12.7	0
$2b2j\not{p}$ high-Σp_T	15	21.8 ± 2.8	0
$2b2j\gamma$	32	39.7 ± 6.2	0
$2b2j\mu^{\pm}\not{p}$	14	17.3 ± 1.9	0
$2b2j\mu^{\pm}$	22	21.8 ± 2	0
$2b\mu^{\pm}\not{p}$	11	14.4 ± 2.1	0
$2bj$ high-Σp_T	891	967.1 ± 13.2	0
$2bj\not{p}$ high-Σp_T	25	31.3 ± 3.1	0
$2bj\gamma$	71	54.5 ± 7.1	0
$2bj\mu^{\pm}\not{p}$	12	10.7 ± 1.9	0
$2be^{\pm}2j\not{p}$	30	27.3 ± 2.2	0
$2be^{\pm}2j$	72	66.5 ± 2.9	0
$2be^{\pm}\not{p}$	22	19.1 ± 2.2	0
$2be^{\pm}j\not{p}$	19	19.4 ± 2.2	0
$2be^{\pm}j$	63	63 ± 3.4	0
$2be^{\pm}$	96	92.1 ± 4.1	0
$\tau^{\pm}\tau^{\mp}$	856	872.5 ± 19	0
$\gamma\not{p}$	3793	3770.7 ± 127.3	0
$\mu^{\pm}\tau^{\mp}$	381	440.9 ± 7.3	0
$\mu^{\pm}\not{p}\tau^{\mp}$	60	75.7 ± 3.4	0
$\mu^{\pm}\not{p}\tau^{\pm}$	15	12 ± 2	0
$\mu^{\pm}\not{p}$	734290	734296 ± 4897.8	0
$\mu^{\pm}\gamma$	475	469.8 ± 12.5	0
$\mu^{\pm}\mu^{\mp}\not{p}$	169	198.5 ± 8.2	0
$\mu^{\pm}\mu^{\mp}\gamma$	83	60 ± 3.1	0
$\mu^{\pm}\mu^{\mp}$	25283	25178.5 ± 86.5	0
$j2\gamma\not{p}$	36	30.4 ± 4.2	0
$j2\gamma$	1822	1813.2 ± 27.4	0
$j\tau^{\pm}$ high-Σp_T	52	56.2 ± 2.5	0
$j\tau^{\pm}\tau^{\mp}$	203	252.2 ± 8.7	0
$j\not{p}$ high-Σp_T	4432	4431.7 ± 45.2	0
$j\gamma\tau^{\pm}$	526	476 ± 9.3	0
$j\gamma\not{p}$	1882	1791.9 ± 72.3	0
$j\gamma$	103319	102124 ± 570.6	0
$j\mu^{\pm}\tau^{\mp}$	71	98 ± 3.9	0
$j\mu^{\pm}\tau^{\pm}$	15	12 ± 2	0
$j\mu^{\pm}\not{p}\tau^{\mp}$	26	30.8 ± 2.6	0
$j\mu^{\pm}\not{p}$	109081	108323 ± 707.7	0
$j\mu^{\pm}\gamma\not{p}$	171	171.1 ± 31	0
$j\mu^{\pm}\gamma$	152	190 ± 39.3	0
$j\mu^{\pm}\mu^{\mp}\not{p}$	32	32.2 ± 10.9	0
$j\mu^{\pm}\mu^{\mp}\gamma$	14	11.5 ± 2.6	0
$j\mu^{\pm}\mu^{\mp}$	4852	4271.2 ± 185.4	0
$j\mu^{\pm}$	77689	76987.5 ± 930.2	0
$e^{\pm}4j\not{p}$	903	830.6 ± 13.2	0
$e^{\pm}4j\gamma$	25	29.2 ± 3.6	0
$e^{\pm}4j$	15750	16740.4 ± 390.5	0
$e^{\pm}3j\tau^{\mp}$	15	21.1 ± 2.2	0
$e^{\pm}3j\not{p}$	4054	4077.2 ± 63.6	0
$e^{\pm}3j\gamma$	108	79.3 ± 5	0
$e^{\pm}3j$	60725	60409.3 ± 723.3	0
$e^{\pm}2\gamma$	41	34.2 ± 2.6	0
$e^{\pm}2j\tau^{\pm}$	37	47.2 ± 2.2	0
$e^{\pm}2j\tau^{\mp}$	109	95.9 ± 6.8	0
$e^{\pm}2j\not{p}$	25725	25403.1 ± 209.4	0
$e^{\pm}2j\gamma\not{p}$	30	31.8 ± 4.8	0
$e^{\pm}2j\gamma$	398	342.8 ± 15.7	0
$e^{\pm}2j\mu^{\mp}\not{p}$	22	14.8 ± 1.9	0
$e^{\pm}2j\mu^{\mp}$	23	15.8 ± 2	0
$e^{\pm}\tau^{\pm}$	437	387 ± 5.3	0
$e^{\pm}\tau^{\mp}$	1333	1266 ± 12.3	0
$e^{\pm}\not{p}\tau^{\mp}$	109	106.1 ± 2.7	0
$e^{\pm}\not{p}$	960826	956579 ± 3077.7	0
$e^{\pm}\gamma\not{p}$	497	496.8 ± 10.3	0
$e^{\pm}\gamma$	3578	3589.9 ± 24.1	0
$e^{\pm}\mu^{\pm}\not{p}$	31	29.9 ± 1.6	0
$e^{\pm}\mu^{\mp}\not{p}$	109	99.4 ± 2.4	0
$e^{\pm}\mu^{\pm}$	45	28.5 ± 1.8	0
$e^{\pm}\mu^{\mp}$	350	313 ± 5.4	0
$e^{\pm}j2\gamma$	13	16.1 ± 3.9	0
$e^{\pm}j\tau^{\mp}$	386	418 ± 18.9	0
$e^{\pm}j\tau^{\pm}$	160	162.8 ± 3.5	0
$e^{\pm}j\not{p}\tau^{\mp}$	48	44.6 ± 3.3	0
$e^{\pm}j\not{p}\tau^{\pm}$	11	8.3 ± 1.5	0
$e^{\pm}j\not{p}$	121431	121023 ± 747.6	0
$e^{\pm}j\gamma\not{p}$	159	192.6 ± 10.9	0
$e^{\pm}j\gamma$	1389	1368.9 ± 38.9	0
$e^{\pm}j\mu^{\mp}\not{p}$	42	33 ± 2.9	0
$e^{\pm}j\mu^{\pm}\not{p}$	16	9.2 ± 1.9	0
$e^{\pm}j\mu^{\mp}$	62	63.8 ± 3.2	0
$e^{\pm}j\mu^{\pm}$	13	8.2 ± 2	0
$e^{\pm}e^{\mp}4j$	148	159.1 ± 7	0
$e^{\pm}e^{\mp}3j$	717	743.6 ± 24.4	0
$e^{\pm}e^{\mp}2j\not{p}$	32	41.4 ± 5.6	0
$e^{\pm}e^{\mp}2j\gamma$	10	11.4 ± 2.9	0
$e^{\pm}e^{\mp}2j$	3638	3566.8 ± 72	0
$e^{\pm}e^{\mp}\tau^{\pm}$	18	16.1 ± 1.7	0
$e^{\pm}e^{\mp}\not{p}$	822	831.8 ± 13.6	0
$e^{\pm}e^{\mp}\gamma$	191	221.9 ± 5.1	0
$e^{\pm}e^{\mp}j\not{p}$	155	170.8 ± 12.4	0
$e^{\pm}e^{\mp}j\gamma$	48	45 ± 3.9	0
$e^{\pm}e^{\mp}j$	17903	18258.2 ± 204.4	0
$e^{\pm}e^{\mp}$	98901	99086.9 ± 147.8	0
$b6j$	51	42.3 ± 3.8	0
$b5j$	237	192.5 ± 7.1	0
$b4j$ high-Σp_T	26	23.4 ± 2.6	0
$b4j$ low-Σp_T	836	821.7 ± 15.9	0
$b3j$ high-Σp_T	12081	12071 ± 84.1	0
$b3j$ low-Σp_T	2974	2873 ± 31	0

Table 4.3: A subset of the comparison between Tevatron Run II data and Standard Model prediction.

prediction. All events have been partitioned in exclusive final states. The number of events observed is compared to the number expected from the Standard Model, taking into account the uncertainty due to finite Monte Carlo statistics, and the trials factor due to examining 399 final states. The final states are ordered in decreasing discrepancy.

No final state is found to have a population discrepancy that is considered significant after accounting for the trials factor. The largest population discrepancy is a 2.7σ deficit (including trials factor) observed in final state $be^{\pm}\not{p}$. Fig. 4-1 summarizes in a histogram the distribution of discrepancies observed in final state populations. Qualitatively, shape discrepancies give us the same information we had in the first round of the analysis.

Discrepant distributions are flagged using the Kolmogorov-Smirnov (KS) statistic. [2] Fig. 4-1 shows a histogram of the disagreement seen in all kinematic distributions. 19,650 distributions are considered in 2 fb^{-1}, and 559 are found to have a significant disagreement. However, as in the first round with 1.0 fb^{-1}, no indication of new physics is found amongst these discrepant distributions; all are attributed to the "3-jet effect", difficulties with intrinsic k_T or residual coarseness of the correction model.

Evolution of the VISTA Global Comparison since 1 fb^{-1}

Table 4.4 displays the VISTA final states which newly appeared in the present analysis. A large number involve b-jets; this is a result of changes in our offline event selection criteria, which now accept more events containing b-tagged jets (previously events with a leading b-jet with $p_T < 200$ were prescaled offline by a factor of 10; we also introduced a new tri-b offline selection).

---

[2] The KS statistic is defined in terms of the cumulative distributions of two populations. Given a particular distribution, such as the invariant mass mass(j1,j2) of the two jets in the 1e+2j1pmiss final state, the Standard Model prediction and the data are both normalized to unit integral, and the cumulative distributions are drawn. The maximal separation of the two cumulative distributions is the KS statistic, a number between 0 and 1. This statistic can be translated into a probability for the data to have been pulled from the Standard Model distribution, with the translation depending only on the value of the statistic and the number of data events. This KS probability KS$_p$ can then be converted into units of standard deviations KS$_\sigma$ by solving $\int_{-\infty}^{KS_\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = KS_p$.

Final State	Data	Background	σ
4j$\tau^{\pm}\tau^{\mp}$	1	2.4 ± 1.5	0
4j$\mu^{\pm}\tau^{\mp}$	1	0.7 ± 1.1	0
4j$\mu^{\pm}\mu^{\mp}p$	1	1.1 ± 1.5	0
4b4j	1	0 ± 1	0
4b2j	1	0.9 ± 1.3	0
4b	3	1.3 ± 1.3	0
3j2$\tau^{\pm}$	6	8.1 ± 1.8	0
3j2γp	1	2 ± 1.6	0
3j$\mu^{\pm}\tau^{\pm}$	3	0.8 ± 1.2	0
3b4j	2	2.9 ± 1.5	0
3b3j	8	8.2 ± 2	0
3b2jp	1	0.7 ± 1.2	0
3b2j	23	27.2 ± 4.8	0
3bγ	1	0.4 ± 1.2	0
3bj$p$	1	3 ± 1.6	0
3bjγ	1	1.8 ± 1.5	0
3bj$\mu^{\pm}p$	1	1.1 ± 1.2	0
3be$^{\pm}p$	1	0.4 ± 1.2	0
2$\mu^{\pm}p$	3	0.8 ± 1.1	0
2j$\gamma p\tau^{\pm}$	1	1.1 ± 1.3	0
2j$\mu^{\pm}p\tau^{\pm}$	1	1.7 ± 1.4	0
2b6j	2	0.3 ± 1.2	0

Final State	Data	Background	σ
2b$\mu^{\pm}\mu^{\mp}$	2	1.1 ± 1.2	0
2bj$\tau^{\pm}$	1	0.8 ± 1.2	0
2bj$\mu^{\pm}\mu^{\mp}$	3	0.3 ± 1.1	0
2be$^{\pm}p\tau^{\mp}$	1	0.2 ± 1.1	0
2be$^{\pm}\mu^{\mp}p$	1	2.2 ± 1.3	0
$\gamma 2\tau^{\pm}$	2	0.1 ± 1.1	0
j2$\tau\tau^{\pm}$	2	1.8 ± 1.4	0
j2$\mu^{\pm}p$	1	0.6 ± 1.2	0
j$\mu^{\pm}2\gamma p$	1	0.1 ± 1.1	0
j$\mu^{\pm}\gamma\tau^{\mp}$	1	0.1 ± 1.1	0
e$^{\pm}4j\tau^{\pm}$	2	3.1 ± 1.2	0
e$^{\pm}4j\mu^{\mp}p$	1	0.6 ± 1.2	0
e$^{\pm}4j\mu^{\pm}p$	1	0 ± 1	0
e$^{\pm}4j\mu^{\mp}$	1	0.7 ± 1.2	0
e$^{\pm}3j\mu^{\mp}$	4	3 ± 1.4	0
e$^{\pm}\gamma\tau^{\pm}$	1	0.9 ± 1.1	0
e$^{\pm}\gamma p\tau^{\mp}$	1	0.5 ± 1.2	0
e$^{\pm}\mu^{\pm}\mu^{\mp}p$	1	0.6 ± 1.1	0
e$^{\pm}$j2γp	1	0.2 ± 1.1	0
e$^{\pm}$j$\mu^{\pm}\mu^{\mp}$	1	0.8 ± 1.1	0
e$^{\pm}$e$^{\mp}$2j$\mu^{\pm}p$	1	0 ± 1	0
e$^{\pm}$e$^{\mp}$j$\mu^{\pm}p$	1	0.2 ± 1	0

Final State	Data	Background	σ
b6jp	1	0.1 ± 1.1	0
b4jp 400+	3	1.6 ± 1.4	0
b3j$\mu^{\pm}\tau^{\pm}$	1	0.1 ± 1	0
b2j$\tau^{\pm}\tau^{\mp}$	1	0.1 ± 1.1	0
b2j$\mu^{\pm}\gamma$	1	0.9 ± 1.3	0
b$\tau^{\pm}\tau^{\mp}$	2	1.6 ± 1.3	0
b$\mu^{\pm}p\tau^{\mp}$	1	1.1 ± 1.3	0
b$\mu^{\pm}\gamma$	1	0.7 ± 1.2	0
b$\mu^{\pm}\mu^{\mp}p$	3	0.7 ± 1.3	0
bj$\tau^{\pm}\tau^{\mp}$	1	0.6 ± 1.2	0
bj$\mu^{\pm}\tau^{\mp}$	1	0.5 ± 1.2	0
be$^{\pm}$3jγ	1	1.4 ± 1.2	0
be$^{\pm}$3j$\mu^{\mp}p$	1	0.8 ± 1.2	0
be$^{\pm}$2γ	2	0.2 ± 1.1	0
be$^{\pm}$2j$\tau^{\mp}$	2	1.6 ± 1.2	0
be$^{\pm}$2j$p\tau^{\mp}$	2	0.9 ± 1.2	0
be$^{\pm}$2jγp	3	0.4 ± 1.2	0
be$^{\pm}\tau^{\pm}$	1	1.2 ± 1.1	0
be$^{\pm}\gamma p$	3	2.7 ± 1.5	0
be$^{\pm}$j$p\tau^{\mp}$	1	1.5 ± 1.3	0

Table 4.4: New VISTA final states which appeared in the analysis of 2 fb^{-1}.

There are also 11 final states which were populated in the 1.0 fb^{-1} analysis, but are not now: 1b1e+3j1tau- 1b3j2ph 1e+ 1e+1e-1ph1tau+ 1e+3j2ph 1j1pmiss2tau+ 1j3ph 2b2ph 3j1mu+1pmiss1tau+ 3j1pmiss1tau+1tau- 1b1e+3j1ph1pmiss These events were generally found to contain an object (usually a τ or plug photon) which now fails our tighter identification requirements.

A final reason for the increase of VISTA final states from 344 in 1.0 fb^{-1} to 399, is that jet-tau final states have been divided into high-p_T and low-p_T states.

The 3j$\tau^{\pm}$ and 2j$\tau^{\pm}$ final states remain among the 'top ten' most discrepant states, but their significance has decreased compared to the first round. The improvement in agreement was achieved after slight changes in modeling jets faking taus in events with large activity. Other final states from the first round's top ten now exhibit zero discrepancy (after accounting for the trials factor). We attribute this to a combination of general improvements in modeling and statistical fluctutations.

4.3 SLEUTH

SLEUTH algorithm was not modified in the second round.
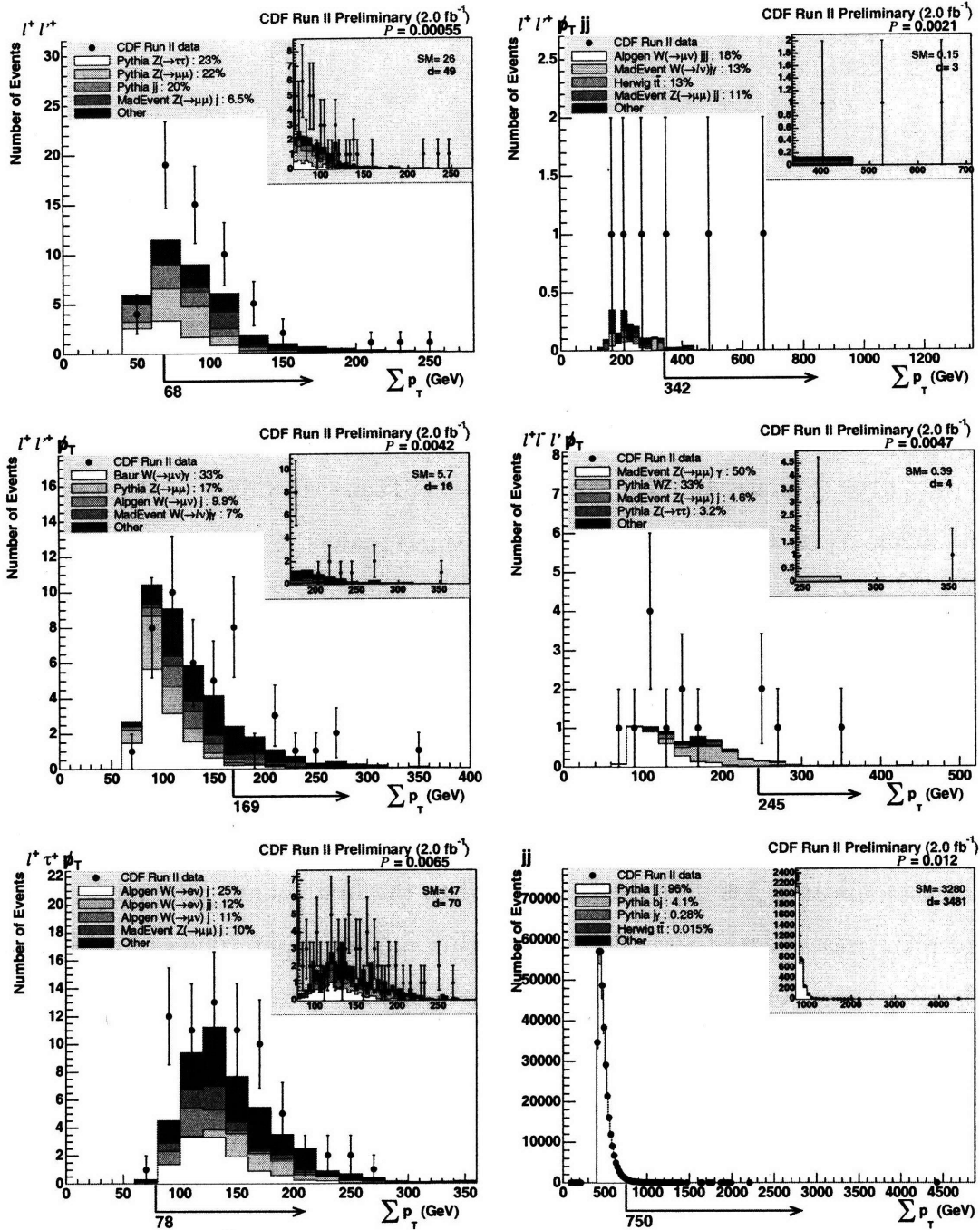
Figure 4-2: The most interesting final states identified by SLEUTH in 2 fb^{-1}.
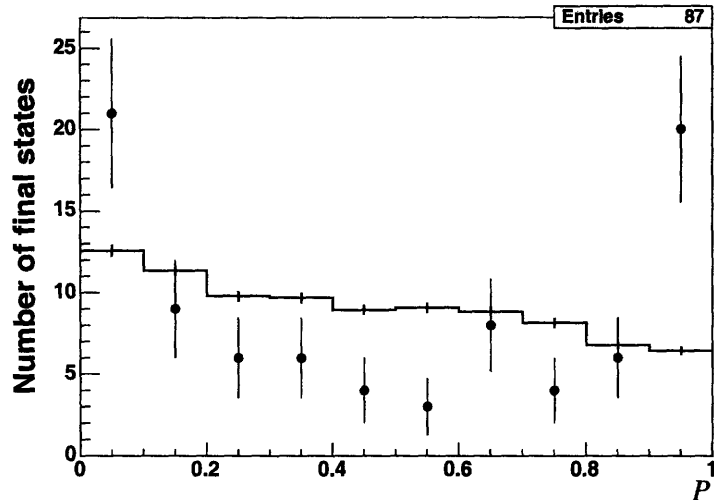
Figure 4-3: *Blue points:* The $\mathcal{P}$ distribution observed in 1990 pb^{-1}, with one entry for each of the 87 SLEUTH final states with at least 3 data. There are 153 SLEUTH final states with non-zero background and less than 3 data, which are assigned $\mathcal{P} = 1$. *Black histogram:* The expected $\mathcal{P}$ distribution from all 240 SLEUTH final states with non-zero background, if instead of actual data we use pseudo-data pulled from the expected $\sum p_T$ distribution of each final state, and omit the final states where pseudo-data are less than 3 and therefore have $\mathcal{P} = 1$. As explained in Sec. 3.3.1, footnote 5, the $\mathcal{P}$ of final states with expected population $\lesssim 10$ is not uniformly distributed. Of the 240 final states SLEUTH considers in 1990 pb^{-1}, 171 have Standard Model background of less than 10 events, which causes the expected $\mathcal{P}$ distribution to slightly favor smaller values.

4.3.1 Results

The most interesting final states highlighted by SLEUTH are shown in Fig. 4-2. The region chosen by SLEUTH is shown by the (blue) arrow, extending up to infinity. CDF Run II data are shown as filled circles; Standard Model prediction is shown as a histogram. SLEUTH final state labels are in the upper left corner of each panel. The number at upper right in each panel is $\mathcal{P}$, the fraction of hypothetical similar experiments in which something as interesting as the region shown would be seen in this final state. The inset in each panel shows an enlargement of the region selected by SLEUTH, together with the number of events (SM) predicted by the Standard Model in this region, and the number of data events (d) observed in that region.

The distribution of $\mathcal{P}$ for the final states considered by SLEUTH in the CDF Run II data is shown in Fig. 4-3.

In these CDF data, SLEUTH finds $\tilde{\mathcal{P}} = 0.085$. This is sufficiently far above the SLEUTH discovery threshold of $\tilde{\mathcal{P}} < 0.001$ that no discovery claim can be made on the basis of SLEUTH for 2 fb^{-1}.

Study of Same-Sign SLEUTH States

The top SLEUTH final states appear a common trend to involve same-sign leptons. We first consider the 2nd and 3rd SLEUTH final states, which both contain same-sign electron and muon, significant missing energy, and varying numbers of jets. The relevant VISTA final states are:

Final State	data	background
$e^+\mu^+\not{p}$	31	29.9 ± 1.6
$e^+j\mu^+\not{p}$	16	9.2 ± 1.9
$e^+2j\mu^+\not{p}$	6	1.7 ± 1.2
$e^+3j\mu^+\not{p}$	0	0.26 ± 0.07

The primary backgrounds for all these final states are similar, although the relative proportions vary with the number of reconstructed jets. The three main backgrounds are: $W(\to \mu\nu)$+jets, with a jet faking the electron; $Z(\to \mu^+\mu^-)$+jets, where 1 μ is not

reconstructed, creating missing energy, and a jet fakes the electron; and $W\gamma$(+jets), where the photon fakes the electron.

All these processes involve real muons – there is no significant Standard Model contribution to these final states from fake muons. Therefore we can discard any explanation for the excess in data which involves charge assignment to muons faked by jets.

We can be confident that the charge-sign of a real muon is well-measured by the CDF tracking system. The curvature resolution of the chamber is $\sigma_C = 3.6 \times 10^{-6}$ cm^{-1}. The curvature corresponding to a track with momentum of 100 GeV/c is 2.1×10^{-5} cm^{-1}. The sign of the curvature of such a track, and hence the charge of such a particle, is thus typically determined with a significance of better than five standard deviations [81]. VISTA supports this conclusion, since we reconstruct $\sim$25,000 $\mu^+\mu^-$ events but only a single $\mu^+\mu^+$ event (and even then, the $\mu^+\mu^+$ invariant mass is $\sim$150 GeV, making it unlikely to be a Z decay with wrong charge-reconstruction).

We can assume the muon charge is correct therefore, and focus on the electron. This is a fake electron from a jet. This fake rate is well-determined from the electron+jet(s) events, and similarly the k-factors for the boson+jets processes are well-determined from other final states. We expect the contribution from these processes to these particular final states to therefore be accurate. Indeed, the most populous state 1e+1mu+1pmiss is well described, and the mild excesses seen by SLEUTH arise from the 1e+1j1mu+1pmiss and 1e+2j1mu+1pmiss final states. Examination of the kinematic distributions from thse final states yields nothing further (the electron η_{det} distributions for these final states are shown in Figs. 4-4 and 4-5), so, following the above reasoning and given that the effect is not statistically very signficant, we ascribe the presence of these two states towards the top of SLEUTH's list as likely just due to a fluctuation.

The 1$^{\text{st}}$ SLEUTH final state 1e+1mu+ also has same sign electron and muon, but no missing energy, and 0 or 1 jets. The potentially relevant VISTA final states are:
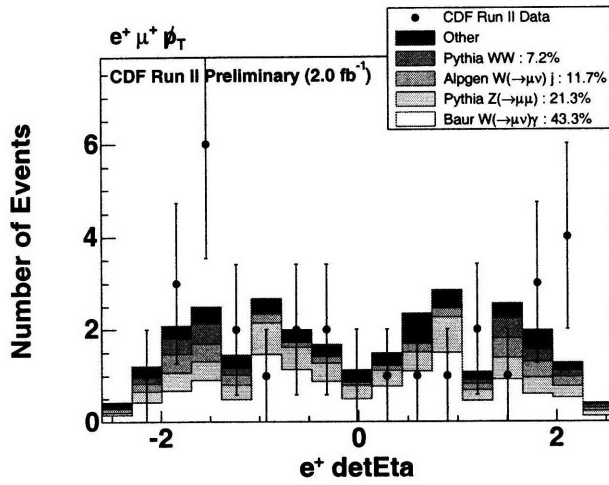
Figure 4-4: Detector η distribution for the electron in 1e+1mu+1pmiss.
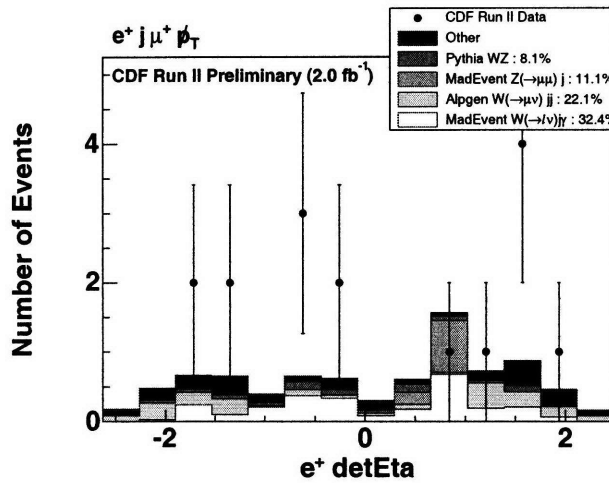
Figure 4-5: Detector η distribution for the electron in 1e+1j1mu+1pmiss.

Final State	data	background
$e^+\mu^+$	45	28.5 ± 1.8
$e^+j\mu^+$	13	8.2 ± 2
$e^+2j\mu^+$	2	2.6 ± 1.6
$e^+3j\mu^+$	2	0.6 ± 1.2

So only the data excess in $e^+\mu^+$ needs any potential investigation for evidence of Standard Model background mismodeling. The largest background is from $Z \rightarrow (\mu^+\mu^-)$+jets, with one muon lost and a jet faking an electron. As explained earlier, this process is well-constrained and cannot explain the excess in data.

The next largest background is $Z \rightarrow \tau^+\tau^-$, with one τ decaying to an electron and the other to a muon. As discussed above, we trust the muon charge, so the electron must be reconstructed with the wrong charge. For central electrons, this occurs at a rate on the order of 1 in 10^{-4}, through electron bremstrahlung to a photon with an asymmetric conversion that half the time results in an opposite charge electron, and therefore is too small to play a role here. For plug electrons, however, the track charge has a false-reconstruction rate of order 10% [82]. Fig. 4-6 shows the η_{det} of the electron, and we indeed observe that the $Z \rightarrow \tau\tau$ contribution is almost entirely in the plug. However, Fig. 4-7, which shows electron η_{det} for the 2e+ final state (dominated by real electrons from Z with phoenix track charge mis-assignment), demonstrates that this charge misidentification is quite well modeled – there is certainly no room for the factor of two increase that would be needed to explain the data excess. The only other large background is from QCD dijet events where both electron and muon are fakes. Both of these total fake rates are very well constrained from the electron+jets(s) and muon+jet(s) final states, so the only possible flexibility is in the charge assignment to the fakes, which would shift background events between the 1e+1mu+ and 1e+1mu- final states. However, with our current modeling, this process contributes an approximately equal number of expected events (~ 5) to each of these states. It is implausible to argue that the combination of QCD Feynman diagrams and faking mechanisms could be such as to significantly *anti-correlate* the fake electron and muon charge signs, so this cannot contribute to the
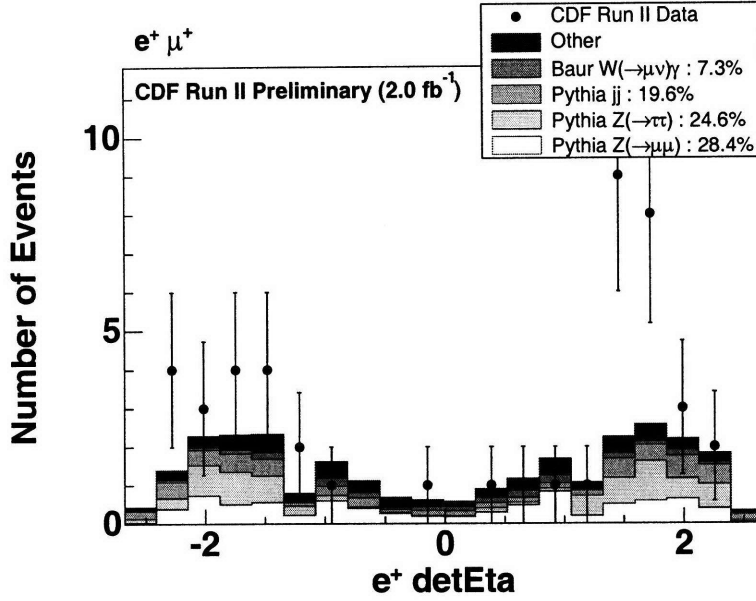
Figure 4-6: η_{det} distribution for the electron in 1e+mu+

data excess. In conclusion, after examining the possibilities and reminding ourselves that the similar final states but with additional jets are actually well described, we have no explanation for this excess other than a statistical fluctuation.

The 5$^{\text{th}}$ most discrepant state in SLEUTH is $\ell^+\tau^+$. Since SLEUTH combines electrons and muons, the relevant VISTA final states are:

Final State	data	background
$e^+\not{p}\tau^+$	36	17.2 ± 1.7
$e^+j\not{p}\tau^+$	11	8.3 ± 1.5
$\mu^+\not{p}\tau^+$	15	12 ± 2
$j\mu^+\not{p}\tau^+$	8	9.4 ± 3.1

One sees that the excess comes only from $e^+\not{p}\tau^+$. This is actually among most discrepant final states in VISTA, with a significance of 1.4σ after accounting for the trials factor. The primary background is $W \to e\nu$+jet, where the jet ends up faking a τ with the same charge as the electron. This is rarer than the other case where the fake τ has opposite sign to the electron. However, we appear to be modeling this process quite well, because it equally applies in the case when the W decays to muon and neutrino, and VISTA predicts those final states correctly. We believe the excess in $e^+\not{p}\tau^+$ is therefore likely just a fluctuation.

128
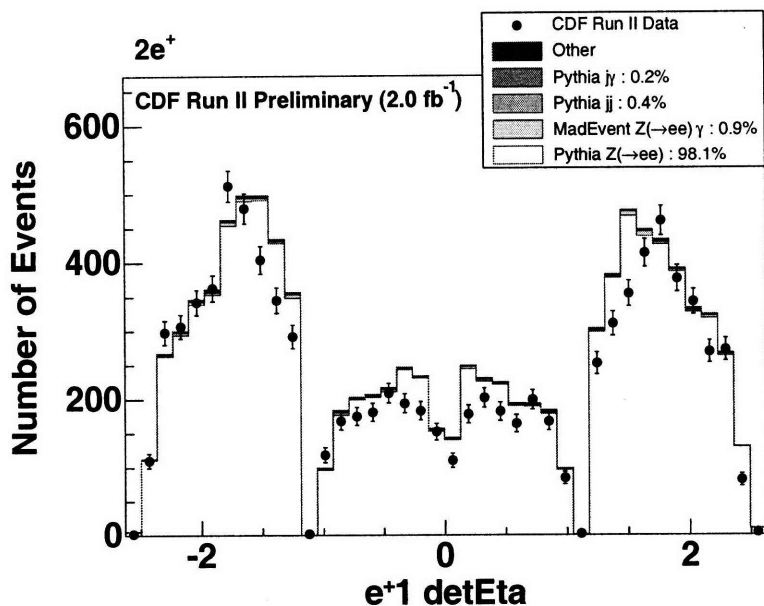
Figure 4-7: η_{det} distribution for the electron in 2e+

In conclusion, although the top SLEUTH states all involve same-sign leptons, we find no explanation that can simultaneously account for all. More data would help us see to what extent this is mismodeling, and to what statistical fluctuation.

Evolution of the Top SLEUTH Final States from 1 fb^{-1}

The 1bb final state which was at the top of the list of SLEUTH discrepancies has now gone down the list. The reason is that the region selected previously had been selected based on a relatively small excess in a particular region of $\sum p_T$. Doubling the data caused that region to exceed the upper limit of 10,000 events. This upper limit is designed to reject excesses found in regions of high statistics where even a small systematic error would cause SLEUTH to give a large discrepancy.

The discrepancy in the $j\not{p}$ final state, which is dominated by cosmic events, has been corrected by the additional quality criteria cuts on the cosmic background.

The 3rd, 4th and 6th most discrepant SLEUTH final states from the first round were same sign dilepton final states. These final states have become more discrepant in this round of the analysis as discussed in Sec. 4.3.1.

The 5th most discrepant SLEUTH final state from the first round of the analysis was the $\not{p}\tau^+$. Then, we a major background contribution was missing, $W(\to \tau\nu) + jets$,

which has been added.

The remaining discrepancies were all corrected either by improving the background modelling, or were simply fluctuations.

4.3.2 Sensitivity

For the 2 fb^{-1} analysis, we have performed an additional test of the sensitivity of SLEUTH to Standard Model single top production.

Final State	Events	Acceptance (%)
Wjj	5149	5.1
$Wb\bar{b}$	3231	3.2
W	1977	2.0
$W4j$	298	0.3
$Wb\bar{b}jj$	219	0.2
$b\bar{b}\rlap{/}p$	128	0.1
jj	109	0.1
$b\bar{b}$	96	0.1
$jj\rlap{/}p$	59	0.1
$b\bar{b}2j$	41	0.0

Table 4.5: Partitioning of events in Single Top into SLEUTH final states. The most populous final states are shown. The offline selection filter accepts % of the pseudo-signal events. The acceptance is shown for each individual final state.

cost	Final state	$\tilde{\mathcal{P}}$
3600	$Wb\bar{b}$	=0.0009669
4800	$Wb\bar{b}$	=0.0003004
3800	$Wb\bar{b}$	=0.0002808
3600	Wjj	=0.0008754
3600	$Wb\bar{b}$	=0.0002843
3800	$Wb\bar{b}$	=0.0007113
5000	$Wb\bar{b}$	=0.0007072
3800	$Wb\bar{b}$	=0.0003327
5400	$Wb\bar{b}$	=0.0003309
2800	$Wb\bar{b}$	=0.0004739

Table 4.6: Summary of "discoveries" for single top. Cost is the number of pseudo-signal events required to obtain $\tilde{\mathcal{P}} < 0.001$. The second column contains the final state in which the most interesting region is found at the point of discovery. The third column contains $\tilde{\mathcal{P}}$ at discovery.

130
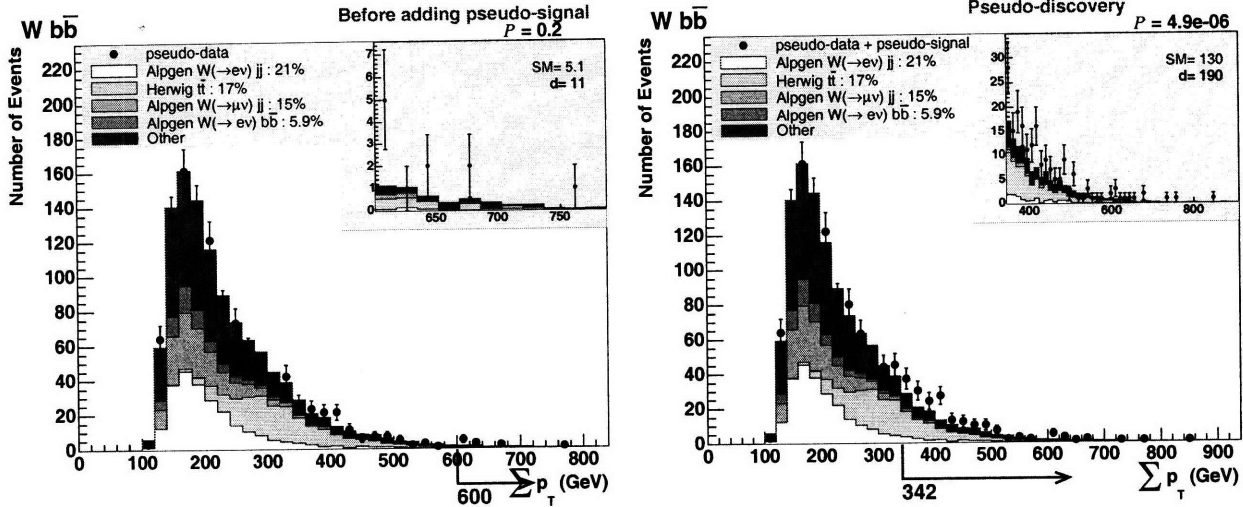
Figure 4-8: *(Left)* The final state in which single top first appears, as it is before the addition of any pseudo signal. *(Right)* The same final state, after the addition of pseudo signal required for its discovery by SLEUTH. For this discovery, 3600 pseudo signal events yields $\tilde{\mathcal{P}} = 0.0009669$.

This sensitivity test is performed by injecting 'signal' single top events into pseudo-data generated from the background. Single top events are obtained from the CDF Top Group Monte Carlo samples stop00 and stop01 (*s*-channel and *t*-channel production respectively), run through our standard event reconstruction. The acceptance for the signal events into SLEUTH final states is shown in Table 4.5.

Signal events are added to the pseudo-data in chunks, until SLEUTH's discovery threshold of $\tilde{\mathcal{P}} < 0.001$ is reached. To account for random fluctuations, ten such trials are performed and the final result is averaged from all trials. Table 4.6 summarizes the result of each trial.

As expected, SLEUTH's 'golden' final state for discovering single top is $W b\bar{b}$. The $\sim 4\%$ acceptance into this final state is consistent with the numbers obtained for dedicated single top searches [83]. Note that due to the definition of final states in SLEUTH, $W b\bar{b}$ contains events with 2 or 3 jets, with at least 1 *b*-tag. This merges somewhat the standard single top separation into distinct 2-jet and 3-jet bins, and this is why the $t\bar{t}$ background contribution is relatively large.

An example 'discovery' is illustrated in Fig. 4-8. This shows the combined background prediction in the absence of signal, and the $\sum p_T$ distribution after adding
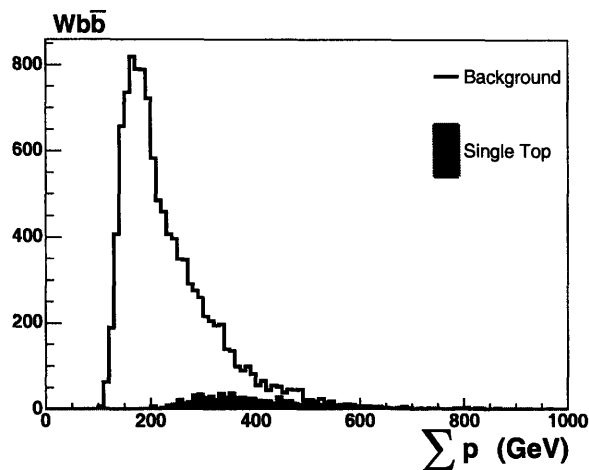
Figure 4-9: Relative $\sum p_T$ distributions from single top signal and combined background prediction.

sufficient signal to trigger SLEUTH's discovery threshold. Fig. 4-9 illustrates the $\sum p_T$ distribution from single top signal relative to the combined background prediction.

The result of this sensitivity test is that SLEUTH would be expected to discover single top at the 5σ level in 2 fb^{-1} if it had a cross-section of 5.9 ± 1.1 pb. The Standard Model expected cross-section is 2.86 pb (combined s- and t-channel). A naive extrapolation therefore leads to an expected luminosity for SLEUTH discovery of $2.0 \times (5.9/2.86)^2 = 8.5 \pm 3.1$ fb^{-1}.

This conclusion seems perhaps surprising given the effort devoted to sophisticated tools such as Matrix Elements and Neural Networks for dedicated single top searches. The apparent sensitivity of SLEUTH stems from the fact that it treats the background as being absolutely fixed. Any addition is therefore considered pure signal, allowing 'discovery' of single top with relatively few extra events. In practice this is unrealistic, since $\sum p_T$ alone would find it hard to distinguish between single top production and excess W+heavy flavour relative to ALPGEN predictions, which have a large uncertainty. In a realistic test, we would probably have to introduce a separate k-factor for W+heavy flavour, which would swallow up much of the single top signal, since there is no other populous final state that could constrain the W+heavy flavor k-factor independently of possible single top contributions. For the dedicated single top searches, the total backgrounds are generally allowed to float, and more sophisticated

purely 'shape-based' variables are used to discriminate signal from background.

4.4 Bump Hunter

The bump hunter is a new feature added in the second round of this analysis, to enhance the sensitivity of the search to new physics involving narrow mass resonances.

4.4.1 Strategy

The idea is to scan the spectrum of most mass variables with a sliding window. The window needs to vary in width to follow the changing detector resolution. As the window drifts accross a mass distribution, it evaluates the probability that the amount of data therein, or even more, could have emerged by fluctuation from the predicted population. The window where this probability is smallest contains the most interesting local excess of data.

In each final state there are typically several mass variables to scan. On average there are $5036/399 \simeq 13$. They include masses of all combinations of reconstructed objects, such as pairs, triplets, or bigger ensembles.

The width of the sliding window equals two times the characteristic mass resolution for the given combination of objects and at the given mass value. Mass uncertainty results from uncertainties about the specific energies and momenta of all objects involved. It is possible to have combinations of four-momenta that result in the same mass, but different mass uncertainties. For example, if a Z^0 decays to e^+e^-, the mass of that pair will always be close to the nominal $m_Z \simeq 91$ GeV, though its resolution will depend on the boost of the decaying Z^0. Obviously, each event has a different mass uncertainty, so we need to estimate the characteristic mass resolution at each value of mass and for each mass variable. That characteristic mass resolution will be representative of the mass resolution of the events there. To estimate it, we assume that all objects in the ensemble have equal momentum, negligible mass, and their momenta balance on a plane[3]. Then, we

---

[3]If the (equal) momenta are two, to balance they have to be back-to-back. If they are three, they

assign to each involved individual energy the appropriate uncertainty, depending on what object it belongs to, since different objects are measured with different energy resolutions. For electrons and photons, the uncertainty is assumed to be $\Delta E_{\text{EM}} = 0.14\sqrt{E} + 0.015E$, determined by the electromagnetic calorimeter. For jets and τs it is taken to be $\Delta E_{\text{HAD}} = E\sqrt{0.457/E + 20.3/E^2 + 0.00834}$, determined by the hadronic calorimeter. For (beam constrained) muons it is $\Delta E_\mu = 0.0005E^2$, determined by the COT track curvature resolution. In cases of transverse mass involving $\not{p}$, we assume roughly $\Delta E_{MET} = 3\sqrt{\not{p}}$. We propagate those ΔEs corresponding to the members of the ensemble into the system's total mass. For example, if we want to find the characteristic mass resolution for a (e^+, μ^-, j) triplet at system mass 90 GeV, we have $m = \sqrt{(E_e + E_\mu + E_j)^2 - (\vec{p_e} + \vec{p_\mu} + \vec{p_j})^2}$. We assume $E_e = E_\mu = E_j \equiv E$ and the planar configuration with zero net momentum, to obtain that $m = 3E$, hence $E = 30$ GeV for each object. We use the above formulas for the three different ΔEs, keeping in mind the different resolutions for electrons, muons and jets, and then we propagate those uncorrelated uncertainties to the mass, to find $\Delta m = \sqrt{(\Delta E_e)^2 + (\Delta E_\mu)^2 + (\Delta E_j)^2} = 6.57$ GeV.

The step size by which the window drifts equals half a characteristic mass resolution, therefore it varies along the mass spectrum, as the width does too. That way there are no gaps left between consecutive windows. Instead, consecutive windows partly overlap.

Each window comes with two sidebands, extending on each side as far as the window's width. The region of the spectrum that is scanned is slightly narrower than the whole spectrum's span (defined as the interval between the highest-mass and the lowest-mass event in both data and background), so that all considered windows have sidebands lying within the spectrum.

As the window drifts along a mass spectrum, its p-val is calculated at each location. That is defined as the Poisson probability that the Standard Model events expected

---

have to be on the same plane, each separated by 120° from its first neighbors. If we have $N \geq 4$ equal, balancing momenta in 3 dimensions, then their angular configurations can be significantly more complicated, as there are many possible arrangements that satisfy the condition of ballance. To avoid such complexity, we choose to constrain all N vectors in one plane, and assume the solution where all vectors have separation $\frac{2\pi}{N}$ from their nearest neighbors.

in the window (b) would fluctuate up to or above the observed data (d), i.e. p-val = $\sum_{i=d}^{\infty} \frac{b^i}{i!} e^{-b}$.

A window qualifies as a bump if it satisfies the following criteria:

- The central region must contain at least 5 data events.

- Both sidebands must be less discrepant than the central region, i.e. both must have larger p-val.

- If the background in a sideband is non-zero, then it must have p-val $> \int_5^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx$, namely it must not exhibit a significant (5σ) discrepancy. If the background is zero, then it must have less than 5 data[4].

- The above criteria need to hold even when we consider the possible effects of low Monte Carlo statistics in the background. This is explained next.

It can happen to have a great excess of data in the central window, and simultaneously non-discrepant sidebands, but realize that the sidebands contain only a couple of very large-weight events in the Standard Model background. These large-weight events are called "spikes", and are the result of limited Monte Carlo statistics. That bump would potentially pass all quality criteria, and appear to be statistically significant, but it would be prudent to treat conservatively the presence of spikes in the sidebands, and consider that these Monte Carlo events could easily have been in the central window instead. In that case, the p-val of the central window would be larger (less significant) and the sidebands would have a higher probability to be discrepant, hence the bump could disqualify. Since limited Monte Carlo statistics are a practical limitation, we have to be conservative and eliminate, if necessary, this bump. To do that, we first need to define what we consider as a spike in each sideband, and reevaluate the p-val and the quality of the bump, assuming the spikes from both sidebands transfered into the central window. To define the weight of spikes in a

---

[4]This special treatment of the zero-background case is to be able to spot excesses of data that may be isolated at mass values where there is no Standard Model background at all. If we had, for example, 6 events in the central window and 1 event in the sideband, we wouldn't like this band to disqualify due to having a discrepant sideband.

sideband, we look for outliers among the Monte Carlo events, namely for events with significantly larger weights than the average weight of the events in the sideband. We find the average weight and the standard deviation of weights in the sideband, including in the calculation all Monte Carlo events therein. If there is an event whose weight lies beyond 3 standard deviations from the average, then we gradually reduce its weight. As we reduce it, we reevaluate the average and standard deviation of weights. If along its path towards smaller weight it meets another event of same weight, then their weights are bound to be equal from then on, and keep being gradually reduced together. To visualize this process, imagine the axis of weights as an horizontal stretched string, and the weight of each event represented by the position of a tiny bead along this string; the larger the weight, the farther on the right the bead is located. If there are significant outliers, namely beads very far on the right, we start pushing the rightmost bead slowly from right to left, to bring it closer to the others. On its way, the rightmost bead drags with it any beads it meets, since beads can not pass through each other. We stop this reduction of weights when they are all within 3 standard deviations from their average. Then, we compare the total initial weight to the total final weight in the sideband. The difference is weight attributed to spikes. If this difference turns out to be smaller than the largest single weight in the sideband, then we define the latter as spike instead. For the sake of saving time, we do not apply the anti-spike treatment described above, unless the p-value of a qualifying bump candidate is smaller than $\int_5^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, since it is not crucial to be conservative, when a bump is not significant to begin with. A demonstration of the effect of the anti-spike treatment is shown in Fig. 4-10.

When a variable's spectrum is scanned from one end to the other, the qualifying bump with the smallest p-val is the most interesting within that variable. Its statistical significance is quantified on first level by its p-val; the smaller the more significant. It is crucial, though, to account for the trials factor due to examining many windows within that spectrum. We need, therefore, to estimate the probability that a qualifying bump candidate of such a small (or smaller) p-val would appear anywhere along the spectrum, if instead of the actual data we had pseudo-data pulled from the
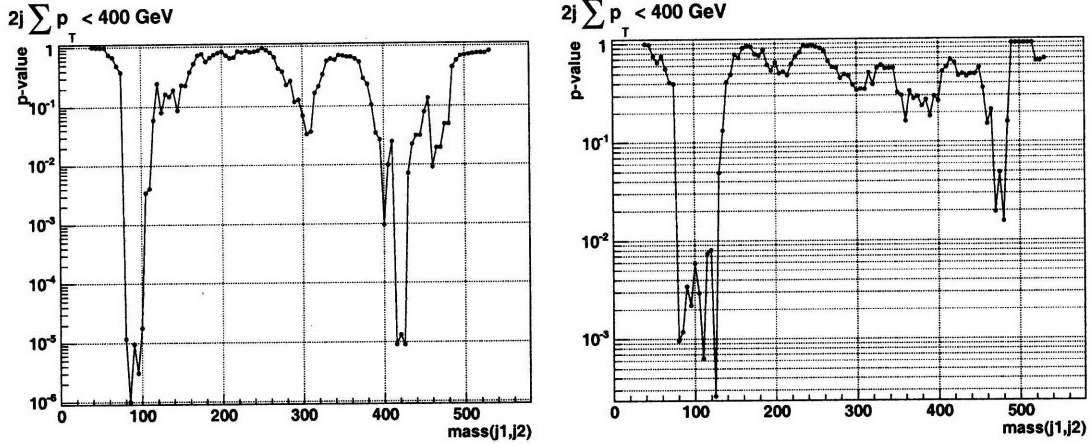
Figure 4-10: *(Left)* The p-val of each bump candidate, as a function of the location of each window's center, along mass($j1, j2$) in final state $2j \sum p_T < 400$ GeV. Bump candidates failing quality criteria have p-val=1. The most significant bump has p-val $\sim 10^{-6}$, which translates to $P_a \sim 3 \times 10^{-5}$ and $P_b \sim 0.15$, therefore all local excesses are insignificant. (Right) For demonstration, we apply the conservative anti-spike treatment to all bump candidates. The result of anti-spike treatment is to have larger p-values and the reduction of significance is greater in regions like around 400 GeV, where Monte Carlo statistics are poorer, therefore spikes contribute more.

Standard Model distribution. We denote this probability P_a, and it can be estimated either experimentally (by producing many sets of pseudo-data and scanning them for more interesting bumps), or using a semi-analytic calculation.

The semi-analytic method, whose goal is to save the enormous time-cost of using Monte Carlo to experimentally evaluate P_a for all mass variables, proceeds as follows: For each window and its sidebands, we estimate with Monte Carlo the probability that it would satisfy quality criteria $(P(Q))$, if the data populations in the center and in the sidebands were pulled randomly from the respective expected populations therein. Let's denote the p-val of the most interesting bump in the actual data p-val$_{min}$. Denote the probability that a window would have p-val $\leq p$-val$_{min}$ as $P(S)$. The probability that a window would qualify and simultaneously have p-val $\leq p$-val$_{min}$ is $P(Q \wedge S) \simeq P(Q) \, P(S)$, where we assumed that Q and S are independent. This is not generally true, but holds approximately in most cases. In fact $P(Q|S) = \frac{P(Q \wedge S)}{P(S)} \geq P(Q)$, because if S is true then we have a significant excess of data in the central window, which makes it somewhat less likely for the sidebands to exhibit a bigger discrepancy

than the center, hence it's more likely that quality standards (Q) will also be met. So, $P(Q) \, P(S) \leq P(Q|S) \, P(S) = P(Q \wedge S)$, i.e. we slightly underestimate $P(Q \wedge S)$ by approximating it with $P(Q) \, P(S)$. $P(S)$ is approximately $pval_{\min}$, but that is exactly correct only as long as there is an *integer* number of data that, given the background in the window, would result in a p-val of *exactly* p-val$_{\min}$. If that is not the case, then $P(S) \leq pval_{\min}$, because to exceed in significance the most interesting bump, this window would need to exhibit a p-val not just equal to p-val$_{\min}$, but smaller. For example, if p-val$_{\min} = 0.01$ and the background is $b = 3.2$, then to exceed p-val$_{\min}$ in significance we need the data to be at least $d = 9$. If $d = 8$ then p-val $= 0.016 > 0.01$. However, if $d = 9$ then p-val $= 0.0057$, which means that the true $P(S)$ in this example would be 0.0057 instead of 0.01. This difference becomes negligible for large backgrounds, where one event more or less changes p-val negligibly.

We find, as described, $P(Q \wedge S)$ for all windows considered along the spectrum, and set $P_a = 1 - \prod(1 - P(Q \wedge S))$, namely the probability that at least one window would qualify and surpass in significance the most interesting bump in the actual data. Here, another assumption is implicit: that windows are independent.

A comparison between the semi-analytic (fast) and the experimental method to estimate P_a is shown in Fig. 4-11. Pseudo-data were pulled from all mass distributions, and then both the slow and the fast methods were used to estimate P_a. The comparison shows that, for pseudo-data, the fast method returns a P_a which is, when translated into units of standard deviation, within about 1σ from the P_a determined by the slow method. This difference reflects on the expected distributions of P_a from all mass variables when using the two methods. While the slow method returns a P_a with uniform expected distribution, the fast method's P_a is distributed as shown in Fig. 4-12.

The slow method does not rely on any approximation, therefore its answer is more representative of the true P_a. It is only limited by the number of pseudo-data sets that we can generate. Its disadvantage is that even when applied on just one mass variable to estimate the significance of its most interesting bump, it can take prohibitively long. How long depends on the number of expected events in the final state where

the mass variable belongs, but more importantly on the smallness of p-val$_{min}$. For really significant bumps (p-val$_{min} \lesssim 10^{-8}$) it may take millions of sets of pseudo-data to start resolving P_a experimentally. The slow method returns the best estimate of P_a it could obtain within the amount of time it was allowed to run. If during this amount of time it is clear at 95% confidence level that P_a is either greater or smaller than what corresponds to a 5σ effect ($\int_5^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 2.87 \times 10^{-7}$), then the slow method returns the estimated value of P_a at that time, since the conclusion is clear and additional accuracy would be of no use. Due to its great time cost, we employ the slow method only if the fast (semi-analytic) method has returned a significant enough P_a, i.e. smaller than what corresponds to a 4.5σ effect. The final significance of a bump is not quantified by P_a, but by P_b (defined later), which includes the whole trials factor. For P_a equivalent to 4.5σ, P_b is 2.1σ, safely away from the discovery threshold of 3σ in P_b, which corresponds to P_a of 5σ. This is mentioned to explain that the slow and more accurate estimator for P_a is employed not just beyond the discovery threshold, but safely earlier, when a bump starts being mildly significant.

Since P_a encompasses the trials factor from examining multiple windows within the mass variable, it characterizes the significance of the mass viariable in terms of its most interesting bump. The next question is what the probability is that in a pseudo-experiment, where data are pulled from the Standard Model epxectation, any mass variable would appear with a P_a smaller than the actual P_a of the mass variable. We denote this probability as P_b. We estimate it assuming all mass viariables are statistically independent trials, therefore $P_b = 1 - (1 - P_a)^N$, where N is the total population of scanned mass variables from all VISTA final states.

In summary, for each mass variable the most interesting bump is the one with the smallest p-val, and with all trials factor accounted for, its significance is approximately given by P_b. Then P_b is converted to units of standard deviations, and if it corresponds to a 3σ effect or more, then we consider it a discrepancy worth pursuing.
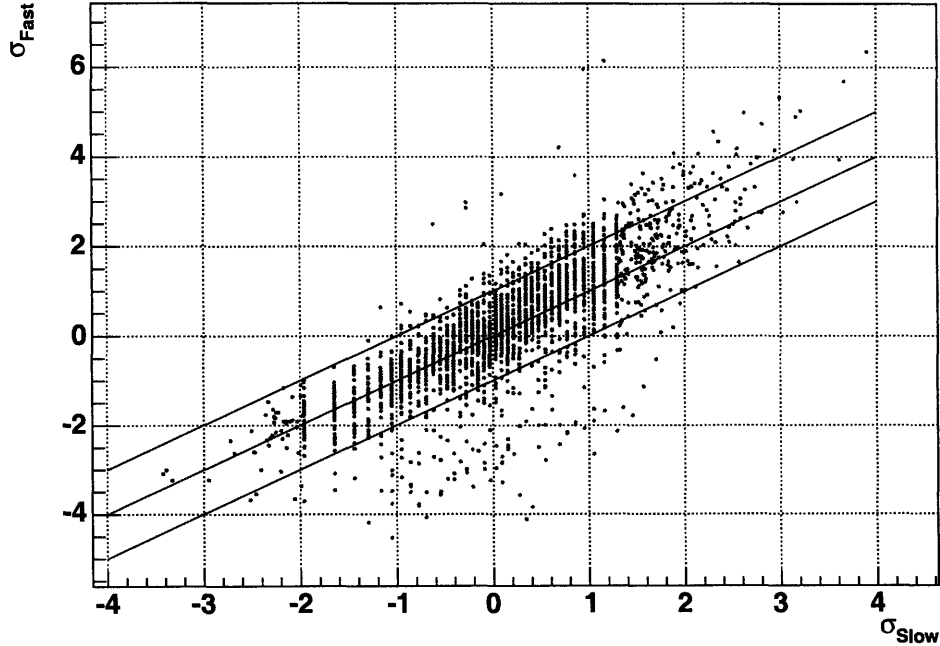
Figure 4-11: Comparison of fast versus slow method to estimate P_a. Each point corresponds to a mass variable with at least one qualifying bump in pseudo-data. The three lines indicate the locus where the fast estimate of P_a is equal to, or $\pm 1\sigma$ away from the slow estimate of P_a. Slow P_a can be only a rational number, since it is the fraction of two integers, namely the number of pseudo-data distributions with a more interesting bump and the total number of tried pseudo-data distributions. That is why the slow P_a appears to assume a discrete spectrum of values.
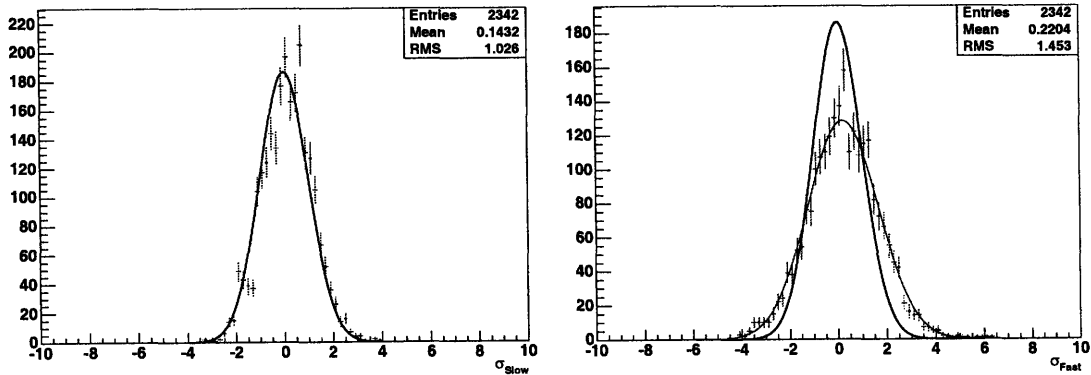
Figure 4-12: Expected distribution of the fast and the slow estimator of P_a, when applied on pseudo-data. The slow estimator (left) is distributed according to a normal distribution (except for some recurrent values which reflect that the slow estimator can only be a rational number), while the fast one (right) follows a Gaussian probability density function with mean 0.2204 and standard deviation 1.453. In the right plot, the Normal distribution has been drawn for comparison.

140

4.4.2 Results

The summary of the most interesting bump in each mass variable is shown in Fig. 4-13.

The only mass variable with its most significant bump exceeding the discovery threshold is the mass of all four jets in the final states with four jets of $\sum p_T < 400$ GeV, shown in Fig. 4-14. This is attributed to the "3-jet" effect, the main cause of all shape discrepancies in this analysis. Fig. 4-15 shows another instance of the same effect in that final state. The same effect is observed in final states of different jet multiplicity, as shown in Fig. 4-16.

Although no discovery-level bumps were found in other mass variables, it is interesting to present the most interesting bumps found in some mass distributions.

In the mass of the (e^+, e^-) pair in the final state with two opposite sign electrons $(e^+ e^-)$ the most significant bump corresponds to a 2.7σ effect, which is though exactly at the Z boson resonance. The number of expected events there is so high, that even the slightest systematic mismodeling would appear as very statistically significant. From Fig. 4-17 it is clear that this "bump" is not due to new physics, but a tiny systematic mismodelling of the Z-peak, with no visible effect anywhere else.

The mass of the two muons in the $\mu^+ \mu^-$ final state does not have any significant bump either, not even of the mundane kind found in $e^+ e^-$. That is shown in Fig. 4-18.

Another potentially interesting mass variable is the dijet mass in the final state with two high $\sum p_T$ jets. That is shown in Fig. 4-19. Unfortunately, no high-mass di-jet resonance was observed.

4.4.3 Sensitivity

To test the sensitivity of the Bump Hunter, we generate some specific new physics signal, pass it through the full CDF detector simulation, and inject it gradually on top of pseudo-data pulled from the Standard Model background, until the Bump Hunter identifies a discovery-level bump.

CDF Run II Preliminary (2 fb⁻¹)

Mass distributions: 5036
With bumps: 2316
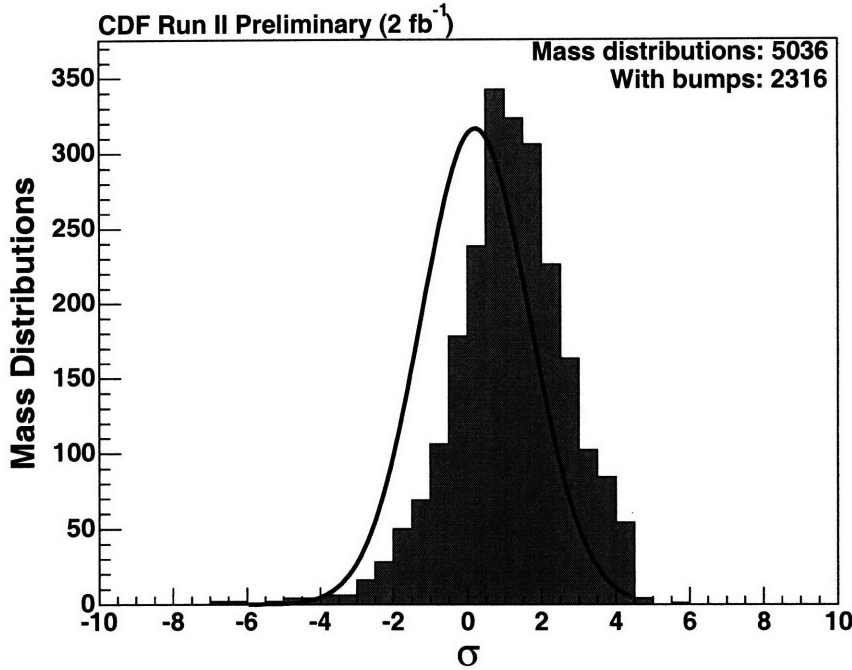
Figure 4-13: Significance of the most interesting bump in each mass variable. Each entry corresponds to one mass distribution found to contain at least one bump satisfying quality criteria. The quantity distributed is P_a, transformed to units of standard deviation (σ), using the formula $P_a = \int_\sigma^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx$. Large P_a translates to a small number of σ and signifies an insignificant effect. The discovery threshold corresponds to 5σ. The entries under 4.5σ have been estimated using the semi-analytic (fast) method, which yields values distributed according to the black curve when applied on pseudo-data agreeing with the Standard Model background. Values above 4.5σ are estimated using the slow, more accurate method. Therefore, values of P_a corresponding to more than 4.5σ can be translated directly to significance, since their expected values follow the Normal distribution. About 5000 mass distributions are considered, which means that to have an effect of significance 3σ after trials factor, it needs to have a significance of 5σ or more in this scale of Pa. Only one mass distribution has its most significant bump exceed this discovery threshold. More details in the text.
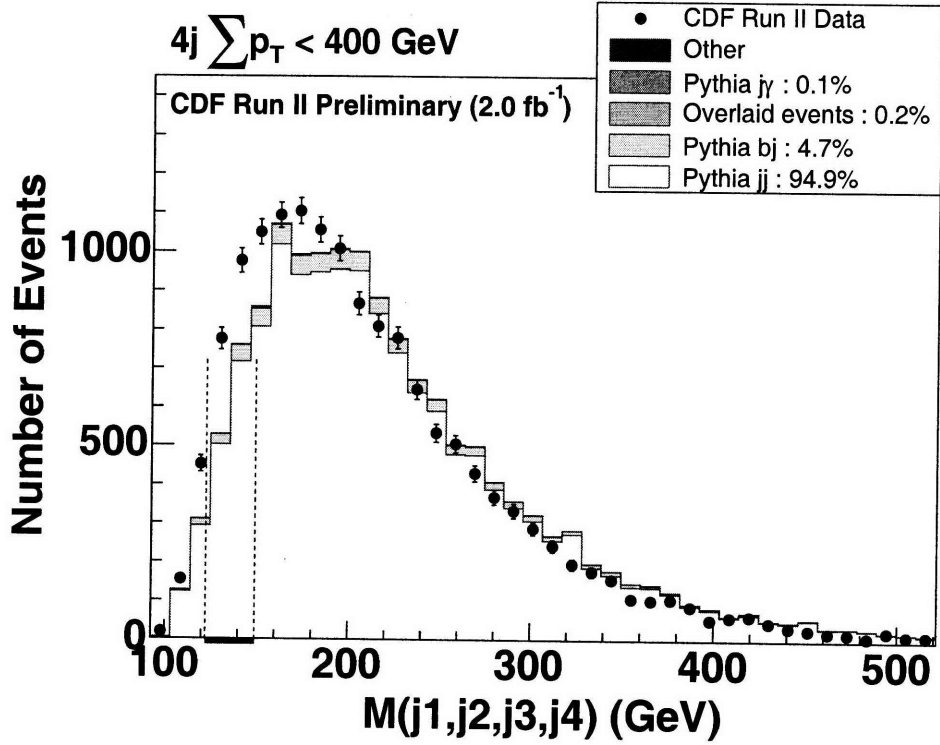
Figure 4-14: The most significant bump found in the $4j \sum p_T < 400$ GeV final state, indicated by the blue lines. Its P_b translates to 4.1σ.

120 GeV Higgs in association with W

The pseudo-signal use for this test contains a Standard Model Higgs of mass 120 GeV, allowed to decay to $b\bar{b}$, which has branching ratio 68% [84]. The associated W decays to e or μ or τ plus neutrino, with total branching ratio $\sim \frac{1}{3}$.

About 6500 signal events are required to obtain the first bump beyond discovery threshold. Events passing selection criteria are distributed in several final states, and 15 of them make it to the $2be^+\not{p}$ final state, producing the bump in Fig. 4-20. Compensating for the branching ratio, we find that the required cross section of $WH_{120\text{GeV}}$ to have this 5σ level discovery would be about 14.4 pb, which is $\sim$90 times larger than the predicted Standard Model cross section.

$Z' \to \ell^+\ell^-$ at mass 250 GeV

Pseudo-signal of a 250 GeV Z' boson was generated, where Z' may decay to $\ell^+\ell^-$, where ℓ can be e, μ, or τ. The first discovery-level bump caused after injecting about
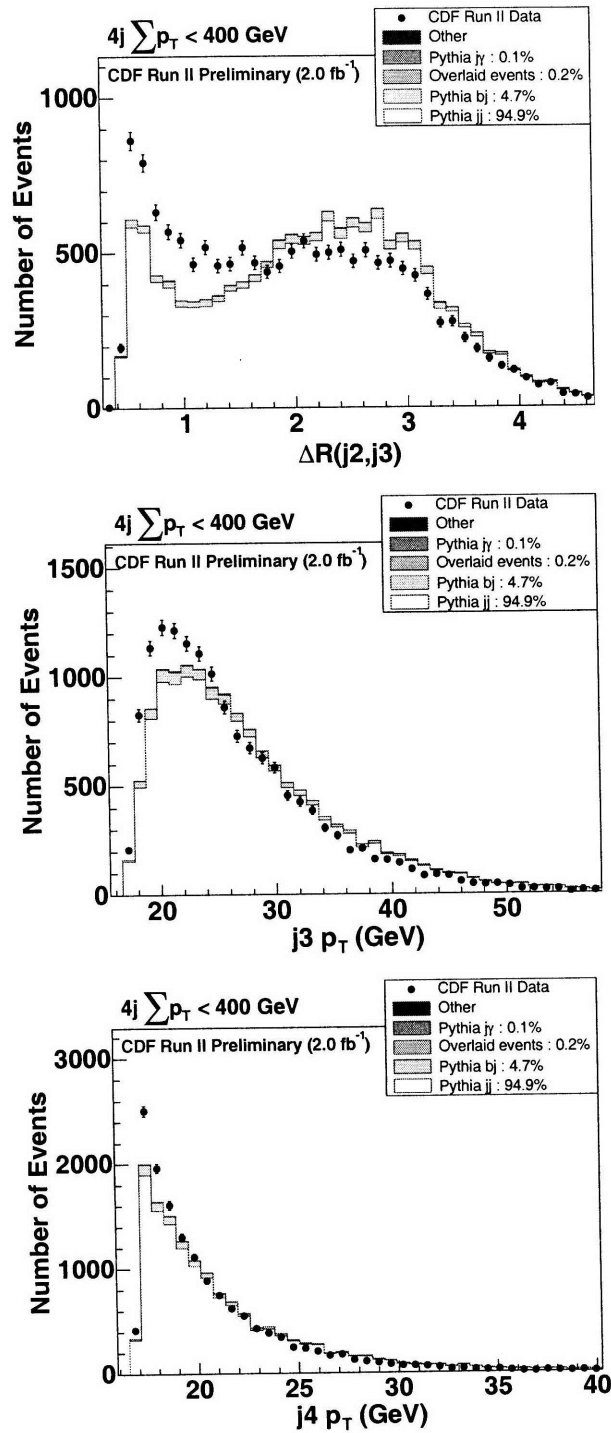
Figure 4-15: *(Upper)* The "3-jet" effect appearing in the angular separation between the second ($j2$) and the third ($j3$) leading jets, in final state $4j$ $\sum p_T < 400$ GeV. There is an excess of soft final state radiated jets emitted at small angles. The *lower two* distributions from the same final state demonstrate exactly this excess, which is not present in the p_T of the first and second leading jets.
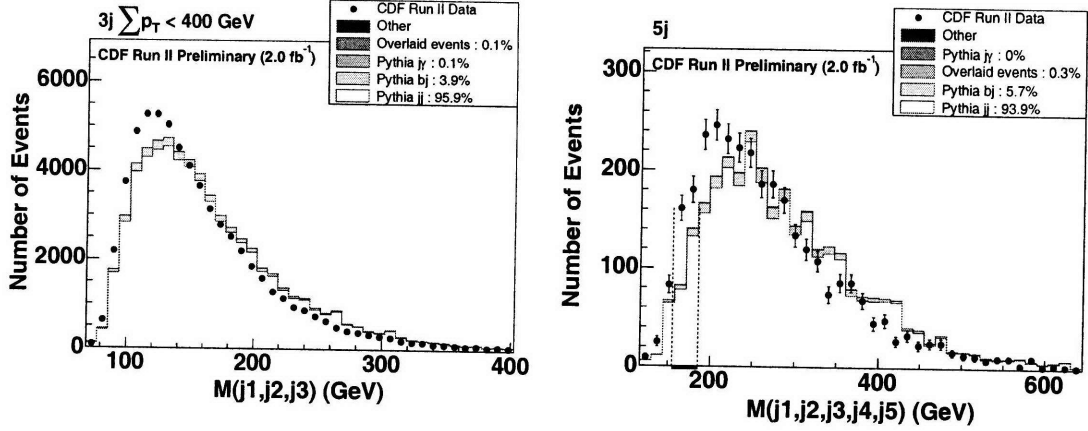
Figure 4-16: The "3-jet" effect appearing in the mass of all jets in the final state with three *(left)* and five *(right)* jets. The excess is similar to the one identified as a bump in the $4j \sum p_T < 400$ GeV final state. The difference in the case of $3j \sum p_T < 400$ GeV is that the excess is wide and the sidebands are discrepant, making this bump candidate disqualify, while in the case of $5j$ the excess satisfies bump quality criteria, but has P_b corresponding to only 1.5σ.

700 events of this pseudo-signal. 55 events end up in the 1e+1e- final state, and form the bump shown in Fig. 4-21.

With 700 injected events the significance found is 3.7σ, which is higher than the discovery threshold of 3σ. That is because the pseudo-signal is injected in bunches of 100 events, so the actual requirement is between 600 and 700 events. Dividing this number of generated events by our integrated luminosity shows that we would need the cross section times branching ratio of this signal to be approximately 0.325 pb.

$Z' \to t\bar{t}$ at mass 500 GeV

For this test we generated Z' events of mass 500 GeV, where the heavy boson decays to a $t\bar{t}$ pair. Injecting 5000 such events causes simultaneously two significant bumps in the $be^+3j\not{p}$ final state; one is in the transverse mass between $\not{p}$ and the second highest p_T jet ($j2$), with significance 3σ; the other is in the transverse mass of the third highest p_T ($j3$) and $\not{p}$, with significance 3.2σ. The latter is shown in Fig. 4-22.

In another instance, after injecting 4600 different pseudo-signal events, a 3.3σ effect after trials factor was created in the same final state ($be^+3j\not{p}$), but this time in the variable $m_{t\bar{t}}$, where one would more easily interpret the excess as due to resonant
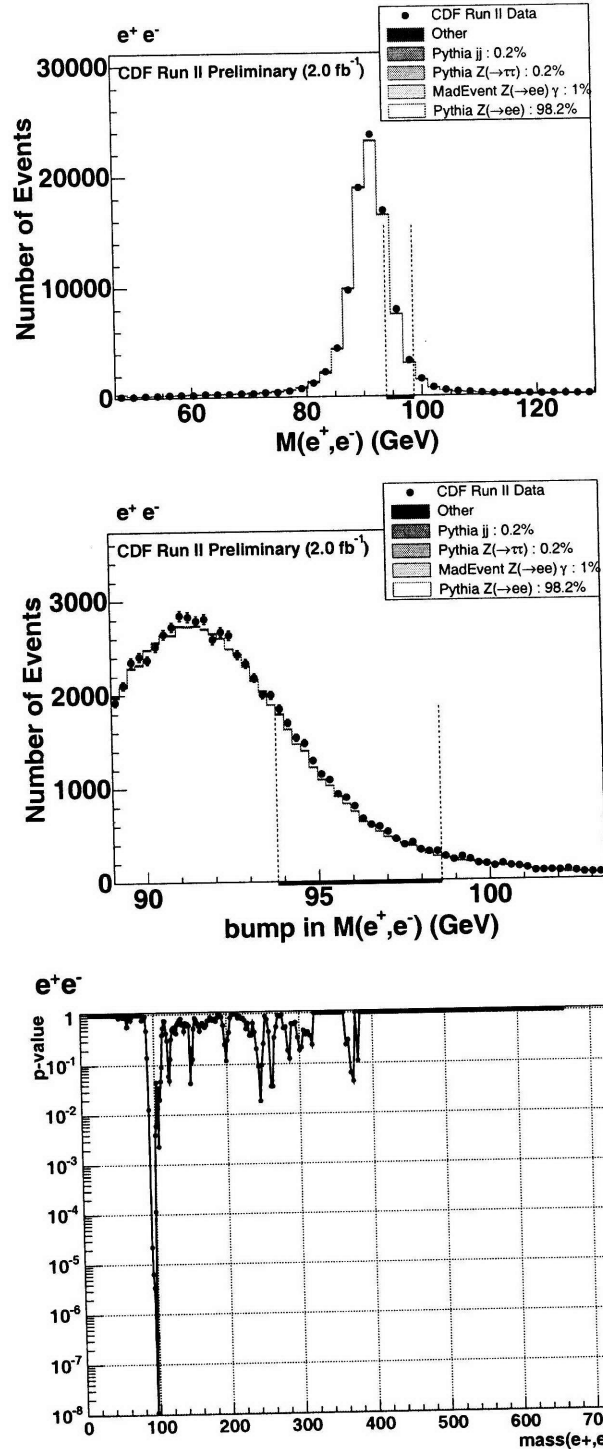
Figure 4-17: *(Upper two)* The most interesting bump found in final state e^+e^-. *(Bottom)* The p-val of all bumps accross the mass spectrum of the two leptons. Apart from this discrepancy at the Z-peak, which corresponds to a 2.7σ effect after trials factor and reflects only a tiny mismodeling in a region with very high statistics, no other significant bump was found.
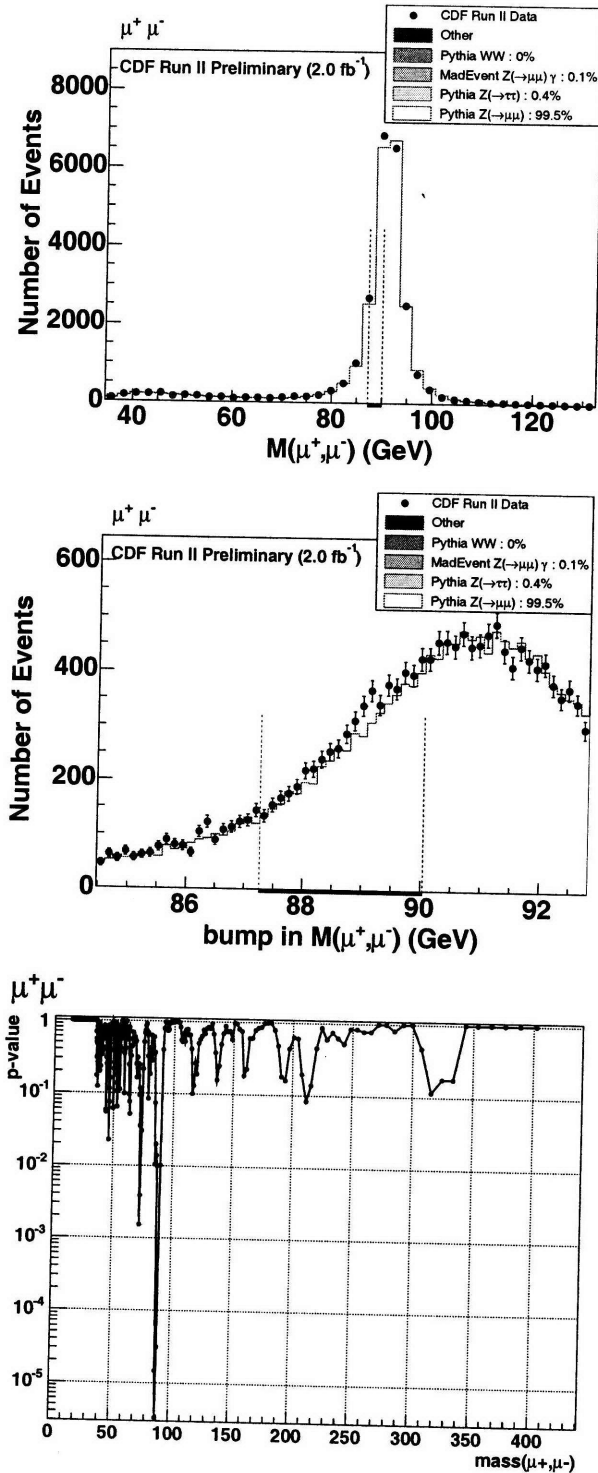
Figure 4-18: *(Upper two)* The most interesting bump found in final state $\mu^+\mu^-$. *(Bottom)* The p-val of all bumps accross the mass spectrum of the two leptons. Even the most significant bump, at the Z-peak, has $P_b \simeq 0.74$, therefore is completely insignificant.
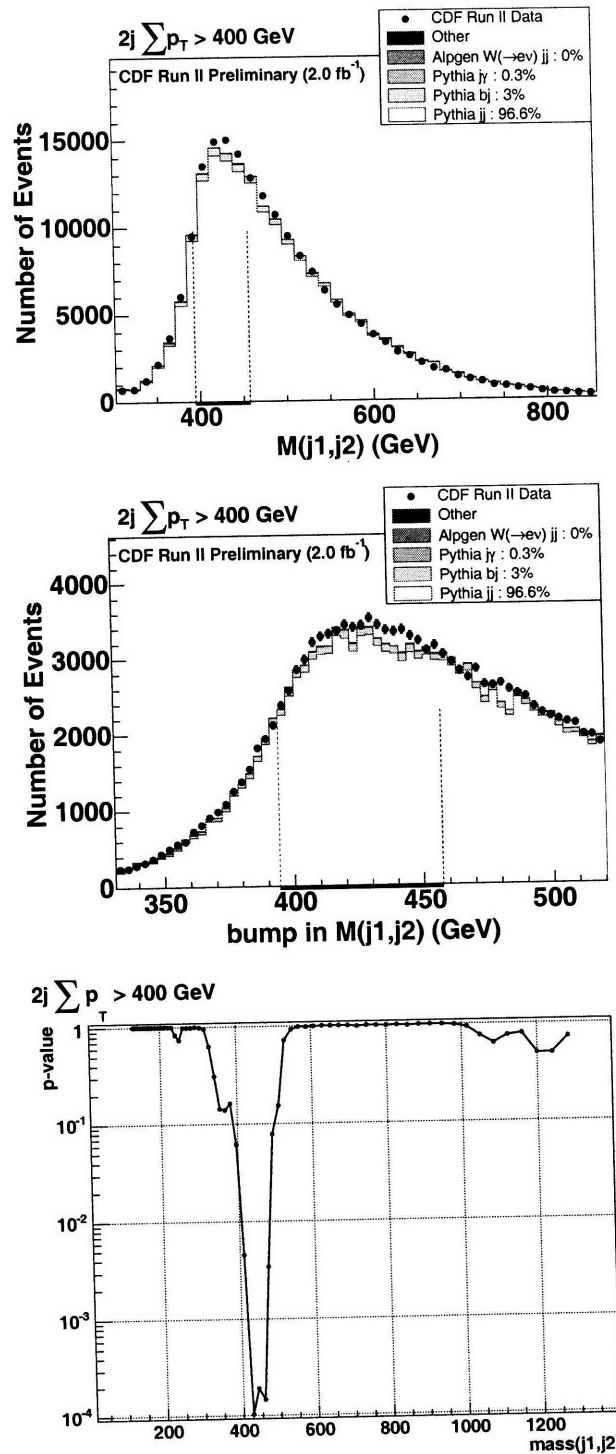
Figure 4-19: *(Upper two)* The most interesting bump found in final state $2j \sum p_T > 400$ GeV. *(Bottom)* The p-val of all bumps accross the di-jet mass spectrum. Even the most significant bump, yields $P_b \simeq 0.99$, therefore is completely insignificant.
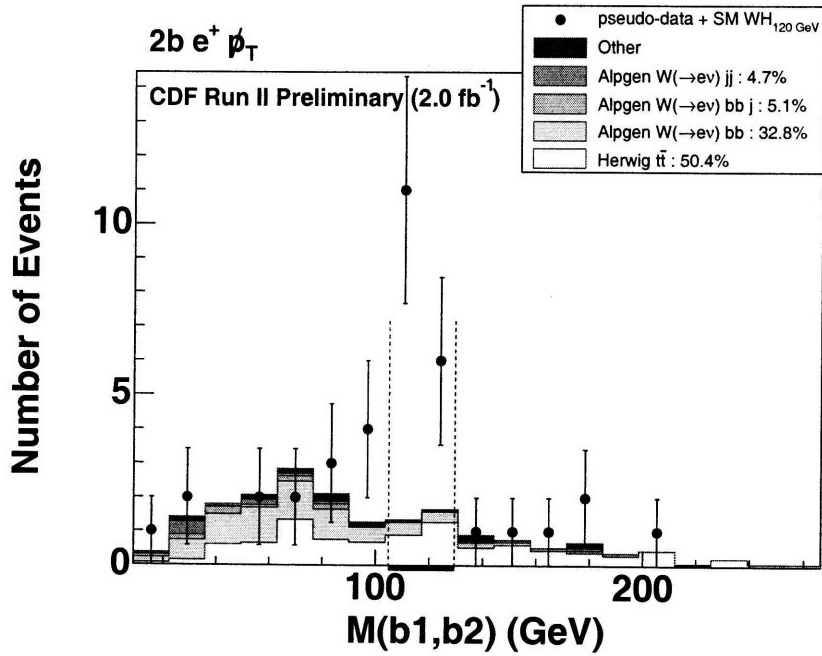
Figure 4-20: Example of a pseudo-discovery of the Standard Model Higgs boson ($m_H = 120$ GeV), produced in association with a W boson. Out of 7000 generated $WH(\to \ell\nu b\bar{b})$ events, 15 populate the $2be^+\not{p}$ final state. They cause this local excess which is identified by the Bump Hunter algorithm and its significance is estimated a 3.4σ after trials factor.
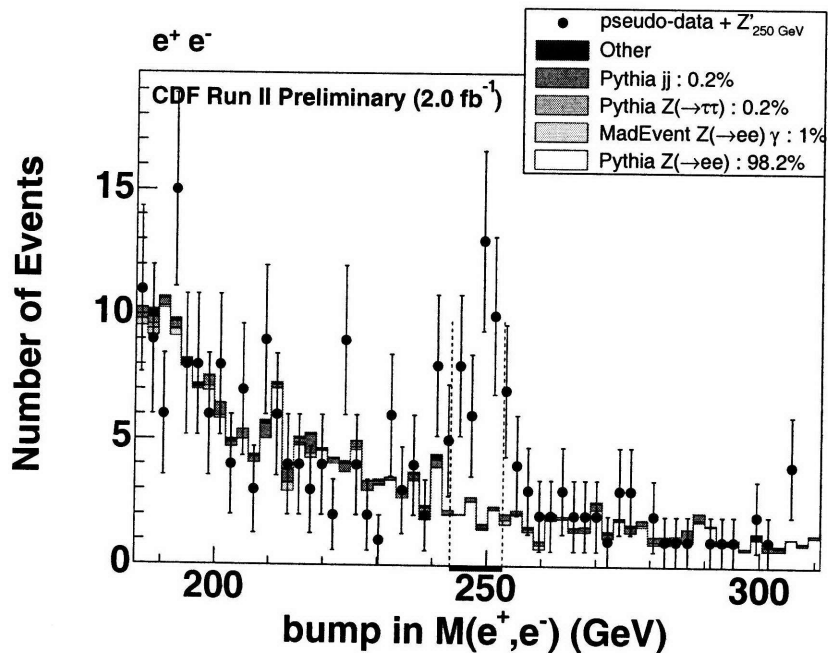
Figure 4-21: Example of a pseudo-discovery of a 250 GeV Z' decaying to charged leptons. Out of 700 generated events, 55 populate the e^+e^- final state, where the most significant bump appears. The significance of this bump is estimated at 3.7σ after trials factor.
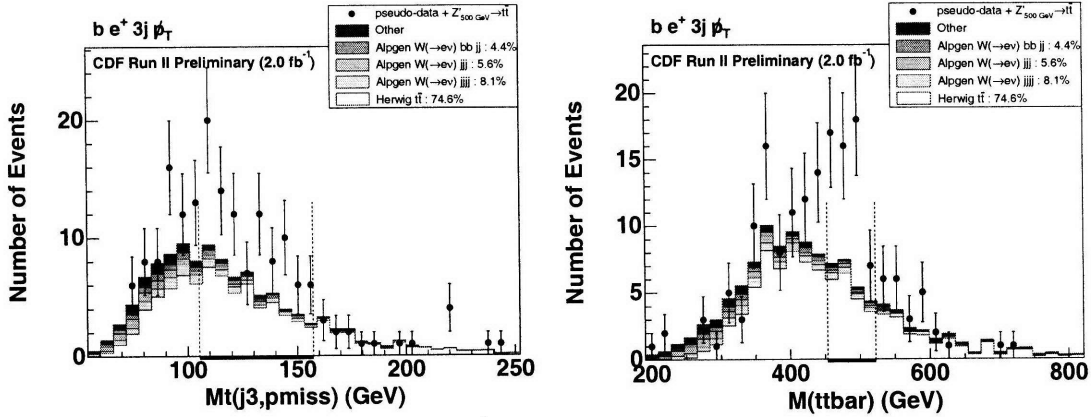
Figure 4-22: *(Left)* Most significant bump after injecting 5000 $Z'_{500GeV} \to t\bar{t}$ events. 47 signal events make it to the $be^+3j\not{p}$ final state, which cause this bump of significance 3.2σ after trials factor in transverse mass of the third highest p_T jet and $\not{p}$. *(Right)* Most significant bump after injecting a different 4600 $Z'_{500GeV} \to t\bar{t}$ events, on a background that was allowed to fluctuate anew. 41 signal events make it to the $be^+3j\not{p}$ final state, which cause this bump of significance 3.3σ after trials factor in $m_{t\bar{t}}$, consistent naturally with the mass of the introduced Z'.

production of $t\bar{t}$. That is shown in Fig. 4-22 as well.

With discovery cost of approximately 4800 events, the required cross section is approximately 2.4 pb.

4.5 Summary of second round with 2 fb^{-1}

VISTA and SLEUTH search for outliers, representing significant discrepancies between data and Standard Model prediction. Unfortunately, the result obtained is that no signficant outliers have been found either in the total number of events in the VISTA exclusive final states, or in SLEUTH's search of the $\sum p_T$ tails. Disregarding effects from tuning corrections to the data, SLEUTH's $\tilde{\mathcal{P}}$ provides a rigorous statistical calculation of the likelihood that the most discrepant SLEUTH final state seen would have arisen purely by chance from the Standard Model prediction and correction model constructed within VISTA.

VISTA's correction model does not explicitly include some sources of systematic uncertainty, including those associated with parton distribution functions and showering parameters in the event generators used; these sources of uncertainty are included

150

implicitly, in that they would be considered if necessary in the event of a possible discovery. Other uncertainties related to the modeling of the CDF detector response and object identification criteria are determined as part of VISTA but are not included in the calculation of $\tilde{\mathcal{P}}$. For the correction model used, SLEUTH finds $\tilde{\mathcal{P}} = 0.085$.

The Bump Hunter, a new algorithm for identification of mass resonances, did not find any significant mass bumps either, except for one that is attributed to PYTHIA not modeling perfectly parton showering.

Although the VISTA correction model could presumably be improved further to show even better agreement with Standard Model prediction, finding $\tilde{\mathcal{P}} \gg 0.001$ indicates that even the most discrepant $\sum p_T$ tail is not of statistical interest. The correction model used is thus good enough (even without considering effect of systematic uncertainties on the SLEUTH final states) to conclude this search for outliers using VISTA and SLEUTH in 2 fb^{-1}.

This analysis does not prove that there is no new hint of physics buried in these data; merely that this search does not find any.

Chapter 5

Grand Summary and Conclusion

This thesis presents the first model-independent search for new physics of such scope.

The Standard Model was implemented using a simplified set of corrections.

New physics was sought that would cause significant discrepancies in (a) populations of exclusive final states, (b) shapes of kinematic distributions, (c) mass spectra, and (d) high-$\sum p_T$ events.

The search was first conducted in 1 fb^{-1} of CDF II data, revealing no ground on which to support a discovery claim. It was then repeated in 2 fb^{-1} of data, improved and enhanced with the Bump Hunter, an algorithm to locate narrow resonances due to new massive particles.

Unfortunately and surprisingly, even with 2 fb^{-1} the result was null, in the sense that no new physics could be claimed with the findings. The discrepancies seen were attributed mainly to the difficulty in modeling soft radiated parton showers with PYTHIA. This issue was suspected to be problematic, but no other analysis had illustrated so clearly its repercussion.

Although no single analysis can guarantee that new physics is nowhere in the data, it is highly informative that in a search of this scope nothing exploitable was found. This is complemented and consented by the numerous searches, dedicated to specific signals, which so far have failed too to reveal what lies beyond the Standard Model.

Even with a null result, the value of this technique is great in providing an overview of all data, even those nobody ever considers. It can make a big difference at the later

stages of the LHC, or in any experiment where there is a proliferation of data, and a fairly accurate theoretical prediction analogous to what our event generators and detector simulation provide.

Appendix A

Correction Model Details

Some aspects of the correction model are fixed, rather than dynamically adjusted by the global fit, which is viewed as just a tool to provide reasonable values for some parameters of the correction model. Not every parameter needs to be determined by a fit, as long as it is reasonable or estimated beforehand, through a MC study for instance.

Implementation details of the correction model will be described in this chapter in some extra detail.

A.1 Fake rate physics

The following facts begin to build a unified understanding of fake rates for electrons, muons, taus, and photons. This understanding is woven throughout the correction model, and significantly informs and constrains the VISTA correction process. Explicit constraints derived from these studies are provided in Appendix A.3. The underlying physical mechanisms for these fakes lead to simple and well justified relations among them.

Table A.1 shows the response of the CDF detector simulation, reconstruction, and object identification algorithms to single particles. Using a single particle gun, 10^5 particles of each type shown at the left of the table are shot with $p_T = 25$ GeV into the CDF detector, uniformly distributed in θ and in ϕ. The resulting reconstructed object

	e^+	e^-	μ^+	μ^-	τ^+	τ^-	γ	j	b
e^+	62228	33	0	0	182	0	2435	28140	0
e^-	24	62324	0	0	0	192	2455	28023	1
μ^+	0	0	50491	0	6	0	0	606	0
μ^-	0	1	0	50294	0	6	0	577	0
γ	1393	1327	0	0	1	1	67679	21468	0
π^0	1204	1228	0	0	5	8	58010	33370	0
π^+	266	0	115	0	41887	6	95	54189	37
π^-	1	361	0	88	13	41355	148	54692	44
K^+	156	1	273	0	42725	7	37	52317	24
K^-	1	248	0	165	28	41562	115	53917	22
B^+	100	0	77	1	100	10	40	66062	25861
B^-	2	85	3	68	11	99	45	66414	25621
B^0	88	27	87	17	77	32	21	65866	25046
$\bar{B}^0$	17	79	11	71	41	77	21	66034	25103
D^+	126	6	62	0	1485	67	207	79596	11620
D^-	4	134	3	74	64	1400	234	79977	11554
D^0	60	13	27	2	312	1053	248	88821	5487
$\bar{D}^0$	15	46	5	28	1027	253	237	89025	5480
K_L^0	1	4	0	0	71	60	202	96089	26
K_S^0	26	31	2	1	170	525	9715	76196	0
τ^+	1711	13	1449	0	4167	2	673	50866	607
τ^-	12	1716	0	1474	6	3940	621	51125	580
u	8	10	1	0	446	31	247	94074	26
d	3	4	0	0	64	308	191	94322	22
g	2	0	0	0	17	14	12	81865	99

Table A.1: Central single particle misidentification matrix. Using a single particle gun, 10^5 particles of each type shown at the left of the table are shot with $p_T = 25$ GeV into the central CDF detector, uniformly distributed in θ and in ϕ. The resulting reconstructed object types are shown at the top of the table, labeling the table columns. Thus the rightmost element of this matrix in the fourth row from the bottom shows $p(\tau^- \to b)$, the number of negatively charged tau leptons (out of 10^5) reconstructed as a b-tagged jet.

e⁺ e⁻ μ⁺ μ⁻ τ⁺ τ⁻ γ j b

e⁺

e⁻

μ⁺

μ⁻

γ

π⁰

π⁺

π⁻

K⁺

K⁻

B⁺

B⁻

B⁰

B̄⁰

D⁺

D⁻

D⁰

D̄⁰

K⁰_L

K⁰_S

τ⁺

τ⁻

u

d

g
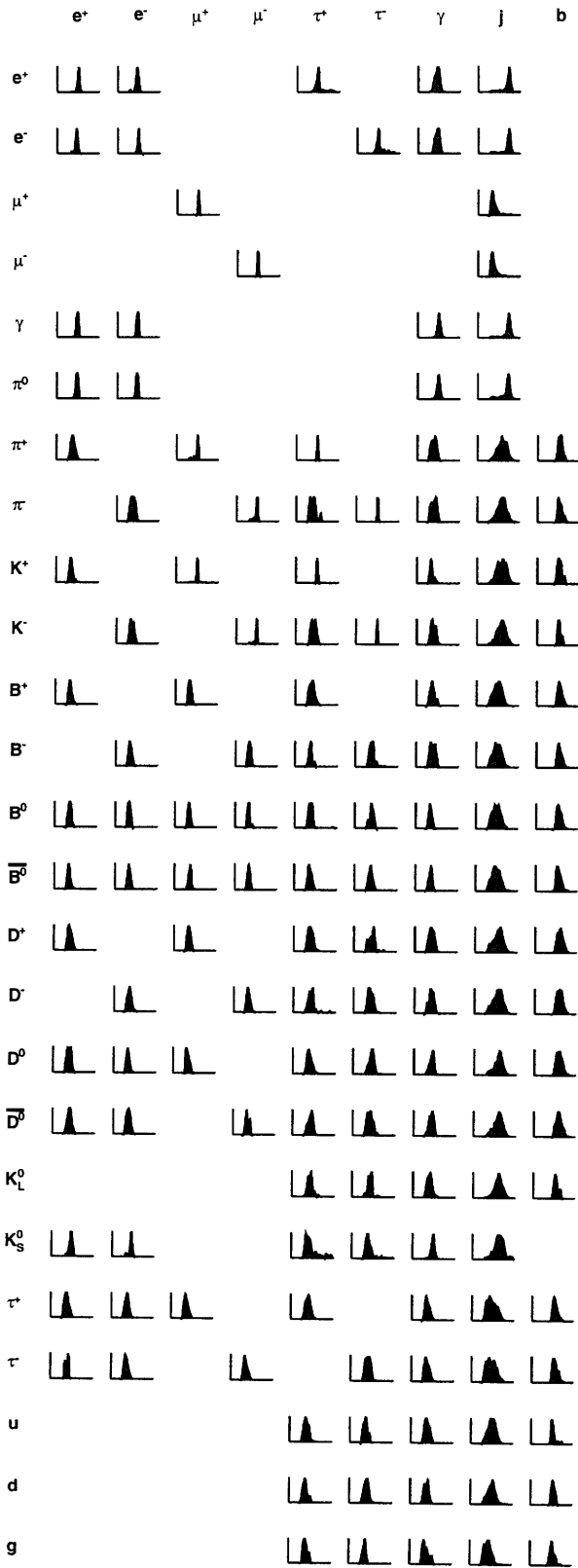
Figure A-1: Transverse momentum distribution of reconstructed objects (labeling columns) arising from single particles (labeling rows) with p_T = 25 GeV shot from a single particle gun into the central CDF detector. The area under each histogram is equal to the number of events in the corresponding misidentification matrix element of Table A.1, with the vertical axis of each histogram scaled to the peak of each distribution. A different vertical scale is used for each histogram, and histograms with fewer than ten events are not shown. The horizontal axis ranges from 0 to 50 GeV.

types are shown at the top of the table, labeling the columns. The first four entries on the diagonal at upper left show the efficiency for reconstructing electrons and muons [1]. The fraction of electrons misidentified as photons (top row, seventh column) is seen to be roughly equal to the fraction of photons identified as electrons or positrons (fifth row, first and second columns), and measures the number of radiation lengths in the innermost regions of the CDF tracker. The fraction of B mesons identified as electrons or muons, primarily through semileptonic decay, are shown in the four left columns, eleventh through fourteenth rows. Other entries provide similarly useful information, most easily comprehensible from simple physics.

The transverse momenta of the objects reconstructed from single particles are displayed in Fig. A.1. The relative resolutions for the measurement of electron and muon momenta are shown in the first four histograms on the diagonal at upper left. The histograms in the left column, sixth through eighth rows, show that single neutral pions misreconstructed as electrons have their momenta well measured, while single charged pions misreconstructed as electrons have their momenta systematically undermeasured, as discussed below. The histogram in the top row, second column from the right, shows that electrons misreconstructed as jets have their energies systematically overmeasured. Other histograms in Fig. A.1 contain similarly relevant information, easily overlooked without the benefit of this study, but understandable from basic physics considerations once the effect has been brought to attention.

Here and below $p(q \rightarrow X)$ denotes a quark fragmenting to X carrying nearly all of the parent quark's energy, and $p(j \rightarrow X)$ denotes a parent quark or gluon being misreconstructed in the detector as X.

---

[1]The electron and muon efficiencies shown in this table are different from the correction factors 0025 and 0027 in Table 4.2, which show the ratio of the object efficiencies in the data to the object identification efficiencies in CDFSIM.
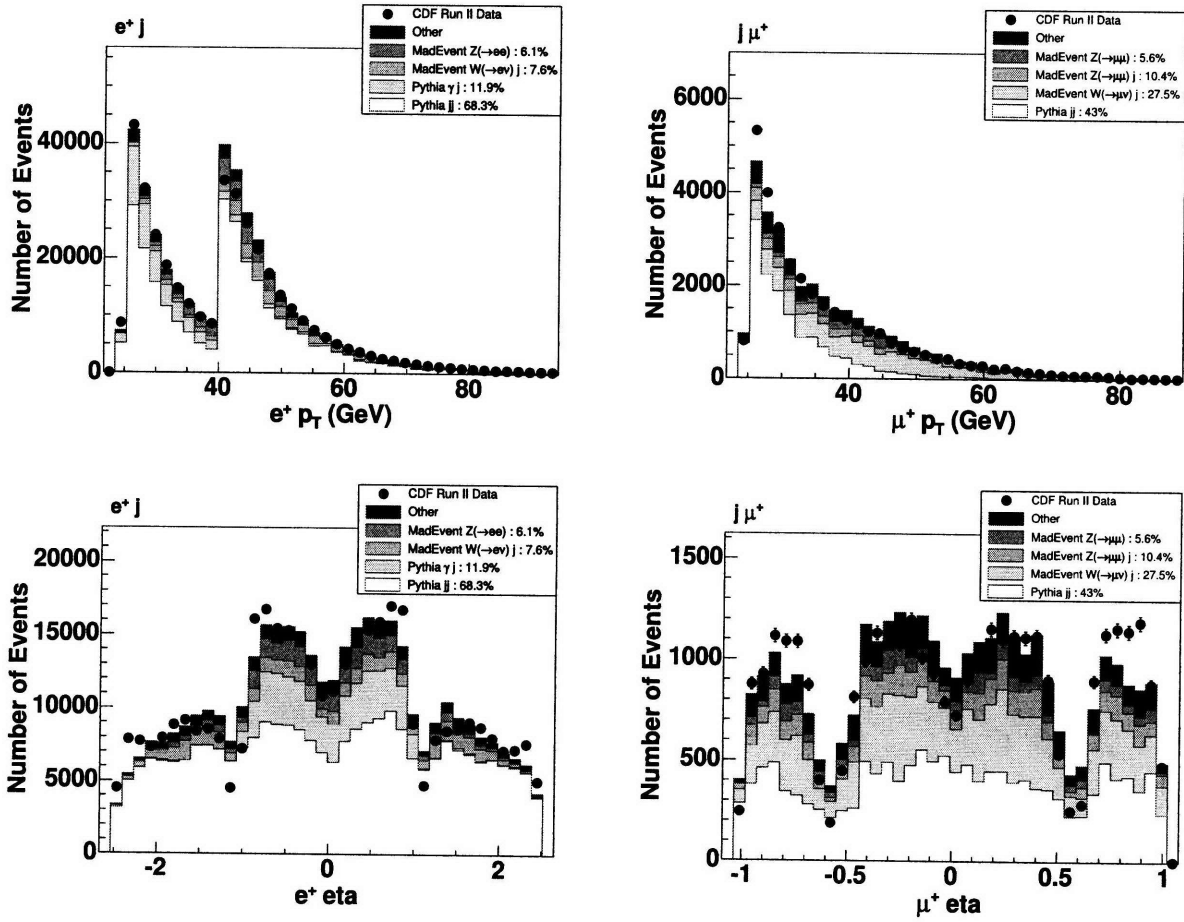
Figure A-2: A few of the most discrepant distributions in the final states ej and $j\mu$, which are greatly affected by the fake rates $p(j \to e)$ and $p(j \to \mu)$, respectively. These distributions are among the 13 significantly discrepant distributions identified as resulting from coarseness of the correction model employed.
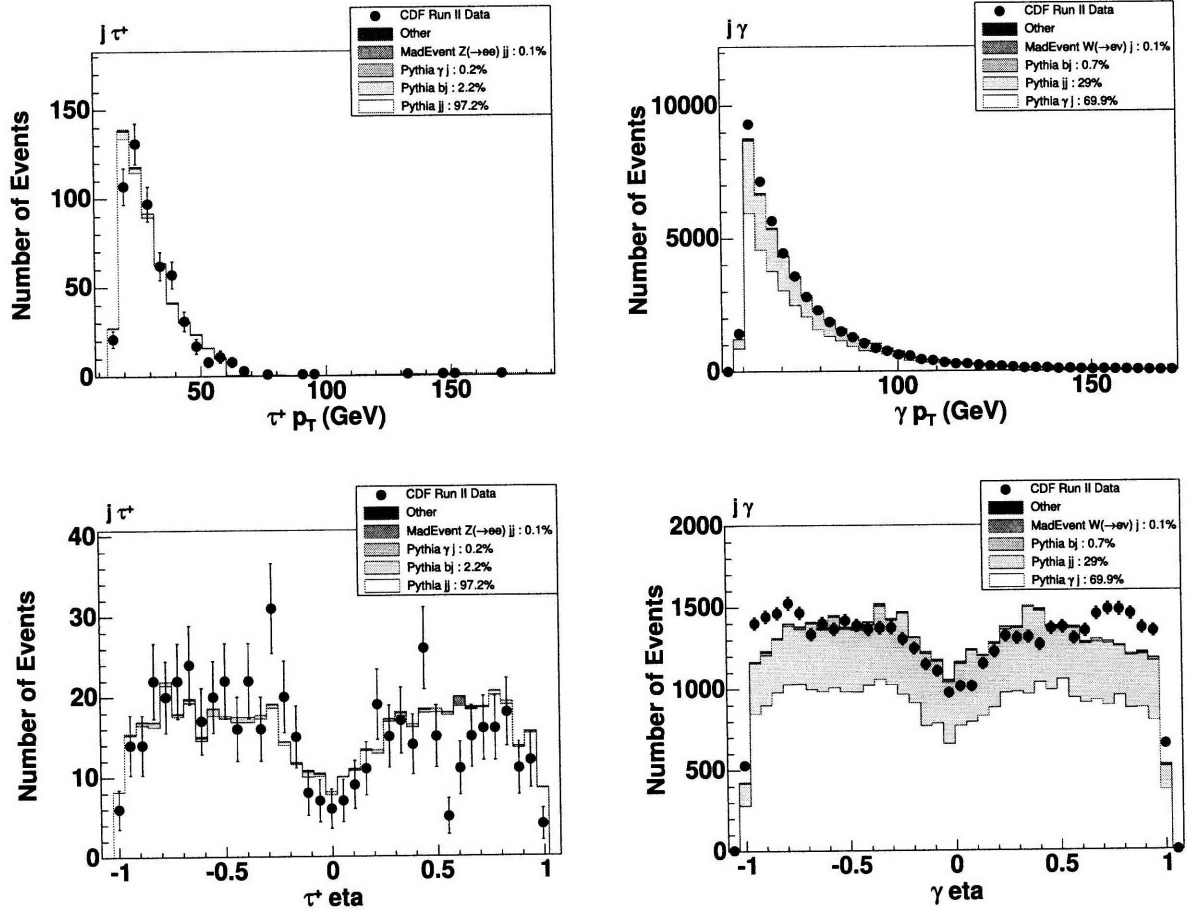
Figure A-3: A few of the most discrepant distributions in the final states $j\tau$ and $j\gamma$, which are greatly affected by the fake rates $p(j \to \tau)$ and $p(j \to \gamma)$, respectively. The distributions in the $j\gamma$ final state are among the 13 significantly discrepant distributions identified as resulting from coarseness of the correction model employed.

The probability for a light quark jet to be misreconstructed as an e^+ can be written

$$\begin{aligned} p(j \to e^+) = \quad & p(q \to \gamma)\, p(\gamma \to e^+) + \\ & p(q \to \pi^0)\, p(\pi^0 \to e^+) + \\ & p(q \to \pi^+)\, p(\pi^+ \to e^+) + \\ & p(q \to K^+)\, p(K^+ \to e^+). \end{aligned} \tag{A.1}$$

A similar equation holds for a light quark jet faking an e^-.

The probability for a light quark jet to be misreconstructed as a μ^+ can be written

$$\begin{aligned} p(j \to \mu^+) = \quad & p(q \to \pi^+)\, p(\pi^+ \to \mu^+) + \\ & p(q \to K^+)\, p(K^+ \to \mu^+). \end{aligned} \tag{A.2}$$

Here $p(\pi \to \mu)$ denotes pion decay-in-flight, and $p(K \to \mu)$ denotes kaon decay-in-flight; other processes contribute negligibly. A similar equation holds for a light quark jet faking a μ^-.

The only non-negligible underlying physical mechanisms for a jet to fake a photon are for the parent quark or gluon to fragment into a photon or a neutral pion, carrying nearly all the energy of the parent quark or gluon. Thus

$$\begin{aligned} p(j \to \gamma) = \quad & p(q \to \pi^0)\, p(\pi^0 \to \gamma) + \\ & p(q \to \gamma)\, p(\gamma \to \gamma). \end{aligned} \tag{A.3}$$

Up and down quarks and gluons fragment nearly equally to each species of pion; hence

$$\begin{aligned} \frac{1}{3} p(q \to \pi) \ = p(q \to \pi^+) \ = p(q \to \pi^-) \\ = p(q \to \pi^0), \end{aligned} \tag{A.4}$$

where $p(q \to \pi)$ denotes fragmentation into any pion carrying nearly all of the par-

ent quark's energy. Fragmentation into each type of kaon also occurs with equal probability; hence

$$\frac{1}{4}p(q \to K) \quad = p(q \to K^+) = p(q \to K^-)$$
$$= p(q \to K^0) = p(q \to \bar{K}^0), \tag{A.5}$$

where $p(q \to K)$ denotes fragmentation into any kaon carrying nearly all of the parent quark's energy.

PYTHIA contains a parameter that sets the number of string fragmentation kaons relative to the number of fragmentation pions. The default value of this parameter, which has been tuned to LEP I data, is 0.3; for every 1 up quark and every 1 down quark, 0.3 strange quarks are produced. Strange particles are produced perturbatively in the hard interaction itself, and in perturbative radiation, at a ratio larger than 0.3:1:1. This leads to the inequality

$$0.3 \lesssim \frac{p(q \to K)}{p(q \to \pi)} < 1, \tag{A.6}$$

where $p(q \to K)$ and $p(q \to \pi)$ are as defined above.

The probability for a jet to be misreconstructed as a tau lepton can be written

$$p(j \to \tau^+) = p(j \to \tau_1^+) + p(j \to \tau_3^+), \tag{A.7}$$

where $p(j \to \tau_1^+)$ denotes the probability for a jet to fake a 1-prong tau, and $p(j \to \tau_3^+)$ denotes the probability for a jet to fake a 3-prong tau. For 1-prong taus,

$$p(j \to \tau_1^+) = \quad p(q \to \pi^+)\,p(\pi^+ \to \tau^+) +$$
$$p(q \to K^+)\,p(K^+ \to \tau^+). \tag{A.8}$$

Similar equations hold for negatively charged taus.

Figure A-4 shows the probability for a quark (or gluon) to fake a one-prong tau, as a function of transverse momentum. Using fragmentation functions tuned on

162
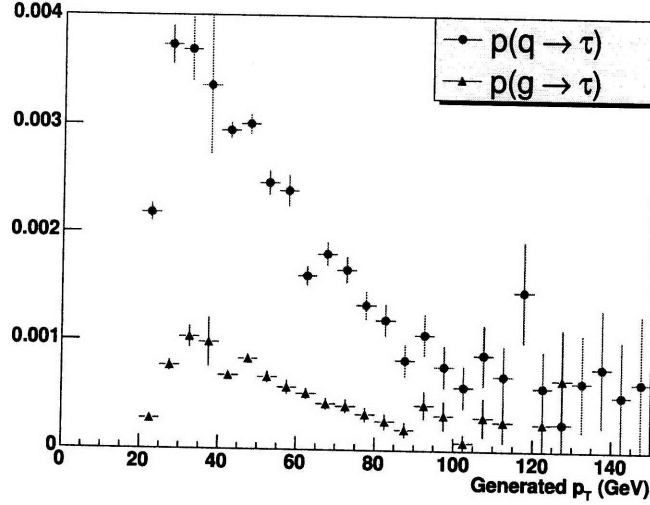
Figure A-4: The probability for a generated parton to be misreconstructed as a one-prong τ, as a function of the parton's generated p_T. Red circles show the probability for a jet arising from a parent quark to be misreconstructed as a one-prong tau. Blue triangles show the probability for a jet arising from a parent gluon to be misreconstructed as a one-prong tau.

LEP 1 data, PYTHIA predicts the probability for a quark jet to fake a one-prong tau to be roughly four times the probability for a gluon jet to fake a one-prong tau. This difference in fragmentation is incorporated into VISTA's treatment of jets faking electrons, muons, taus, and photons. The VISTA correction model includes such correction factors as the probability for a jet with a parent quark to fake an electron (0033 and 0034) and the probability for a jet with a parent quark to fake a muon (0035); the probability for a jet with a parent gluon to fake an electron or muon is then obtained by dividing the values of these fitted correction factors by four.

This effect is investigated using fake one-prong taus reconstructed in PYTHIA dijet samples.

Figure A-5 shows that the reconstructed fake tau has about $75 \pm 18\%$ of the p_T of the prominent generated particle, defined to be the generated particle carrying the greatest p_T and being within a cone of $\Delta R < 0.4$ centered on the reconstructed tau. The p_T of the misreconstructed tau is on average more undermeasured if the generated parton is a gluon than if it is a quark. This reduction in the p_T of the fake tau is implemented in VISTA when a jet is made to fake a τ during the misreconstruction
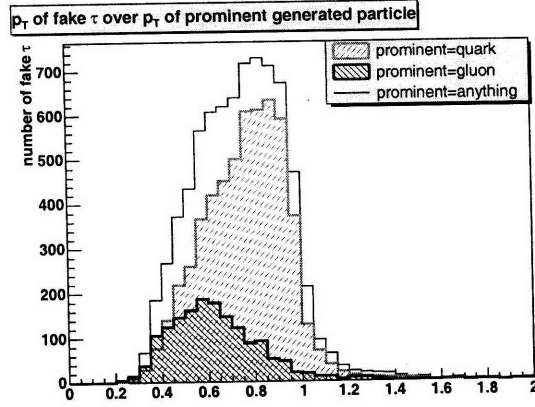
Figure A-5: Distribution of the p_T of the fake τ over the p_T of the prominent generated particle (pgp), which is defined as the generated particle within $\Delta R < 0.4$ from the reconstructed τ with the greatest p_T. The pgp is almost always a quark or a gluon, and more likely to be a quark by a factor of four.

process.

Figure A-6 shows the remaining generated p_T to be carried by neutral particles: mostly π^0's, followed by K_L^0's and η's decaying to photons or to three neutral pions. The p_T of the fake tau is determined by the track and reconstructed π^0's.

The physical mechanism underlying the process whereby an incident photon or neutral pion is misreconstructed as an electron is a conversion in the material serving as the support structure of the silicon vertex detector. This process produces exactly as many e^+ as e^-, leading to

$$\frac{1}{2} p(\gamma \to e) = p(\gamma \to e^+) = p(\gamma \to e^-)$$
$$\frac{1}{2} p(\pi^0 \to e) = p(\pi^0 \to e^+) = p(\pi^0 \to e^-), \tag{A.9}$$

where e is an electron or positron.

From Fig. A.1, the average p_T of electrons reconstructed from 25 GeV incident photons is 23.9 ± 1.4 GeV. The average p_T of electrons reconstructed from incident 25 GeV neutral pions is 23.7 ± 1.3 GeV.

The charge asymmetry between $p(K^+ \to e^+)$ and $p(K^- \to e^-)$ in Table A.1 arises because K^- can capture on a nucleon, producing a hyperon ($\Sigma^\pm$), which K^+ does not produce, due to baryon number and strangeness conservation. Among the products
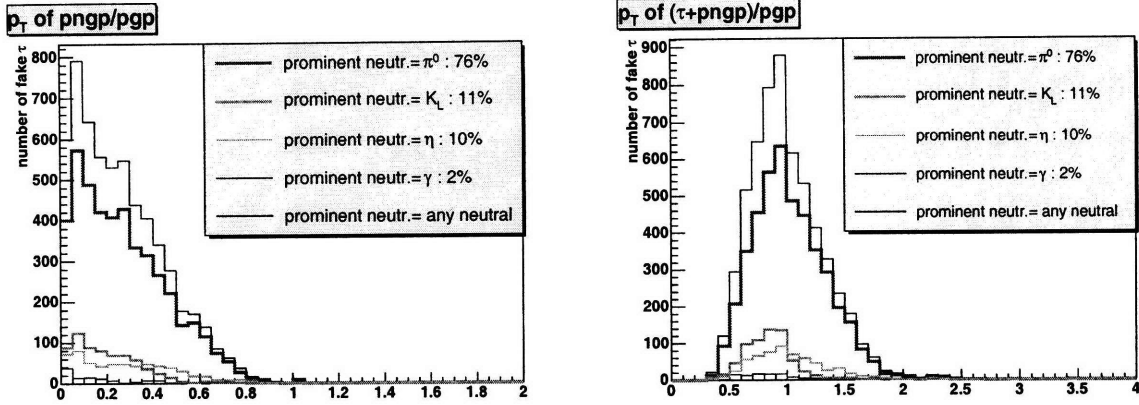
Figure A-6: Upper: The distribution of the p_T of the prominent neutral generated particle (pngp), which is the neutral generated particle with the greatest p_T within a cone of $\Delta R < 0.4$ from the fake one-prong τ, divided by the p_T of the prominent generated particle (pgp), which happens to be either a quark or a gluon. Lower: p_T of the pngp plus the p_T of the reconstructed τ, divided by the p_T of the pgp. The fact that this distribution peaks around 1 shows that the generated p_T that is missing from the fake τ was carried by the pngp. Most of the times the pngp is a π^0.

of the hyperon decay are neutral pions, which decay electromagnetically and deposit in the electromagnetic calorimeter the energy needed to have a fake e^-. The absense of this process in $K^+ + N$ interaction reduces the $pooK^+ e^+$ relative $pK^- e^-$ by roughly a factor of two.

The physical process primarily responsible for $\pi^\pm \to e^\pm$ is inelastic charge exchange

$$\pi^- p \to \pi^0 n$$
$$\pi^+ n \to \pi^0 p \qquad (A.10)$$

occurring within the electromagnetic calorimeter. The charged pion leaves the "electron's" track in the CDF tracking chamber, and the π^0 produces the "electron's" electromagnetic shower. No true electron appears at all in this process, except as secondaries in the electromagnetic shower originating from the π^0.

The average p_T of reconstructed "electrons" originating from a single charged pion is 18.8 ± 2.2 GeV, indicating that the misreconstructed "electron" in this case is measured to have on average only 75% of the total energy of the parent quark or

165

gluon. This is expected, since the recoiling nucleon from the charge exchange process carries some of the incident pion's momentum.

An additional small loss in energy for a jet misreconstructed as an electron, photon, or muon is expected since the leading π^+, K^+, π^0, or γ takes only some fraction of the parent quark's energy.

The cross sections for $\pi^- p \to \pi^0 n$ and $\pi^+ n \to \pi^0 p$, proceeding through the isospin I conserving and I_3 independent strong interaction, are roughly equal. The corresponding particles in the two reactions are related by interchanging the signs of their z-components of isospin.

The probability for a 25 GeV π^+ to decay to a μ^+ can be written

$$p(\pi^+ \to \mu^+) = \quad p(\text{decays within tracker}) +$$
$$p(\text{decays within calorimeter}). \tag{A.11}$$

The probability for the pion to decay within the tracking volume is

$$p(\text{decays within tracker}) = 1 - e^{-R_\text{tracker}/\gamma(c\tau)}, \tag{A.12}$$

where $\gamma = 25$ GeV / 140 MeV = 180 is the pion's Lorentz boost, the proper decay length of the charged pion is $(c\tau) = 7.8$ meters, and the radius of the CDF tracking volume is $R_\text{tracker} = 1.5$ meters, giving $p(\text{decays within tracker}) = 0.001$. The probability for the pion to decay within the calorimeter volume is

$$p(\text{decays within calorimeter}) \approx \lambda_I/\gamma(c\tau), \tag{A.13}$$

where $\lambda_I \approx 0.4$ meters is the nuclear interaction length for charged pions on lead or iron and the path length through the calorimeter is $L_\text{cal} \approx 2$ meters, leading to $p(\text{decays within calorimeter}) \approx 0.00025$. Summing the contributions from decay within the tracking volume and decay within the calorimeter volume, $p(\pi^+ \to \mu^+) \approx 0.00125$.

The primary physical mechanism by which a jet fakes a photon is for the par-

ent quark or gluon to fragment into a leading π^0 carrying nearly all the momentum. The highly boosted π^0 decays within the beam pipe to two photons that are sufficiently collinear to appear in the preshower, electromagnetic calorimeter, and shower maximum detector as a single photon. Thus

$$p(j \to \gamma) = p(q \to \pi^0)\,p(\pi^0 \to \gamma). \tag{A.14}$$

An immediate corollary is that the misreconstructed "photon" carries the energy of the parent quark or gluon, and is well measured.

Since $p(q \to \pi^0) \gg p(q \to \gamma)$, it follows from Eq. A.4 and Table A.1 that the conversion contribution to $p(j \to e)$ is $\approx 75\%$, and the charge exchange contribution is $\approx 25\%$:

$$
\begin{aligned}
\frac{0.75}{0.25} = (\quad & p(q \to \gamma)\,p(\gamma \to e^+)+ \\
& p(q \to \pi^0)\,p(\pi^0 \to e^+) \quad)\,/ \\
(\quad & p(q \to \pi^+)\,p(\pi^+ \to e^+)+ \\
& p(q \to K^+)\,p(K^+ \to e^+) \quad).
\end{aligned}
\tag{A.15}
$$

The number of $e^+ j$ events in data is 0.9 times the number of $e^- j$ events. This charge asymmetry arises from $p(K^+ \to e^+)$ and $p(K^- \to e^-)$ in Table A.1. Quantitatively,

$$\frac{p(j \to e^+)}{p(j \to e^-)} = \frac{0.9 + 0.2\,p(K^+ \to e^+)/p(K \to e)}{0.9 + 0.2\,p(K^- \to e^-)/p(K \to e)}, \tag{A.16}$$

where 0.9 is the sum of 0.75 from Eq. A.15 and $0.15 \approx 0.25 \times 0.6$ from Eq. A.6, and 0.2 is twice $1 - 0.9$. From $p(K^+ \to e^+)$ and $p(K^- \to e^-)$ in Table A.1, $p(K^+ \to e^+)/p(K \to e) = 1/3$ and $p(K^- \to e^-)/p(K \to e) = 2/3$, predicting $p(j \to e^+)/p(j \to e^-) = 0.935$, in reasonable agreement with the ratio of the observed number of events in the $e^+ j$ and $e^- j$ final states.

167

The number of $j\,\mu^+$ events observed in CDF Run II is 1.1 times the number of $j\,\mu^-$ events observed. This charge asymmetry arises from $p(K^+ \to \mu^+)$ and $p(K^- \to \mu^-)$ in Table A.1.

The physical mechanism by which a prompt photon fakes a tau lepton is for the photon to convert, producing an electron or positron carrying most of the photon's energy, which is then misreconstructed as a tau. The probability for this to occur is equal for positively and negatively charged taus,

$$\frac{1}{2}p(\gamma \to \tau) = p(\gamma \to \tau^+) = p(\gamma \to \tau^-), \tag{A.17}$$

and is related to previously defined quantities by

$$p(\gamma \to \tau) = p(\gamma \to e)\,\frac{1}{p(e \to e)}\,p(e \to \tau), \tag{A.18}$$

where $p(\gamma \to e)$ denotes the fraction of produced photons that are reconstructed as electrons, $p(e \to e)$ denotes the fraction of produced electrons that are reconstructed as electrons, and hence $p(\gamma \to e)/p(e \to e)$ is the fraction of produced photons that pair produce a single leading electron.

Note $p(e \to \gamma) \approx p(\gamma \to e)$ from Table A.1, as expected, with value of ≈ 0.03 determined by the amount of material in the inner detectors and the tightness of isolation criteria. A hard bremsstrahlung followed by a conversion is responsible for electrons to be reconstructed with opposite sign; hence

$$\begin{aligned} p(e^\pm \to e^\mp) =&\ p(e^+ \to e^-) = p(e^- \to e^+) \\ \approx&\ \tfrac{1}{2}p(e^\pm \to \gamma)p(\gamma \to e^\mp), \end{aligned} \tag{A.19}$$

where the factor of 1/2 comes because the material already traversed by the $e^\pm$ will not be traversed again by the γ. In particular, track curvature mismeasurement is not responsible for erroneous sign determination in the central region of the CDF detector.

From knowledge of the underlying physical mechanisms by which jets fake elec-

trons, muons, taus, and photons, the simple use of a reconstructed jet as a lepton or photon with an appropriate fake rate applied to the weight of the event needs slight modification to correctly handle the fact that a jet that has faked a lepton or photon generally is measured more accurately than a hadronic jet. Rather than using the momentum of the reconstructed jet, the momentum of the parent quark or gluon is determined by adding up all Monte Carlo particle level objects within a cone of $\Delta R = 0.4$ about the reconstructed jet. In misreconstructing a jet in an event, the momentum of the corresponding parent quark or gluon is used rather than the momentum of the reconstructed jet. A jet that fakes a photon then has momentum equal to the momentum of the parent quark or gluon plus a fractional correction equal to $0.01 \times$ (parent $p_T - 25$ GeV)/(25 GeV) to account for leakage out of the cone of $\Delta R = 0.4$, and a further smearing of $0.2 \sqrt{\text{GeV}} \times \sqrt{\text{parent} \, p_T}$, reflecting the electromagnetic resolution of the CDF detector. The momenta of jets that fake photons are multiplied by an overall factor of 1.12, and jets that fake electrons, muons, or taus are multiplied by an overall factor of 0.95. These numbers are determined by the $\ell \not{p}$, ℓj, and γj final states. The distributions most sensitive to these numbers are the missing energy and the jet p_T.

A b quark fragmenting into a leading b hadron that then decays leptonically or semileptonically results in an electron or muon that shares the p_T of the parent b quark with the associated neutrino. If all hadronic decay products are soft, the distribution of the momentum fraction carried by the charged lepton can be obtained by considering the decay of a scalar to two massless fermions. Isolated and energetic electrons and muons arising from parent b quarks in this way are modeled as having p_T equal to the parent b quark p_T, multiplied by a random number uniformly distributed between 0 and 1.

A.2   Additional background sources

This appendix provides additional details on the estimation of the Standard Model prediction.

A.2.1 Cosmic ray and beam halo muons

There are four dominant categories of events caused by cosmic ray muons penetrating the detector: $\mu\slashed{p}$, $\mu^+\mu^-$, $\gamma\slashed{p}$, and $j\slashed{p}$. There is negligible contribution from cosmic ray secondaries of any particle type other than muons.

A cosmic ray muon penetrating the CDF detector whose trajectory passes within 1 mm of the beam line and within $-60 < z < 60$ cm of the origin may be reconstructed as two outgoing muons. In this case the cosmic ray event is partitioned into the final state $\mu^+\mu^-$. If one of the tracks is missed, the cosmic ray event is partitioned into the final state $\mu\slashed{p}$. The standard CDF cosmic ray filter, which makes use of drift time information in the central tracking chamber, is used to reduce these two categories of cosmic ray events.

CDF data events with exactly one track (corresponding to one muon) and events with exactly two tracks (corresponding to two muons) are used to estimate the cosmic ray muon contribution to the final states $\mu\slashed{p}$ and $\mu^+\mu^-$ after the cosmic ray filter. This sample of events is used as the SM background process `cosmic` μ. The `cosmic` μ sample does not contribute to the events passing the analysis offline trigger, whose cleanup cuts require the presence of three or more tracks.

The remaining two categories are $\gamma\slashed{p}$ and $j\slashed{p}$, resulting from a cosmic ray muon that penetrates the CDF electromagnetic or hadronic calorimeter and undergoes a hard bremsstrahlung in one calorimeter cell. Such an interaction can mimic a single photon or a single jet, respectively. The reconstruction algorithm infers the presence of significant missing energy balancing the "photon" or "jet." If this cosmic ray interaction occurs during a bunch crossing in which there is a $p\bar{p}$ interaction producing three or more tracks, the event will be partitioned into the final state $\gamma\slashed{p}$ or $j\slashed{p}$.

CDF data events with fewer than three tracks are used to estimate the cosmic ray muon contribution to the final states $\gamma\slashed{p}$ and $j\slashed{p}$. These samples of events are used as SM background processes `cosmic` γ and `cosmic` j for the modeling of this background, corresponding to offline triggers requiring a photon with $p_T > 60$ GeV, or a jet with $p_T > 40$ GeV (prescaled) or $p_T > 200$ GeV (unprescaled), respectively.
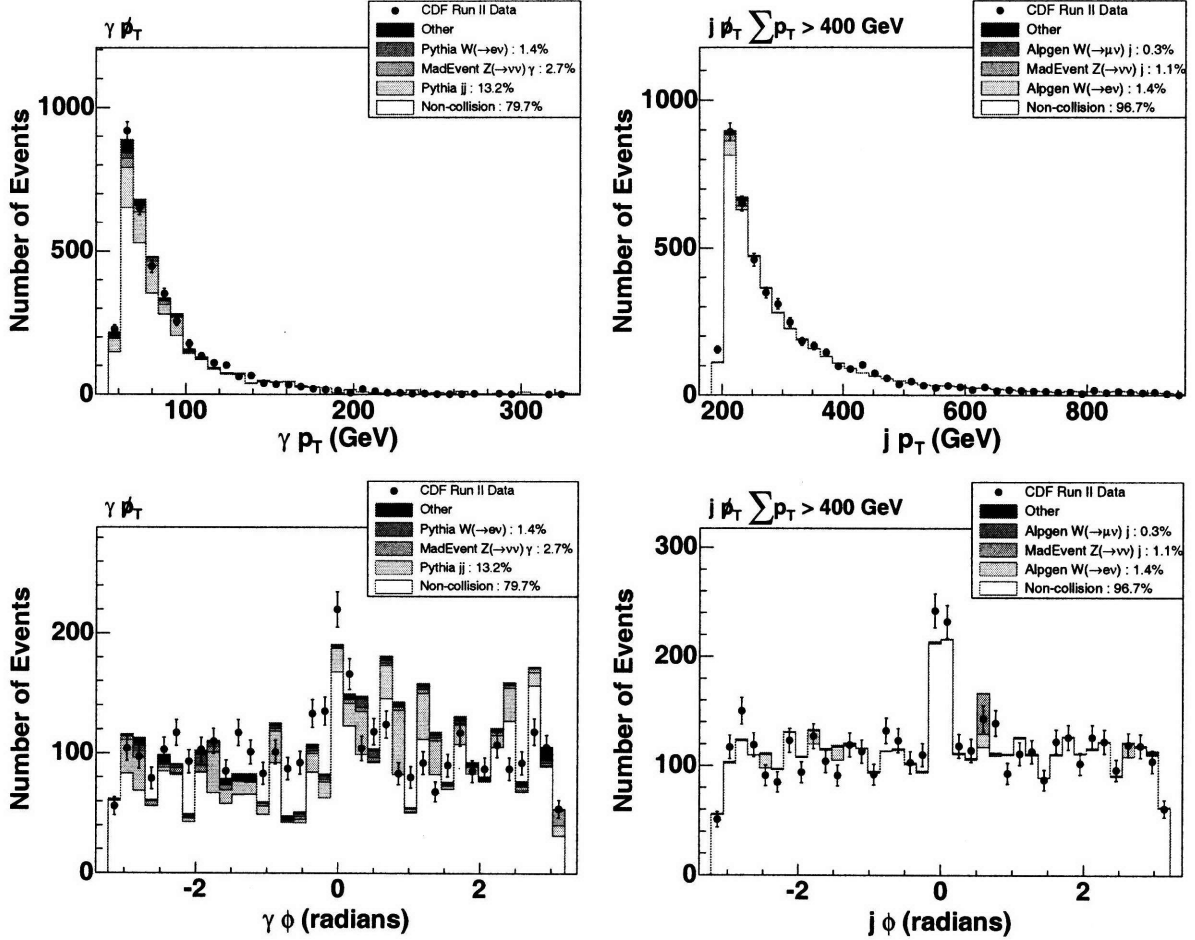
Figure A-7: The distribution of transverse momentum and azimuthal angle for photons and jets in the $\gamma\displaystyle{\not}p$ and $j\displaystyle{\not}p$ final states, dominated by cosmic ray and beam halo muons. The vertical axis shows the number of events in each bin. Data are shown as filled (black) circles; the SM prediction is shown as the shaded (red) histogram. The prediction includes contributions from cosmic ray and beam halo muons, estimated using events containing fewer than three reconstructed tracks. The contribution from cosmic ray muons is flat in ϕ, while the contribution from beam halo is localized to $\phi = 0$. The only degrees of freedom for the background to these final states are the cosmic γ and cosmic j correction factors, whose values are determined from the global fit (Table 4.2).

These samples do not contribute to the events passing the analysis offline trigger, whose cleanup cuts require three or more tracks. The contribution of these events is adjusted with correction factors that are listed as `cosmic` γ and `cosmic` j "k-factors" in Table 4.2, but which are more properly understood as reflecting the number of bunch crossings with zero $p\bar{p}$ interactions (resulting in zero reconstructed tracks) relative to the number of bunch crossings with one or more interactions (resulting in three or more reconstructed tracks).

The cosmic ray muon contribution to the final states $\gamma\not{p}$ and $j\not{p}$ is uniform as a function of the CDF azimuthal angle ϕ. Consider the CDF detector to be a thick cylindrical shell, and consider two arbitrary infinitesimal volume elements at different locations in the material of the shell. Since the two volume elements have similar overburdens, the number of cosmic ray muons with $E \gtrsim 20$ GeV penetrating the first volume element is very nearly the same as the number of cosmic ray muons with $E \gtrsim 20$ GeV penetrating the second volume element. Since the material of the CDF calorimeters is uniform as a function of CDF azimuthal angle ϕ, it follows that the cosmic ray muon contribution to the final states $\gamma\not{p}$ and $j\not{p}$ should also be uniform as a function of ϕ. In particular, it is noted that the ϕ dependence of this contribution depends solely on the material distribution of CDF calorimeter, which is uniform in ϕ, and has no dependence on the distribution of the horizon angle of the muons from cosmic rays.

The final states $\gamma\not{p}$ and $j\not{p}$ are also populated by beam halo muons, traveling horizontally through the CDF detector in time with a bunch. A beam halo muon can undergo a hard bremsstrahlung in the electromagnetic or hadronic calorimeters, producing an energy deposition that can be reconstructed as a photon or jet, respectively. These beam halo muons tend to lie in the horizontal plane and outside of the Tevatron ring, as if centrifugally hurled away from the beam; they horizontally penetrate the CDF detector along $\hat{z}$ at $y = 0$ and $x > 0$, hence at $\phi = 0$.

Fig. A-7 shows the $\gamma\not{p}$ and $j\not{p}$ final states, in which events come primarily from cosmic ray and beam halo muons.

A.2.2 Multiple interactions

In order to estimate event overlaps, consider an interesting event observed in final state C, which looks like an overlap of two events in the final states A and B. An example is C=e+e-4j, A=e+e- and B=4j. It is desired to estimate how many C events are expected from the overlap of A and B events, given the observed frequencies of A and B.

Let $\mathcal{L}(t)$ be the instantaneous luminosity as a function of time t; let

$$L = \int_{\text{RunII}} \mathcal{L}(t)dt = 1993 \text{ pb}^{-1} \tag{A.20}$$

denote the total integrated luminosity; and let

$$\bar{\mathcal{L}} = \frac{\int_{\text{RunII}} \mathcal{L}(t)\mathcal{L}(t)dt}{\int_{\text{RunII}} \mathcal{L}(t)dt} \approx 10^{32} \text{ cm}^{-2}\text{s}^{-1} \tag{A.21}$$

be the luminosity-averaged instantaneous luminosity. Denote by t_0 the time interval of 396 ns between successive bunch crossings. The total number of effective bunch crossings X is then

$$X = \frac{L}{\bar{\mathcal{L}}t_0} \approx 5 \times 10^{13}. \tag{A.22}$$

Letting A and B denote the number of observed events in final states A and B, it follows that the number of events in the final state C expected from overlap of A and B is

$$C = \frac{A}{X}\frac{B}{X}X = \frac{AB}{X}. \tag{A.23}$$

Overlap events are included in the SM background estimate, although their contribution is generally negligible.

A.2.3 Intrinsic k_T

Significant discrepancy is observed in many final states containing two objects o1 and o2 in the variables $\Delta\phi$(o1,o2), uncl p_T, and $\not{p}_T$. These discrepancies are ascribed to the sum of two effects: (1) an intrinsic Fermi motion of the colliding partons within

the proton and anti-proton, and (2) soft radiation along the beam axis. The sum of these two effects appears to be larger in Nature than predicted by PYTHIA with the parameter tunes used for the generation of the samples employed in this analysis. This discrepancy is well known from previous studies at the Tevatron and elsewhere, and affects this analysis similarly to other Tevatron analyses.

The W and Z electroweak samples used in this analysis have been generated with an adjusted PYTHIA parameter that increases the intrinsic k_T. For all other generated Standard Model events, the net effect of the Fermi motion of the colliding partons and the soft non-perturbative radiation is hypothesized to be described by an overall "effective intrinsic k_T," and the center of mass of each event is given a transverse kick. Specifically, for every event of invariant mass m and generated summed transverse momentum $\sum p_T$, a random number k_T is pulled from the probability distribution

$$p(k_T) \propto (k_T < m/5) \times \quad [\tfrac{4}{5}g(k_T; \mu = 0, \sigma_1) +$$
$$\tfrac{1}{5}g(k_T; \mu = 0, \sigma_2)], \qquad (A.24)$$

where $(k_T < m/5)$ evaluates to unity if true and zero if false; $g(k_T; \mu, \sigma)$ is a Gaussian function of k_T with center at μ and width σ; $\sigma_1 = 2.55\,\mathrm{GeV} + 0.0085 \sum p_T$ is the width of the core of the double Gaussian; and $\sigma_2 = 5.25\,\mathrm{GeV} + 0.0175 \sum p_T$ is the width of the second, wider Gaussian. The event is then boosted to an inertial frame traveling with speed $\left|\vec{\beta}\right| = k_T/m$ with respect to the lab frame, in a direction transverse to the beam axis, where m is the invariant mass of all reconstructed objects in the event, along an azimuthal angle pulled randomly from a uniform distribution between 0 and 2π. The momenta of identified objects are recalculated in the lab frame. Sixty percent of the recoil kick is assigned to unclustered momentum in the event. The remaining forty percent of the recoil kick is assumed to disappear down the beam pipe, and contributes to the missing transverse momentum in the event. This picture, and the particular parameter values that accompany this story, are determined primarily by the uncl p_T and $\not{p}_T$ distributions in highly populated two-object final states, including the low-p_T $2j$ final state, the high-p_T $2j$ final state, and the final states $j\gamma$, e^+e^-, and

$\mu^+\mu^-$.

Under the hypothesis described, reasonable although imperfect agreement with observation is obtained. The result of this analysis supports the conclusions of previous studies indicating that the effective intrinsic k_T needed to match observation is quite large relative to naive expectation. That the data appear to require such a large effective intrinsic k_T may be pointing out the need for some basic improvement to our understanding of this physics.

A.3 Global fit

This section describes the construction of the global χ^2 used in the VISTA global fit.

A.3.1 The χ_k^2

The bins in the CDF high-p_T data sample are labeled by the index $k = (k_1, k_2)$, where each value of k_1 represents a phrase such as "this bin contains events with three objects: one with $17 < p_T < 25$ GeV and $|\eta| < 0.6$, one with $40 < p_T < 60$ GeV and $0.6 < |\eta| < 1.0$, and one with $25 < p_T < 40$ GeV and $1.0 < |\eta|$," and each value of k_2 represents a phrase such as "this bin contains events with three objects: an electron, muon, and jet, respectively." The reason for splitting k into k_1 and k_2 is that a jet can fake an electron (mixing the contents of k_2), but an object with $|\eta| < 0.6$ cannot fake an object with $0.6 < |\eta| < 1.0$ (no mixing of k_1). The term corresponding to the kth bin takes the form of Eq. 3.1, where Data[k] is the number of data events observed in the kth bin, SM[k] is the number of events predicted by the Standard Model in the kth bin, δSM[k] is the Monte Carlo statistical uncertainty on the Standard Model prediction in the kth bin, and $\sqrt{\text{SM}[k]}$ is the statistical uncertainty on the prediction in the kth bin. To legitimize the use of Gaussian errors, only bins containing eight or more data events are considered. The Standard Model prediction SM[k] for the kth

bin can be written in terms of the introduced correction factors as

$$\mathrm{SM}[k] = \mathrm{SM}[(k_1, k_2)] =$$

$$\sum_{k_2' \in \text{objectLists}} \sum_{l \in \text{processes}}$$

$$(\textstyle\int \mathcal{L} \, dt) \cdot (\text{kFactor}[l]) \cdot (\mathrm{SM}_0[(k_1, k_2')][l]) \cdot$$

$$(\text{probabilityToBeSoMisreconstructed}[(k_1, k_2')][k_2]) \cdot$$

$$(\text{probabilityPassesTrigger}[(k_1, k_2)]), \qquad\qquad (\text{A.25})$$

where $\mathrm{SM}[k]$ is the Standard Model prediction for the kth bin; the index k is the Cartesian product of the two indices k_1 and k_2 introduced above, labeling the regions of the detector in which there are energy clusters and the identified objects corresponding to those clusters, respectively; the index k_2' is a dummy summation index; the index l labels Standard Model background processes, such as dijet production or W+1 jet production; $\mathrm{SM}_0[(k_1, k_2')][l]$ is the initial number of Standard Model events predicted in bin (k_1, k_2') from the process labeled by the index l; probabilityToBeMisreconstructedThus$[(k_1, k_2')][k_2]$ is the probability that an event produced with energy clusters in the detector regions labeled by k_1 that are identified as objects labeled by k_2' would be mistaken as having objects labeled by k_2; and probabilityPassesTrigger$[(k_1, k_2)]$ represents the probability that an event produced with energy clusters in the detector regions labeled by k_1 that are identified as objects labeled by k_2 would pass the trigger.

The quantity $\mathrm{SM}_0[(k_1, k_2')][l]$ is obtained by generating some number n_l (say 10^4) of Monte Carlo events corresponding to the process l. The event generator provides a cross section σ_l for this process l. The weight of each of these Monte Carlo events is equal to σ_l/n_l. Passing these events through the CDF simulation and reconstruction, the sum of the weights of these events falling into the bin (k_1, k_2') is $\mathrm{SM}_0[(k_1, k_2')][l]$.

A.3.2 $\chi^2_{\text{constraints}}$

The term $\chi^2_{\text{constraints}}(\vec{s})$ in Eq. 3.2 reflects constraints on the values of the correction factors determined by data other than those in the global high-p_T sample. These constraints include k-factors taken from theoretical calculations and numbers from the CDF literature when use is made of CDF data external to the VISTA high-p_T sample. The constraints imposed are:

- The luminosity (0001) is constrained to be within 6% of the value measured by the CDF Čerenkov luminosity counters.

- The fake rate $p(q \to \gamma)$ (0039) is constrained to be $2.6 \times 10^{-4} \pm 1.5 \times 10^{-5}$, from the single particle gun study of Appendix A.1.

- The fake rate $p(e \to \gamma)$ (0032) plus the efficiency $p(e \to e)$ (0026) for electrons in the plug is constrained to be within 1% of unity.

- Noting $p(q \to \gamma)$ corresponds to correction factor 0039, $p(q \to \pi^{\pm}) = 2\,p(q \to \pi^0)$, and $p(q \to \pi^0) = p(q \to \gamma)/p(\pi^0 \to \gamma)$, and taking $p(\pi^0 \to \gamma) = 0.6$ and $p(\pi^{\pm} \to \tau) = 0.415$ from the single particle gun study of Appendix A.1, the fake rate $p(q \to \tau)$ (0038) is constrained to $p(q \to \tau) = p(q \to \pi^{\pm})p(\pi^{\pm} \to \tau) \pm 10\%$.

- The k-factors for dijet production (0018 and 0019) are constrained to 1.10 ± 0.05 and 1.33 ± 0.05 in the kinematic regions $\hat{p}_T < 150$ GeV and $\hat{p}_T > 150$ GeV, respectively, where $\hat{p}_T$ is the transverse momentum of the scattered partons in the $2 \to 2$ process in the colliding parton center of momentum frame.

- The inclusive k-factor for $\gamma+N$jets (0004–0007) is constrained to 1.25 ± 0.15 [85, 86].

- The inclusive k-factor for $\gamma\gamma+N$jets (0008–0010) is constrained to 2.0 ± 0.15 [87].

- The inclusive k-factors for W and Z production (0011–0014 and 0015–0017) are subject to a 2-dimensional Gaussian constraint, with mean at the NNLO/LO theoretical values [88], and a covariance matrix that encapsulates the highly correlated theoretical uncertainties, as discussed in Appendix A.4.

177

- Trigger efficiency correction factors are constrained to be less than unity.

- All correction factors are constrained to be positive.

A.3.3 Covariance matrix

This section describes the correction factor covariance matrix Σ. The inverse of the covariance matrix is obtained from

$$\Sigma_{ij}^{-1} = \frac{1}{2} \frac{\partial^2 \chi^2(\vec{s})}{\partial s_i \partial s_j}\bigg|_{\vec{s}_0}, \tag{A.26}$$

where $\chi^2(\vec{s})$ is defined by Eq. 3.2 as a function of the correction factor vector $\vec{s}$, vector elements s_i and s_j are the ith and j^{th} correction factors, and $\vec{s}_0$ is the vector of correction factors that minimizes $\chi^2(\vec{s})$. Numerical estimation of the right hand side of Eq. A.26 is achieved by calculating χ^2 at $\vec{s}_0$ and at positions slightly displaced from $\vec{s}_0$ in the direction of the ith and j^{th} correction factors, denoted by the unit vectors $\hat{i}$ and $\hat{j}$. Approximating the second partial derivative

$$\frac{\partial^2 \chi^2}{\partial s_j \partial s_i}\bigg|_{\vec{s}_0} = \frac{\chi^2(\vec{s}_0 + \hat{i}\,\delta s_i + \hat{j}\,\delta s_j) - \chi^2(\vec{s}_0 + \hat{j}\,\delta s_j)}{\delta s_j \delta s_i} - \frac{\chi^2(\vec{s}_0 + \hat{i}\,\delta s_i) - \chi^2(\vec{s}_0)}{\delta s_j \delta s_i}$$

leads to

$$\begin{aligned}
\Sigma_{ij}^{-1} = {}& [\chi^2(\vec{s}_0 + \delta s_i\,\hat{i} + \delta s_j\,\hat{j}) \\
& - \chi^2(\vec{s}_0 + \delta s_i\,\hat{i}) \\
& - \chi^2(\vec{s}_0 + \delta s_j\,\hat{j}) \\
& + \chi^2(\vec{s}_0)]/(2\delta s_i\,\delta s_j), \tag{A.27}
\end{aligned}$$

for appropriately small steps δs_i and δs_j away from the minimum. The covariance matrix Σ is calculated by inverting Σ^{-1}. The diagonal element Σ_{ii} is the variance σ_i^2 of the ith correction factor, and the correlation ρ_{ij} between the ith and j^{th} correction

	0001	0002	0003	0004	0005	0006	0007	0008	0009	0010	0011	0012	0013	0014	0015	0016	0017	0018	0019	0020	0021	0022	0023	0024	0025	0026	0027	0028	0029	0030	0031
0001	1	-.32	-.7	-.56	-.53	-.45	-.26	-.36	-.21	-.14	-.87	-.77	-.51	-.28	-.82	-.55	-.31	-.95	-.96	-.94	-.94	-.88	-.88	-.62	-.54	-.17	-.46	-.37	-.09	-.1	0
0002	-.32	1	+.21	+.37	+.38	+.33	+.2	+.34	+.18	+.12	+.28	+.25	+.17	+.09	+.27	+.18	+.1	+.3	+.31	+.31	+.3	+.28	+.28	+.2	+.18	+.06	+.15	+.12	-.31	+.02	0
0003	-.7	+.21	1	+.39	+.37	+.31	+.18	+.25	+.14	+.1	+.61	+.53	+.35	+.2	+.57	+.38	+.21	+.66	+.66	+.66	+.65	+.61	+.61	+.43	+.38	+.12	+.32	+.26	+.06	+.07	-.01
0004	-.56	+.37	+.39	1	+.9	+.77	+.48	+.61	+.33	+.23	+.49	+.43	+.29	+.16	+.46	+.32	+.18	+.5	+.53	+.53	+.52	+.49	+.49	+.35	+.3	+.1	+.25	+.2	-.46	+.03	0
0005	-.53	+.38	+.37	+.9	1	+.75	+.46	+.62	+.31	+.21	+.46	+.41	+.27	+.15	+.44	+.3	+.17	+.5	+.51	+.48	+.5	+.47	+.46	+.33	+.29	+.1	+.24	+.19	-.49	+.03	-.02
0006	-.45	+.33	+.31	+.77	+.75	1	+.4	+.54	+.29	+.13	+.4	+.35	+.24	+.13	+.38	+.26	+.14	+.43	+.44	+.42	+.42	+.35	+.4	+.28	+.25	+.09	+.2	+.16	-.45	+.02	-.01
0007	-.26	+.2	+.18	+.48	+.46	+.4	1	+.34	+.18	+.09	+.23	+.2	+.13	+.08	+.22	+.15	+.08	+.24	+.25	+.25	+.24	+.21	+.22	+.02	+.14	+.05	+.12	+.09	-.29	+.01	-.02
0008	-.36	+.34	+.25	+.61	+.62	+.54	+.34	1	+.37	+.28	+.32	+.28	+.19	+.1	+.3	+.22	+.12	+.34	+.34	+.34	+.33	+.31	+.31	+.22	+.18	+.06	+.16	+.12	-.61	-.03	0
0009	-.21	+.18	+.14	+.33	+.31	+.29	+.18	+.37	1	+.06	+.19	+.17	+.11	+.06	+.2	+.06	+.11	+.2	+.2	+.19	+.19	+.18	+.18	+.13	+.08	+.03	+.07	+.06	-.31	+.05	0
0010	-.14	+.12	+.1	+.23	+.21	+.13	+.09	+.28	+.06	1	+.13	+.11	+.08	+.06	+.13	+.11	-.03	+.13	+.14	+.13	+.13	+.12	+.12	+.09	+.05	-.01	+.05	+.04	-.19	+.06	0
0011	-.87	+.28	+.61	+.49	+.46	+.4	+.23	+.32	+.19	+.13	1	+.85	+.58	+.32	+.89	+.61	+.33	+.83	+.84	+.82	+.82	+.76	+.77	+.54	+.25	+.09	+.16	+.15	+.07	+.07	0
0012	-.77	+.25	+.53	+.43	+.41	+.35	+.2	+.28	+.17	+.11	+.85	1	+.33	+.35	+.79	+.49	+.33	+.72	+.74	+.74	+.72	+.68	+.67	+.47	+.21	+.08	+.15	+.13	+.06	+.06	+.01
0013	-.51	+.17	+.35	+.29	+.27	+.24	+.13	+.19	+.11	+.08	+.58	+.33	1	-.21	+.52	+.35	+.15	+.5	+.49	+.46	+.48	+.46	+.45	+.36	+.15	+.06	+.1	+.09	+.04	+.04	-.01
0014	-.28	+.09	+.2	+.16	+.15	+.13	+.08	+.1	+.06	+.06	+.32	+.35	-.21	1	+.29	+.26	-.04	+.28	+.27	+.28	+.26	+.21	+.26	+.09	+.08	+.03	+.05	+.05	+.02	+.02	0
0015	-.82	+.27	+.57	+.46	+.44	+.38	+.22	+.3	+.2	+.13	+.89	+.79	+.52	+.29	1	+.58	+.35	+.77	+.78	+.77	+.76	+.71	+.71	+.5	+.09	+.04	+.06	+.05	+.05	+.02	0
0016	-.55	+.18	+.38	+.32	+.3	+.26	+.15	+.22	+.06	+.11	+.61	+.49	+.35	+.26	+.58	1	-.09	+.52	+.53	+.52	+.52	+.49	+.48	+.35	+.1	+.03	+.08	+.07	+.03	+.02	0
0017	-.31	+.1	+.21	+.18	+.17	+.14	+.08	+.12	+.11	-.03	+.33	+.33	+.15	-.04	+.35	-.09	1	+.3	+.3	+.29	+.29	+.25	+.28	+.16	+.03	-.02	+.04	+.04	+.02	+.04	0
0018	-.95	+.3	+.66	+.5	+.5	+.43	+.24	+.34	+.2	+.13	+.83	+.72	+.5	+.28	+.77	+.52	+.3	1	+.91	+.92	+.89	+.85	+.83	+.6	+.51	+.16	+.43	+.35	+.09	+.1	-.07
0019	-.96	+.31	+.66	+.53	+.51	+.44	+.25	+.34	+.2	+.14	+.84	+.74	+.49	+.27	+.78	+.53	+.3	+.91	1	+.91	+.91	+.84	+.85	+.59	+.52	+.16	+.44	+.36	+.09	+.1	+.03
0020	-.94	+.31	+.66	+.53	+.48	+.42	+.25	+.34	+.19	+.13	+.82	+.74	+.46	+.28	+.77	+.52	+.29	+.92	+.91	1	+.87	+.84	+.83	+.6	+.51	+.16	+.43	+.35	+.08	+.1	-.05
0021	-.94	+.3	+.65	+.52	+.5	+.42	+.24	+.33	+.19	+.13	+.82	+.72	+.48	+.26	+.76	+.52	+.29	+.89	+.91	+.87	1	+.82	+.83	+.57	+.51	+.16	+.43	+.35	+.08	+.1	+.04
0022	-.88	+.28	+.61	+.49	+.47	+.35	+.21	+.31	+.18	+.12	+.76	+.68	+.46	+.21	+.71	+.49	+.25	+.85	+.84	+.84	+.82	1	+.73	+.55	+.47	+.15	+.4	+.33	+.08	+.09	-.04
0023	-.88	+.28	+.61	+.49	+.46	+.4	+.22	+.31	+.18	+.12	+.77	+.67	+.45	+.26	+.71	+.48	+.28	+.83	+.85	+.83	+.83	+.73	1	+.53	+.48	+.15	+.4	+.33	+.08	+.09	+.01
0024	-.62	+.2	+.43	+.35	+.33	+.28	+.02	+.22	+.13	+.09	+.54	+.47	+.36	+.09	+.5	+.35	+.16	+.6	+.59	+.6	+.57	+.55	+.53	1	+.33	+.11	+.28	+.23	+.05	+.06	-.01
0025	-.54	+.18	+.38	+.3	+.29	+.25	+.14	+.18	+.08	+.05	+.25	+.21	+.15	+.08	+.09	+.1	+.03	+.51	+.52	+.51	+.51	+.47	+.48	+.33	1	+.23	+.6	+.49	+.05	+.04	-.01
0026	-.17	+.06	+.12	+.1	+.1	+.09	+.05	+.06	+.03	-.01	+.09	+.08	+.06	+.03	+.04	+.03	-.02	+.16	+.16	+.16	+.16	+.15	+.15	+.11	+.23	1	+.18	+.15	+.01	+.01	0
0027	-.46	+.15	+.32	+.25	+.24	+.2	+.12	+.16	+.07	+.05	+.16	+.15	+.1	+.05	+.06	+.08	+.04	+.43	+.44	+.43	+.43	+.4	+.4	+.28	+.6	+.18	1	+.29	+.05	+.1	0
0028	-.37	+.12	+.26	+.2	+.19	+.16	+.09	+.12	+.06	+.04	+.15	+.13	+.09	+.05	+.05	+.07	+.04	+.35	+.36	+.35	+.35	+.33	+.33	+.23	+.49	+.15	+.29	1	+.05	+.08	0
0029	-.09	-.31	+.06	-.46	-.49	-.45	-.29	-.61	-.31	-.19	+.07	+.06	+.04	+.02	+.05	+.03	+.02	+.09	+.09	+.08	+.08	+.08	+.08	+.05	+.05	+.01	+.05	+.05	1	+.06	0
0030	-.1	+.02	+.07	+.03	+.03	+.02	+.01	-.03	+.05	+.06	+.07	+.06	+.04	+.02	+.02	+.02	+.04	+.1	+.1	+.1	+.1	+.09	+.09	+.06	+.04	+.01	+.1	+.08	+.06	1	0
0031	0	0	-.01	0	-.02	-.01	-.02	0	0	0	0	+.01	-.01	0	0	0	0	-.07	+.03	-.05	+.04	-.04	+.01	-.01	-.01	0	0	0	0	0	1
0032	-.24	+.08	+.17	+.13	+.12	+.1	+.06	+.11	+.06	+.07	+.12	+.11	+.07	+.03	+.03	+.04	+.04	+.23	+.23	+.23	+.21	+.15	+.25	-.66	+.27	+.23	+.23	-.13			
0033	+.08	-.14	-.05	-.44	-.43	-.36	-.23	-.29	-.15	-.1	+.1	+.1	+.05	+.01	-.02	-.03	-.03	-.16	-.07	-.13	-.07	-.1	-.06	-.09	-.03	-.03	-.15	-.1	+.31	-.02	+.07
0034	+.17	-.06	-.12	-.09	-.09	-.07	-.04	-.06	-.03	-.03	+.04	-.02	+.07	0	-.03	-.01	-.06	-.23	-.17	-.24	-.16	-.21	-.15	-.16	-.23	+.37	-.25	-.19	-.02	-.03	+.04
0035	+.08	-.03	-.06	-.03	-.04	-.02	-.03	-.01	-.01	+.02	-.09	-.07	-.07	-.06	-.1	-.06	-.16	-.08	-.13	-.07	-.1	-.07	-.07	-.05	-.01	0	+.03	0	0	+.07	
0036	-.01	0	+.01	-.01	0	0	+.01	0	0	0	0	-.01	0	0	0	0	0	+.02	-.06	+.01	-.08	+.01	-.04	+.01	+.01	0	0	0	+.01	0	-.83
0037	-.04	+.01	+.03	+.02	+.02	+.01	+.01	+.01	+.01	+.01	+.01	+.01	0	0	0	+.01	+.01	0	+.04	+.01	+.04	+.02	+.03	+.02	+.04	-.02	+.05	+.04	+.01	+.03	+.03
0038	+.01	-.03	-.01	-.32	-.29	-.24	-.15	-.09	-.04	-.04	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.01	-.02	-.01	-.01	-.01	-.01	-.01	0	-.01	-.01	+.01	0	+.01	
0039	+.02	-.07	-.01	-.62	-.57	-.46	-.3	-.17	-.08	-.07	-.02	-.01	-.01	-.01	-.02	-.02	-.01	-.03	-.02	-.03	-.02	-.03	-.02	-.02	-.02	-.01	-.01	-.01	+.01	0	+.03
0040	-.02	-.07	+.01	-.17	-.16	-.14	-.09	-.28	-.29	-.26	+.01	+.01	+.01	+.01	+.03	+.02	-.02	+.01	+.02	+.01	+.02	+.02	+.02	+.01	+.09	+.19	0	0	+.21	-.76	+.01
0041	-.22	+.07	+.15	+.11	+.11	+.09	+.05	+.07	+.04	+.03	-.13	-.14	-.1	-.05	+.04	0	+.03	+.21	+.21	+.21	+.2	+.19	+.19	+.13	+.12	+.07	+.35	+.26	+.03	+.08	-.01
0042	-.21	+.07	+.14	+.11	+.1	+.09	+.05	+.07	+.04	+.04	-.04	+.01	-.07	-.01	+.03	+.01	+.05	+.18	+.2	+.2	+.19	+.18	+.18	+.13	+.28	-.44	+.3	+.23	+.03	+.05	0
0043	-.13	+.04	+.09	+.07	+.07	+.06	+.03	+.04	+.03	+.01	-.11	-.06	-.06	-.03	+.04	-.02	+.01	+.12	+.12	+.12	+.12	+.11	+.11	+.08	+.19	+.05	-.33	+.32	+.01	+.01	-.01
0044	-.11	+.04	+.08	+.06	+.06	+.05	+.03	+.04	+.02	+.01	-.09	-.05	-.05	-.02	+.03	-.01	+.01	+.1	+.11	+.11	+.1	+.1	+.1	+.07	+.17	+.04	+.33	-.54	+.01	+.01	-.01

Table A.2: Correction factor correlation matrix. The top row and left column show correction fact‹ the matrix shows the correlation between the correction factors corresponding to the column and ro‹ dimensionless; the elements along the diagonal are unity; the matrix is symmetric; positive elements in‹ and negative elements anti-correlation.

179

factors is $\rho_{ij} = \Sigma_{ij}/\sigma_i\sigma_j$. The variances of each correction factor, corresponding to the diagonal elements of the covariance matrix, are shown in Table 4.2. The correlation matrix obtained is shown in Table A.2.

A.4   Correction factor values

This section provides notes on the values of the VISTA correction factors obtained from a global fit of Standard Model prediction to data. The correction factors considered are numbers that can in principle be calculated *a priori*, but whose calculation is in practice not feasible. These correction factors divide naturally into two classes, the first of which reflects the difficulty of calculating the Standard Model prediction to all orders, and the second of which reflects the difficulty of understanding from first principles the response of the experimental apparatus.

The theoretical correction factors considered are of two types. The difficulty of calculating the Standard Model prediction for many processes to all orders in perturbation theory is handled through the introduction of k-factors, representing the ratio of the true all orders prediction to the prediction at lowest order in perturbation theory. Uncertainties in the distribution of partons inside the colliding proton and anti-proton as a function of parton momentum are in principle handled through the introduction of correction factors associated with parton distribution functions, but there are currently no discrepancies to motivate this.

Experimental correction factors correspond to numbers describing the response of the CDF detector that are precisely calculable in principle, but that are in practice best constrained by the high-p_T data themselves. These correction factors take the form of the integrated luminosity, object identification efficiencies, object misidentification probabilities, trigger efficiencies, and energy scales.

A.4.1 k-factors

For nearly all Standard Model processes, k-factors are used as an overall multiplicative constant, rather than being considered to be a function of one or more kinematic
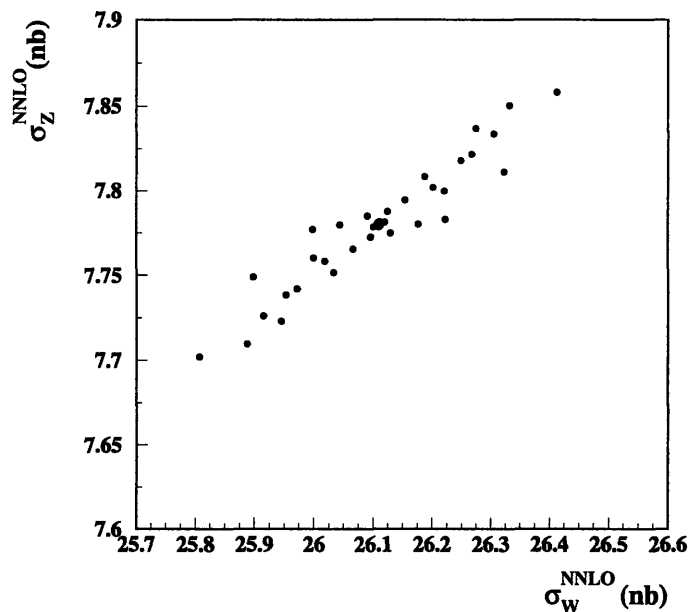
Figure A-8: Variation of the k-factors for inclusive W and Z production under different choices of parton distribution functions, from the Alekhin parton distribution error set [89]. The correlation of the uncertainty on these two k-factors due to uncertainty in the parton distribution functions is 0.955.

variables. The spirit of the approach is to introduce as few correction factors as possible, and to only introduce correction factors motivated by specific discrepancies.

0001. The integrated luminosity of the analysis sample has a close relationship with the theoretically determined values of inclusive W and Z production at the Tevatron. Figure A-8 shows the variation in calculated inclusive W and Z k-factors under changes in the assumed parton distribution functions. Each point represents a different W and Z inclusive cross section determined using modified parton distribution functions. The use of 16 bases to reflect systematic uncertainties results in 32 black dots in Fig. A-8. The uncertainties in the W and Z cross sections due to variations in the renormalization and factorization scales are nearly 100% correlated; varying these scales affects both the W and Z inclusive cross sections in the same way. The uncertainties in the parton distribution functions and the choice of renormalization and factorization scales represent the dominant contributions to the theoretical uncertainty in the total inclusive W and Z cross section calculations at the Tevatron. The term in $\chi^2_{\text{constraints}}$ that reflects our knowledge of the theoretical
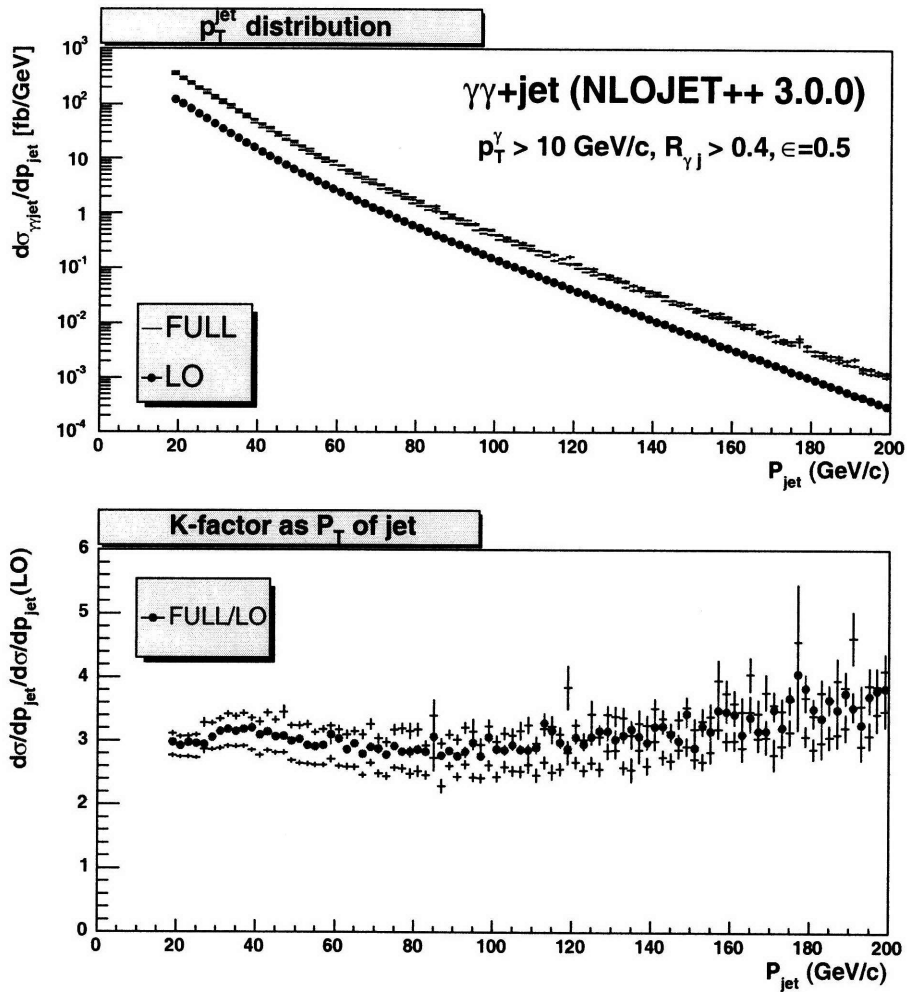
Figure A-9: Calculation of the $\gamma\gamma j$ k-factor, as a function of jet transverse momentum. The effect of changing the factorization scale by a factor of two in either direction is also shown (small black points with error bars).

prediction of the inclusive W and Z cross sections explicitly acknowledges this high degree of correlation.

Theoretical constraints on all other k-factors are assumed to be uncorrelated with each other, not because the uncertainties of these calculations are indeed uncorrelated, but rather because the correlations among these computations are poorly known.

0002, 0003. The `cosmic` γ and `cosmic` j backgrounds are estimated using events recorded in the CDF data with one or more reconstructed photons and with two or fewer reconstructed tracks. The use of events with two or fewer reconstructed tracks is a new technique for estimating these backgrounds. These correction factors are

primarily constrained by the number of events in the VISTA $\gamma\not p$ and $j\not p$ final states. The values are related to (and consistent with) the fraction of bunch crossings with one or more inelastic $p\bar{p}$ interactions, complicated slightly by the requirement that any jet falling in the final state $j\not p$ has at least 5 GeV of track p_T within a cone of 0.4 relative to the jet axis.

0004, 0005, 0006, 0007. The NLOJET++ calculation of the γj inclusive k-factor constrains the cross section weighted sum of the γj, $\gamma 2j$, $\gamma 3j$, and $\gamma 4j$ correction factors to 1.25 ± 0.15 [85, 86].

0008, 0009, 0010. The DIPHOX calculation of the inclusive $\gamma\gamma$ cross section at NLO constrains the weighted sum of these correction factors to 2.0 ± 0.15 [87]. From Table 4.2, the $\gamma\gamma j$ k-factor (0009) appears anomalously large. Figure A-9 shows a calculation of this $\gamma\gamma j$ k-factor using NLOJET++ [85, 86] as a function of summed transverse momentum. The NLO correction to the LO prediction is found to be large, and not manifestly inconsistent with the value for this k-factor determined from the VISTA fit. The cross section for $\gamma\gamma 2j$ production has not been calculated at NLO.

0011, 0012, 0013, 0014. These correction factors correspond to k-factors for W production in association with zero, one, two, and three or more jets, respectively. A linear combination of these correction factors is constrained by the requirement that the inclusive W production cross section is consistent with the NNLO calculation of Ref. [89]. The values of these correction factors, and their trend of decreasing as the number of jets increases, depends heavily on the choice of renormalization and factorization scales. The individual correction factors are not explicitly constrained by a NLO calculation.

0015, 0016, 0017. These correction factors correspond to k-factors for Z production in association with zero, one, and two or more jets, respectively. A linear combination of these correction factors is constrained by the requirement that the inclusive Z production cross section is consistent with the NNLO calculation of Ref. [89].

0018, 0019. The two k-factors for dijet production correspond to two bins in $\hat{p}_T$, the p_T of the hard two to two scattering in the parton center of mass frame. These correction factors are constrained by a NLO calculation [90], and show expected behavior as a function of $\hat{p}_T$.

0020, 0021. The two k-factors for 3-jet production, corresponding to two bins in $\hat{p}_T$, are unconstrained by any NLO calculation, but show reasonable behavior as a function of $\hat{p}_T$.

0022, 0023. The k-factors for 4-jet production, corresponding to two bins in $\hat{p}_T$, are unconstrained by any NLO calculation, but show reasonable behavior as a function of $\hat{p}_T$.

0024. The k-factor for the production of five or more jets, constrained primarily by the VISTA low-p_T $5j$ final state, is found to be close to unity.

A.4.2 Identification efficiencies

The correction factors in this section, although billed as "identification efficiencies," are truly ratios of the identification efficiency in the data relative to the identification efficiency in CDFSIM. A correction factor value of unity indicates a proper modeling of the overall identification efficiency by CDFSIM; a correction factor value of 0.5 indicates that CDFSIM overestimates the overall identification efficiency by a factor of two.

0025. The central electron identification efficiency scale factor is close to unity, indicating the central electron efficiency measured in data is similar (to within 1%) to the central electron efficiency in the CDF detector simulation. This reflects an emphasis within CDF on tuning the detector simulation for central electrons. The determination of this correction factor is dominated by the VISTA final states $e\not{p}$ and e^+e^-, where one of the electrons has $|\eta| < 1$.

0026. The plug electron identification efficiency scale factor is several percent less than unity, indicating that the CDF detector simulation slightly overestimates the electron identification efficiency in the plug region of the CDF detector. The determination of this correction factor is dominated by the VISTA final states $e\not{p}$ and e^+e^-, where one of the electrons has $|\eta| > 1$.

0027, 0028. To reduce backgrounds hypothesized to arise from pion and kaon decays in flight with a substantially mismeasured track, a very good track fit in the CDF tracker is required. Partially due to this tight track fit requirement, CDF muon identification efficiencies in the regions $|\eta| < 0.6$ and $0.6 < |\eta| < 1.5$ are overestimated in the CDF detector simulation by over 10%. The determination of the identification efficiencies $p(\mu \to \mu)$ is dominated by the VISTA final states $\mu\not{p}$ and $\mu^+\mu^-$.

0029. The central photon identification efficiency scale factor is determined primarily by the number of events in the VISTA final states $j\gamma$ and $\gamma\gamma$. The uncertainty on this correction factor is highly correlated with the uncertainties on the γj k-factor, the $p(j \to \gamma)$ fake rate, and the $\gamma\gamma$ k-factor.

0030. The plug photon identification efficiency scale factor is determined primarily by the number of events in the VISTA final state $\gamma\gamma$. The uncertainty on this correction factor is highly correlated with the uncertainty on the plug $p(j \to \gamma)$ fake rate.

0031. The b-jet identification efficiency is determined to be consistent with the prediction from CDFSIM.

A.4.3 Fake rates

0032. The fake rate $p(e \to \gamma)$ for electrons to be misreconstructed as photons in the plug region of the detector is added on top of the significant number of electrons misreconstructed as photons by CDFSIM.

0033. In VISTA, the contribution of jets faking electrons is modeled by applying a fake rate $p(j \to e)$ to Monte Carlo jets. VISTA represents the first large scale Tevatron analysis in which a completely Monte Carlo based modeling of jets faking electrons is employed. Significant understanding of the physical mechanisms contributing to this fake rate has been achieved, as summarized in Appendix A.1. Consistency with this understanding is required; for example, $p(j \to e) \approx p(j \to \gamma)p(\gamma \to e)$. The value of this correction factor is determined primarily by the number of events in the VISTA final state ej, where the electron is identified in the central region of the CDF detector. It is notable that this fake rate is independent of global event properties, and that a consistent simultaneous understanding of the ej, $e2j$, $e3j$, and $e4j$ final states is obtained.

0034. The value of the fake rate $p(j \to e)$ in the plug region of the CDF detector is roughly one order of magnitude larger than the corresponding fake rate $p(j \to e)$ in the central region of the detector, consistent with an understanding of the relative performance of the detector in the central and plug regions for the identification of electrons. This correction factor is determined primarily by the number of events in the VISTA final state ej, where the electron is identified in the plug region of the CDF detector.

0035. In VISTA, the contribution of jets faking muons is modeled by applying a fake rate $p(j \to \mu)$ to Monte Carlo jets. VISTA represents the first large scale Tevatron analysis in which a completely Monte Carlo based modeling of jets faking muons is employed. The value obtained from the VISTA fit is seen to be roughly one order of magnitude smaller than the fake rate $p(j \to e)$ in the central region of the detector, consistent with our understanding of the physical mechanisms underlying these fake rates, as described in Appendix A.1. The value of this correction factor is determined primarily by the number of events in the VISTA final state $j\mu$.

0036. The fake rate $p(j \to b)$ has p_T dependence explicitly imposed. The number of tracks inside a typical jet, and hence the probability that a secondary vertex is

(mis)reconstructed, increases with jet p_T. The values of these correction factors are consistent with the mistag rate determined using secondary vertices reconstructed on the other side of the beam axis with respect to the direction of the tagged jet [91]. The value of this correction factor is determined primarily by the number of events in the VISTA final states bj and bb.

0037, 0038. The fake rate $p(j \to \tau)$ decreases with jet p_T, since the number of tracks inside a typical jet increases with jet p_T. The values of these correction factors are determined primarily by the number of events in the VISTA final state $j\tau$.

0039, 0040. The fake rate $p(j \to \gamma)$ is determined separately in the central and plug regions of the CDF detector. The values of these correction factors are determined primarily by the number of events in the VISTA final states $j\gamma$ and $\gamma\gamma$. The value obtained for **0039** is consistent with the value obtained from a study using detailed information from the central preshower detector. The fake rate determined in the plug region is noticeably higher than the fake rate determined in the central region, as expected.

A.4.4   Trigger efficiencies

0041. The central electron trigger inefficiency is dominated by not correctly reconstructing the electron's track at the first online trigger level.

0042. The plug electron trigger inefficiency is due to inefficiencies in clustering at the second online trigger level.

0043, 0044. The muon trigger inefficiencies in the regions $|\eta| < 0.6$ and $0.6 < |\eta| < 1.0$ derive partly from tracking inefficiency, and partly from an inefficiency in reconstructing muon stubs in the CDF muon chambers.

The value of these corrections factors are consistent with other trigger efficiency measurements made using additional information [92].

187

A.4.5 Energy scales

The VISTA infrastructure also allows the jet energy scale to be treated as a correction factor. At present this correction factor is not used, since there is no discrepancy requiring it.

To understand the effect of introducing such a correction factor, a jet energy scale correction factor is added and constrained to 1 ± 0.03, reflecting the jet energy scale determination at CDF [50]. The fit returns a value with a very small error, since this correction factor is highly constrained by the low-p_T $2j$, $3j$, ej, and $e\,2j$ final states. Assuming perfectly correct modeling of jets faking electrons, as described in Appendix A.1, this is a correct energy scale error. The inclusion of additional correction factor degrees of freedom to reflect possible imperfections in this modeling of jets faking electrons increases the energy scale error. The interesting conclusion is that the jet energy scale (considered as a lone free parameter) is very well constrained by the large number of dijet events; adjustment to the jet energy scale must be accompanied by simultaneous adjustment of other correction factors (such as the dijet k-factor) in order to retain agreement with data.

A.5 SLEUTH details

This appendix elaborates on the SLEUTH partitioning rule, and on the minimum number of events required for a final state to be considered by SLEUTH.

A.5.1 Partitioning

Table A.3 lists the VISTA final states associated with each SLEUTH final state.

A.5.2 Minimum number of events

This section expands on a subtle point in the definition of the SLEUTH algorithm: for purely practical considerations, only final states in which three or more events are observed in the data are considered.

SLEUTH	VISTA Final States
bb	$bj,\ b2j,\ 2bj,\ 2b,\ 3b$
$\bar b b \ell^+\ell^-$	$e^+e^-bj,\ \mu^+\mu^-bj,\ e^+e^-b2j,\ \mu^+\mu^-b2j,$ $\mu^+\mu^-2bj,\ e^+e^-2b,\ \mu^+\mu^-2b$
$\bar b b \ell^+\ell^-2j$	$e^+e^-b3j,\ \mu^+\mu^-b3j$
$\bar b b \ell^+\ell^-2j\not p$	$\mu^+\mu^-2b2j\not p$
$\bar b b \ell^+\ell^-\not p$	$e^+e^-bj\not p,\ \mu^+\mu^-bj\not p,\ e^+e^-b2j\not p,\ \mu^+\mu^-b2j/$ $p,\ e^+e^-2bj\not p,\ e^+e^-2b\not p,\ \mu^+\mu^-2b\not p$
$\bar b b \ell^+2j\ell'^-\not p$	$e^+\mu^-b3j\not p$
$\bar b b \ell^+2j\gamma\not p$	$\mu^+\gamma b3j\not p$
$W\bar b b jj$	$e^+b3j\not p,\ \mu^+b3j\not p,\ e^+2b2j\not p,\ \mu^+2b2j\not p$
$\bar b b \ell^+2j\tau^+$	$\mu^+\tau^+b3j$
$\bar b b \ell^+\ell'^+$	$e^+\mu^+2b$
$\bar b b \ell^+\ell'^-$	$e^+\mu^-bj$
$\bar b b \ell^+\ell'^-\not p$	$e^+\mu^-bj\not p,\ e^+\mu^-b2j\not p,\ e^+\mu^-2b\not p$
$\bar b b \ell^+\gamma\not p$	$e^+\gamma b2j\not p,\ \mu^+\gamma b2j\not p$
$W\bar b b$	$e^+bj\not p,\ \mu^+bj\not p,\ e^+b2j\not p,\ \mu^+b2j\not p,\ e^+2b\not p,$ $e^+2bj\not p,\ \mu^+2bj\not p,\ \mu^+2b\not p,\ e^+3b\not p$
$\bar b b \ell^+\not p\tau^-$	$e^+\tau^-b2j\not p,\ e^+\tau^-bj\not p,\ \mu^+\tau^-bj\not p,\ e^+\tau^-2b\not p$
$\bar b b \ell^+\tau^+$	$e^+\tau^+bj$
$\bar b b \ell^+\tau^-$	$e^+\tau^-bj,\ e^+\tau^-b2j,\ e^+\tau^-2b,\ \mu^+\tau^-bj$
$bb2j$	$b3j,\ 2b2j$
$bb2j\gamma$	$\gamma b3j,\ \gamma 2b2j$
$bb2j\gamma\not p$	$\gamma b3j\not p$
$bb2j\not p$	$b3j\not p,\ 2b2j\not p$
γbb	$\gamma bj,\ \gamma b2j,\ \gamma 2b,\ \gamma 2bj,\ \gamma 3b$
$bb\gamma\not p$	$\gamma bj\not p,\ \gamma b2j\not p,\ \gamma 2b\not p$
$bb\not p$	$b2j\not p,\ 2bj\not p,\ bj\not p,\ 2b\not p$
$bb\not p\tau$	$\tau^+bj\not p$
$bb\tau^+\tau^-$	$\tau^+\tau^-bj,\ \tau^+\tau^-b2j$
$bb4j$	$2b4j$
$2b4j\not p$	$b6j\not p$
$\gamma\gamma bb$	$2\gamma bj,\ 2\gamma b2j$
$2b6j$	$2b6j$
$\ell^+\ell^-$	$e^+e^-,\ \mu^+\mu^-,\ e^+e^-j,\ \mu^+\mu^-j,\ e^+e^-b,$ $\mu^+\mu^-b$
$\ell^+\ell^-2j$	$e^+e^-2j,\ \mu^+\mu^-2j,\ e^+e^-3j,\ \mu^+\mu^-3j$
$\ell^+\ell^-\ell'2j\not p$	$e^+e^-\mu^+2j\not p$
$\ell^+\ell^-2j\gamma$	$e^+e^-\gamma 2j,\ \mu^+\mu^-\gamma 2j,\ e^+e^-\gamma 3j,\ \mu^+\mu^-\gamma 3j$
$\ell^+\ell^-2j\not p$	$e^+e^-2j\not p,\ \mu^+\mu^-2j\not p,\ e^+e^-3j\not p,\ \mu^+\mu^-3j\not p$
$\ell^+\ell^-\tau^+2j\not p$	$e^+e^-\tau^+2j$
$\ell^+\ell^-\ell'\gamma\not p$	$e^+\mu^+\mu^-\gamma j$
$\ell^+\ell^-\ell'\not p$	$e^+\mu^+\mu^-,\ e^+e^-\mu^+\not p,\ e^+e^-\mu^+j\not p,\ e^+e^-\mu^+,$ $e^+\mu^+\mu^-j,\ e^+\mu^+\mu^-\not p$
$\ell^+\ell^-\gamma$	$e^+e^-\gamma,\ \mu^+\mu^-\gamma,\ e^+e^-\gamma j,\ \mu^+\mu^-\gamma j$
$\ell^+\ell^-\gamma\not p$	$e^+e^-\gamma\not p,\ e^+e^-\gamma j\not p$
$\ell^+\ell^-\not p$	$e^+e^-\not p,\ \mu^+\mu^-\not p,\ e^+e^-j\not p,\ \mu^+\mu^-j\not p,$ $\mu^+\mu^-b\not p,\ e^+e^-b\not p$
$\ell^+\ell^-\not p\tau^+$	$e^+e^-\tau^+,\ e^+e^-\tau^+j,\ \mu^+\mu^-\tau^+$
$\ell^+\ell^-4j$	$e^+e^-4j,\ \mu^+\mu^-4j$
$\ell^+\ell^-4j\not p$	$e^+e^-4j\not p,\ \mu^+\mu^-4j\not p$
$\ell^+\ell^-\tau^+4j\not p$	$e^+e^-\tau^+4j$
$\ell^+\ell'^+jj$	$e^+\mu^+2j,\ e^+\mu^+3j$
$\ell^+\ell'^+\not p jj$	$e^+\mu^+2j\not p$
$\ell^+\ell'^-jj$	$e^+\mu^-2j,\ e^+\mu^-3j$
$\ell^+\ell'^-\not p jj$	$e^+\mu^-2j\not p,\ e^+\mu^-3j\not p$
$W\gamma jj$	$\mu^+\gamma 2j\not p,\ e^+\gamma 2j\not p,\ \mu^+\gamma 3j\not p,\ e^+\gamma 3j\not p$
Wjj	$e^+2j\not p,\ \mu^+2j\not p,\ e^+3j\not p,\ \mu^+3j\not p$
$\ell^+\tau^+\not p jj$	$\mu^+\tau^+2j\not p$
$\ell^+\tau^-\not p jj$	$e^+\tau^-2j\not p,\ e^+\tau^-3j\not p,\ \mu^+\tau^-3j\not p,\ \mu^+\tau^-2j\not p$
$\ell^+\tau^+jj$	$e^+\tau^+2j,\ e^+\tau^+3j,\ \mu^+\tau^+2j,\ \mu^+\tau^+3j$
$\ell^+\tau^-jj$	$e^+\tau^-2j,\ \mu^+\tau^-2j,\ e^+\tau^-3j,\ \mu^+\tau^-3j$
$\ell^+\ell'^+$	$e^+\mu^+,\ e^+\mu^+j$
$\ell^+\ell'^+\not p$	$e^+\mu^+\not p,\ e^+\mu^+j\not p$
$\ell^+\ell'^-$	$e^+\mu^-,\ e^+\mu^-j$
$\ell^+\ell'^-\gamma\not p$	$e^+\mu^-\gamma j\not p$

SLEUTH	VISTA Final States
$\ell^+\ell'^-\not p$	$e^+\mu^-\not p,\ e^+\mu^-j\not p,\ e^+\mu^-b\not p$
$W\gamma$	$\mu^+\gamma\not p,\ e^+\gamma\not p,\ \mu^+\gamma j\not p,\ e^+\gamma j\not p,\ e^+\gamma b\not p$
$\ell^+\tau^-\gamma\not p$	$e^+\tau^-\gamma\not p$
$\ell^+\tau^+\gamma$	$e^+\tau^+\gamma$
$\ell^+\tau^-\gamma$	$e^+\tau^-\gamma,\ \mu^+\tau^-\gamma j$
W	$e^+\not p,\ \mu^+\not p,\ e^+j\not p,\ \mu^+j\not p,\ e^+b\not p,\ \mu^+b\not p$
$\ell^+\tau^+\not p$	$e^+\tau^+\not p,\ \mu^+\tau^+\not p,\ e^+\tau^+j\not p,\ \mu^+\tau^+j\not p$
$\ell^+\tau^-\not p$	$e^+\tau^-\not p,\quad \mu^+\tau^-\not p,\quad e^+\tau^-j\not p,\quad \mu^+\tau^-j\not p,$ $\mu^+\tau^-b\not p$
$\ell^+\tau^+$	$e^+\tau^+,\ e^+\tau^+j,\ \mu^+\tau^+,\ \mu^+\tau^+j,\ e^+\tau^+b$
$\ell^+\tau^-$	$e^+\tau^-,\ e^+\tau^-j,\ \mu^+\tau^-,\ \mu^+\tau^-j,\ e^+\tau^-b$
$\ell^+\ell'^+4j\not p$	$e^+\mu^+4j\not p$
$\ell^+\ell'^-4j$	$e^+\mu^-4j$
$\ell^+\ell'^-\not p4j$	$e^+\mu^-4j\not p$
$W\gamma 4j$	$\mu^+\gamma 4j\not p,\ e^+\gamma 4j\not p$
$W4j$	$e^+4j\not p,\ \mu^+4j\not p$
$\ell^+4j\tau^+$	$e^+\tau^+4j$
$\ell^+4j\tau^-$	$e^+\tau^-4j,\ \mu^+\tau^-4j$
$W\gamma\gamma$	$\mu^+2\gamma\not p,\ e^+2\gamma\not p,\ e^+2\gamma j\not p,\ \mu^+2\gamma j\not p$
jj	$2j,\ 3j$
γjj	$\gamma 2j,\ \gamma 3j$
$\gamma\not p jj$	$\gamma 2j\not p,\ \gamma 3j\not p$
$\gamma\not p\tau^+jj$	$\tau^+\gamma 2j\not p$
$jj\not p$	$3j\not p,\ 2j\not p$
$\tau\not p jj$	$\tau^+2j\not p,\ \tau^+3j\not p$
$\tau^+\tau^-2j$	$\tau^+\tau^-2j,\ \tau^+\tau^-3j$
$\gamma\gamma jj$	$2\gamma 2j,\ 2\gamma 3j$
$jj\gamma\gamma\not p$	$2\gamma 2j\not p,\ 2\gamma 3j\not p$
$2\tau^+2j$	$2\tau^+2j,\ 2\tau^+3j$
$\gamma\gamma\gamma jj$	$3\gamma 2j$
γj	$\gamma j,\ \gamma b$
$\gamma\not p$	$\gamma\not p,\ \gamma j\not p,\ \gamma b\not p$
$\tau\not p\gamma$	$\tau^+\gamma j\not p,\ \tau^+\gamma\not p$
$\tau^+\tau^-\gamma$	$\tau^+\tau^-\gamma$
$\tau^+\tau^+\gamma$	$2\tau^+\gamma$
$j\not p$	$j\not p,\ b\not p$
$\tau\not p$	$\tau^+j\not p,\ \tau^+b\not p$
$\tau^+\tau^-\not p$	$\tau^+\tau^-j\not p,\ \tau^+\tau^-\not p$
$\tau^+\tau^-$	$\tau^+\tau^-,\ \tau^+\tau^-j,\ \tau^+\tau^-b$
$bbbb$	$3bj,\ 3b2j,\ 4b$
$Wbbbb$	$e^+3bj\not p,\ \mu^+3bj\not p$
$4b2j$	$3b3j,\ 3b4j,\ 4b2j$
$\gamma bbbb$	$\gamma 3bj$
$bbbb\not p$	$3bj\not p,\ 3b2j\not p$
$4b4j$	$4b4j$
$\ell^+\ell^+$	$2e^+,\ 2e^+j,\ 2\mu^+$
$\ell^+\ell^-\ell^+jj\not p$	$2e^+e^-2j,\ 2e^+e^-3j$
$\ell^+\ell^-\ell^+\not p$	$2e^+e^-,\ 2e^+e^-j,\ 2e^+e^-\not p$
$\ell^+\ell^+jj$	$2e^+2j$
$\ell^+\ell^+\ell'^-\not p$	$e^+2\mu^-\not p$
$\ell^+\ell^+\gamma$	$2e^+\gamma$
$\ell^+\ell^+\gamma\not p$	$2e^+\gamma\not p$
$\ell^+\ell^+\not p$	$2e^+\not p,\ 2\mu^+\not p,\ 2e^+j\not p,\ 2\mu^+j\not p$
$\ell^+\ell^+4j$	$2e^+4j$
$4j$	$4j$
$\gamma 4j$	$\gamma 4j$
$\gamma 4j\not p$	$\gamma 4j\not p$
$4j\not p$	$4j\not p$
$\tau^+\not p4j$	$\tau^+4j\not p$
$\tau^+\tau^-4j$	$\tau^+\tau^-4j$
$\gamma\gamma 4j$	$2\gamma 4j$
$\gamma\gamma$	$2\gamma,\ 2\gamma j,\ 2\gamma b$
$\gamma\gamma\not p$	$2\gamma\not p,\ 2\gamma j\not p$
$\tau^+\tau^+$	$2\tau^+,\ 2\tau^+j$
3γ	3γ

Table A.3: Correspondence between SLEUTH and VISTA final states. The first column shows the SLEUTH final state formed by merging the populated VISTA final states in the second column. Charge conjugates of each VISTA final state are implied.

189

Suppose $\mathcal{P}_{e^+e^-b\bar{b}} = 10^{-6}$; then in computing $\tilde{\mathcal{P}}$ all final states with $b > 10^{-6}$ must be considered and accounted for. (A final state with $b = 10^{-7}$, on the other hand, counts as only ≈ 0.1 final states, since the fraction of hypothetical similar experiments in which $\mathcal{P} < 10^{-6}$ in this final state is equal to the fraction of hypothetical similar experiments in which one or more events is seen in this final state, which is 10^{-7}.) This is a large practical problem, since it requires that all final states with $b > 10^{-6}$ be enumerated and estimated, and it is difficult to do this believably.

To solve this problem, let SLEUTH consider only final states with at least $d_{\min}$ events observed in the data. The goal is to be able to find $\tilde{\mathcal{P}} < 10^{-3}$. There will be some number $N_{\text{fs}}(b_{\min})$ of final states with expected number of events $b > b_{\min}$, writing N_{fs} explicitly as a function of $b_{\min}$; thus $b_{\min}$ must be chosen to be sufficiently large that all of these $N_{\text{fs}}(b_{\min})$ final states can be enumerated and estimated. The time cost of simulating events is such that the integrated luminosity of Monte Carlo events is at most 100 times the integrated luminosity of the data; this practical constraint restricts $b_{\min} > 0.01$. The number of SLEUTH Tevatron Run II final states with $b > 0.01$ is $N_{\text{fs}}(b_{\min} = 0.01) \approx 10^3$.

For small $\mathcal{P}_{\min}$, keeping the first term in a binomial expansion yields $\tilde{\mathcal{P}} = \mathcal{P}_{\min} N_{\text{fs}}(b_{\min})$, where $\mathcal{P}_{\min}$ is the smallest $\mathcal{P}$ found in any final state. From the discussion above, the computation of $\tilde{\mathcal{P}}$ from $\mathcal{P}_{\min}$ can only be justified if $\mathcal{P}_{\min} > (b_{\min}{}^{d_{\min}})$; if otherwise, final states with $b < b_{\min}$ will need to be accounted for. Thus $\tilde{\mathcal{P}}$ can be confidently computed only if $\tilde{\mathcal{P}} > (b_{\min}{}^{d_{\min}}) N_{\text{fs}}(b_{\min})$.

Solving this inequality for $d_{\min}$ and inserting values from above,

$$d_{\min} \geq \frac{\log_{10} \tilde{\mathcal{P}} - \log_{10} N_{\text{fs}}(b_{\min})}{\log_{10} b_{\min}} \approx \frac{-3 - 3}{-2} = 3. \tag{A.28}$$

A believable trials factor can be computed if $d_{\min} \geq 3$.

At the other end of the scale, computational strength limits the maximum number of events SLEUTH is able to consider to $\lesssim 10^4$. Excesses in which the number of events exceed 10^4 are expected to be identified by VISTA's normalization statistic.

A.5.3 p-val$_{\mathrm{min}}$, population and $\mathcal{P}$

SLEUTH estimates $\mathcal{P}$ for a given final state by producing pseudo-data, i.e. $\sum p_T$ values that are distributed according to the Standard Model prediction. It then scans all $\sum p_T$ tails, finds the smallest p-val and compares it to the p-val$_{\mathrm{min}}$ from the actual data. That is repeated with many different distributions of pseudo-data, until the fraction of more interesting pseudo-data distributions (which is $\mathcal{P}$) is determined with 5% relative uncertainty.

In each pseudo-data distribution that is produced, the population of pseudo-data is randomly distributed according to a Poisson distribution, whose mean is the Standard Model predicted total population (B) for the final state.

Each examined $\sum p_T$ tail has a p-val that is not taking into account the statistical uncertainty in the background (b) contained in the tail. The same is true for both data and pseudo-data, therefore the effect in the final $\mathcal{P}$ is negligible.

Regardless of the particular shape of an expected $\sum p_T$ distribution, p-val$_{\mathrm{min}}$ in pseudo-data follows the same distribution. Therefore, $\mathcal{P}$ depends only on the p-val$_{\mathrm{min}}$ observed in data, and on the overall expected population; the larger the population, the bigger the average number of considered $\sum p_T$ tails in pseudo-data, therefore the larger the $\mathcal{P}$. The dependence of $\mathcal{P}$ on p-val$_{\mathrm{min}}$ and B is shown in Fig. A-10. The advantage of having tabulated this dependence, is that then one does not have to produce pseudo-data repeatedly to estimate $\mathcal{P}$; he can simply read it from Fig. A-10, for a given B and p-val$_{\mathrm{min}}$. This technique makes the execution of SLEUTH incredibly fast, allowing for studies such as sensitivity tests, projections to different luminosity, propagation of systematic uncertainties to $\tilde{\mathcal{P}}$, and frequent assessment of the $\sum p_T$ excesses in data.
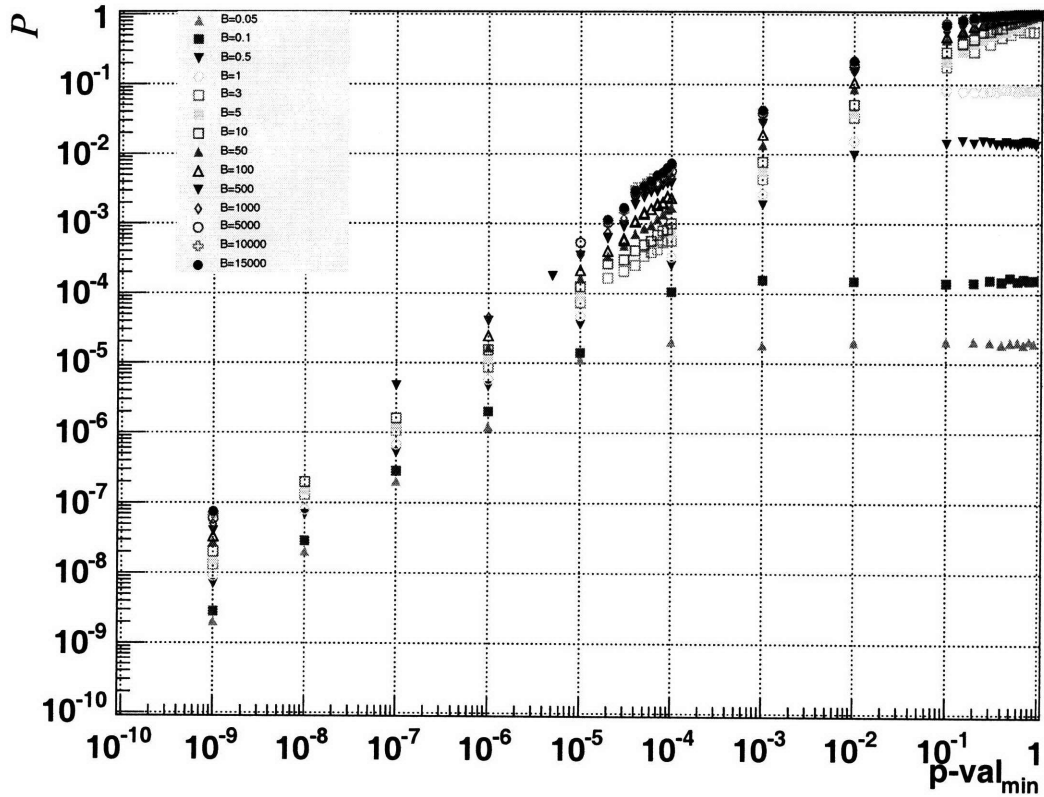
Figure A-10: $\mathcal{P}$ as a function of $p\text{-val}_{\min}$, for final states of different expected populations B. $\mathcal{P}$ reaches a plateau at $p\text{-val}_{\max} = \sum_{i=3}^{\infty} \frac{B^i}{i!} e^{-B}$, which is visible for small B, and reflects the requirement to have at least 3 data events in a $\sum p_T$ tail to consider it. $\mathcal{P}$ values have been estimated to 5% relative uncertainty.

Appendix B

Correction Model Details, reflecting the 2 fb^{-1} analysis

B.1   Details on Event Selection

Although specific online triggers are not explicitly required, it is still possible to identify the primary online triggers which feed this analysis. These are:

- electron_central_18

- muon_central_18

- photon_25_iso

- jet20

- jet100

- susy dilepton triggers: electron_central_8_&_track8 cem4_cmup4 cem4_cmx4 cem4_pem8 cmup4_pem8 cmx4_pem8 dielectron_central_4 dimuon_cmup4_cmx4 dimuon_cmupcmup4

- susy dilepton triggers muon_cmup8_&_track8 and muon_cmx8_&_track8 (introduced in run number 200274, roughly 600 pb^{-1} into Run II)

193

- hadronic ditau trigger (introduced roughly 300 pb^{-1} into Run II)

The following datasets were used:

- HighPt Central Electron stream: bhel0d, bhel0h, bhel0i, bhel0j

- HighPt CMUP and CMX muon stream: bhmu0d, bhmu0h, bhmu0i, bhmu0j

- HighPt Photon stream: cph10d, cph10h, cph10i, cph10j

- SUSY dilepton stream: edil0d, edil0h, edil0i, edil0j

- Ditau stream: etau0d, etau0h, etau0i, etau0j

- Jet20 stream: gjt10d, gjt10h, gjt10i, gjt10j

- Jet100 stream: gjt40d, gjt40h, gjt40i, gjt40j

B.2   Details on Particle Identification

This section contains tables of information related to particle identification. Electron identification is described in Tables B.1 and B.2; muon identification in Tables B.3, B.4, B.5, and B.6; tau identification in Table B.7; and photon identification in Tables B.8 and B.9. Standard fiducial criteria apply. Standard CDF SecVtx algorithm is used to identify b-jets.

Jets are identified using the JetClu [49] clustering algorithm with cone size $\Delta R = 0.4$, unless the event contains one or more jets with $p_T > 200$ GeV and no leptons or photons, in which case cones of $\Delta R = 0.7$ are used. Jets with $p_T > 150$ GeV are required to have at least 5 GeV of track p_T within the cone.

B.3   VISTA: Single Particle Gun Results

Tables B.10 and B.11 show the response of the CDF detector simulation, reconstruction, and particle identification algorithms to single particles in the central and plug regions respectively, with all changes to particle identification criteria discussed in

194

Variable	Cut
Region	Fiducial CES
Track Z_0	$\leq$ 60 cm
COT Ax. Seg.	$\geq$ 3
COT St. Seg.	$\geq$ 2
Signed CES ΔX	$-3.0 < q\Delta X < 1.5$
CES ΔZ	< 3.0 cm
Track p_T	> 10 GeV/c
p_T/E_T (if $p_T < 50$)	> 0.5
Had/Em	$< 0.055 + 0.00045 \times E$
Isolation	$< 0.1 \times E_T$
LShrTrk	< 0.2
CES StripChi2	< 10.0
Conversion	FALSE

Table B.1: Central electron identification criteria used in VISTA and SLEUTH. These correspond to TightCEM electrons as defined in [93], for Gen5 and Gen6. The conversion finder flags a second track with $|\Delta XY| < 0.2$ cm, $|\Delta \cot\theta| < 0.015$, and $p_T > 0$ GeV.

Variable	Cut
Region	$|\eta| < 2.6$
Had/Em	< 0.05
Isolation	$< 0.1 \times E_T$
PEM Chi2	< 10
PES U	> 0.65
PES V	> 0.65
PHX Track	TRUE
N SVX hits	≥ 3
deltaR(PHX Track,EM cluster)	< 0.01

Table B.2: Plug electron identification criteria used in VISTA and SLEUTH. These correspond to Tight Phoenix electrons as defined in [93], except for the cut on ΔR.

Variable	Cut
Larry curvature correction	Applied
Track Z_0	≤ 60 cm
COT Ax. Seg.	≥ 3
COT St. Seg.	≥ 3
Iso/p_T	< 0.1
EM + Had Energy	> 0.1 GeV
EM Energy	$< 2.0 + 0.0115(p - 100) \times (p > 100)$
Had Energy	$< 6.0 + 0.0280(p - 100) \times (p > 100)$
COT χ^2 (gen5)	< 3 for $p_T < 60$; < 2 for $p_T \geq 60$
COT χ^2 (gen6)	< 2
Track With Si hits	$|d_0| < 0.02$
Track Without Si hits	$|d_0| < 0.2$

Table B.3: Common muon identification criteria used in VISTA and SLEUTH.

Variable	Cut
CMU ΔX	< 7 cm
CMP ΔX	< 5 cm
CMUP Fiducial	TRUE
No bluebeam muons	For Runs < 154449

Table B.4: CMUP muon identification criteria used in VISTA and SLEUTH. These are in addition to the criteria common to all muons.

Variable	Cut
CMX ΔX	< 6 cm
COT exit radius	> 140 cm
CMX Fiducial	TRUE
Run	> 150144
Keystone and Miniskirt good	$run \geq 190697$
Exclude wedge 14, west	For $190697 \leq run \leq 209760$

Table B.5: CMX Muon identification criteria used in VISTA and SLEUTH. These are in addition to the criteria common to all muons.

Variable	Cut
BMU ΔX	< 10 cm
BMU Fiducial	TRUE

Table B.6: BMU Muon identification criteria used in VISTA and SLEUTH. These are in addition to the criteria common to all muons.

Variable	Cut
Seed Track p_T	> 10.5
Track $\|z_0\|$	< 60 cm
Fiducial ShowerMax	$9.0 < \|Z_{\text{CES}}\| < 230.0$
π^0 isolation (gen5)	No π^0 ($p_T > 0.5$) in annulus 10-30°
π^0 isolation (gen6)	SumPt of π^0 with $p_T > 0.5$ in annulus 0.15-0.4 rad < 0.6
Track isolation (gen5)	No track ($p_T > 1$) in annulus 10-30°
Track isolation (gen6)	SumPt of tracks in annulus 0.2-0.4 rad < 1.0
Calorimeter Isolation	Iso/$E_T < 0.1$
Calorimeter E_T	Cal $E_T < $ VisPt $+ 1.5\sqrt{\text{VisPt}}$
Visible Mass (tracks+π^0s)	< 1.8 GeV
Track $\|d_0\|$	< 0.2 cm
Seed Track Vertex Consistency	Abs(seedTrack z -
Consistency	Primary Vertex z) < 5 cm
One Prong Tau	N tracks in 10° cone $=1$
Electron removal	$\xi > 0.1$ and EMfraction < 0.925
Not a Muon	No matching muon stubs and Cal $E_T/$ Seed Track $p_T > 0.5$

Table B.7: Table of τ identification criteria used in VISTA and SLEUTH.

Variable	Cut
Fiducial Region (X)	CES $\|X\| < 21$ cm
Fiducial Region (Z)	$9 < $ CES Z < 230 cm
Had/Em	$< 0.125 \|\| < 0.055 + 0.00045 \times E$
Isolation ($E_T \leq 20$)	$< 0.1 \times E_T$
Isolation ($E_T > 20$)	$< 2.0 + 0.02 \times (E_T - 20)$
Track isolation, cone 0.4	SumPt $< 2 + 0.005 \times E_T$
Ntrack (N3D)	N3D ≤ 1
Track p_T (if N3D=1)	$< 1.0 + 0.005 \times E_T$
Chi2 (Strips+Wires)/2.0	< 20
2nd CES clus. $E \sin \theta$ ($E_T \leq 18$)	strip+wire $< 0.14 \times E_T$
2nd CES clus. $E \sin \theta$ ($E_T > 18$)	strip+wire $< 2.52 + 0.01 \times E_T$

Table B.8: Central photon identification criteria used in VISTA and SLEUTH. Here E_T refers to corrected photon E_T. The "2nd CES Cluster" cut is tighter than the standard photon cut.

Variable	Cut
Region	$1.2 < |\eta| < 2.6$
Had/Em ($E_T \leq 100$)	< 0.05
Had/Em ($E_T > 100$)	$< 0.05 + 0.026 \times Log(E_T/100)$
Isolation ($E_T \leq 20$)	$< 0.1 \times E_T$
Isolation ($E_T > 20$)	$< 2.0 + 0.02 \times E_T$
Track Isolation (in a cone of $dR < 0.4$)	$< 2.0 + 0.005 \times E_T$
PEM Chi2	< 10
PES U	> 0.65
PES V	> 0.65

Table B.9: Plug photon identification criteria used in VISTA and SLEUTH. Here E_T refers to corrected photon E_T. These are the standard Joint Physics cuts.

section 4.2.1. We use a single particle gun to shoot 10^5 particles of each type, with $p_T = 25$ GeV, uniformly distributed in θ and ϕ. The types of generated particles label the rows, while the resulting reconstructed objects label the columns of each table. Table B.12 shows a similar study with 10^4 particles at $p_T = 50$ GeV. These results are not directly used in the analysis, but provide a sensible cross-check for the used fake rates and identification efficiencies.

It should be noted that the number of photons reconstructed as electrons decreased compared to the last round of this analysis. As expected, the number of electrons which were identified with the wrong charge has decreased proportionately, as well as the number of π^0 reconstructed as electrons. All these are results of making the conversion filter tighter, by removing the lower p_T threshold that was previously required when looking for sibling tracks coming from conversion.

Figures B.3 and B.3 show the p_T distributions of the reconstructed object (column label), resulting from the initial particle (row label), for the central and plug region of the detector respectively. We note that the p_T resolution of reconstructed τs has worsened, consistently with obtaining p_T from calorimeter E_T rather than visible momentum.

	e^+	e^-	μ^+	μ^-	τ^+	τ^-	γ	j	b
e^+	60940	13	0	0	265	0	2009	33140	0
e^-	8	61124	0	0	0	263	1988	33021	1
μ^+	0	0	53142	0	19	0	0	2079	0
μ^-	0	1	0	53202	0	17	0	2067	0
γ	592	571	0	0	1	2	67197	27110	0
π^0	502	499	0	0	2	5	57655	38244	0
π^+	306	0	122	0	68254	5	86	29195	37
π^-	3	388	0	94	10	67327	133	30100	43
K^+	183	2	289	0	69690	8	25	26930	24
K^-	0	282	0	178	32	68036	98	28846	23
B^+	326	18	231	9	132	17	40	71606	25893
B^-	17	312	6	232	13	117	42	71921	25611
B^0	343	97	267	43	103	63	22	72021	24990
$\bar{B}^0$	83	351	34	262	58	97	20	72031	25094
D^+	269	4	198	2	2221	56	209	83659	11606
D^-	4	275	5	171	80	2292	229	83673	11595
D^0	119	20	93	5	310	1070	250	91599	5473
$\bar{D}^0$	22	95	15	99	1170	270	225	91589	5466
K^0_L	0	2	0	0	68	67	194	97538	27
K^0_S	17	18	2	1	83	445	9647	78364	0
τ^+	6750	20	4919	0	6336	10	652	55677	613
τ^-	17	6623	0	4907	9	6064	615	56201	580
u	12	8	2	0	658	29	247	98645	24
d	4	16	1	1	55	428	181	98916	21
g	10	8	0	4	29	31	12	98190	99

Table B.10: Central single particle misidentification matrix. Using a single particle gun, 10^5 particles of each type shown at the left of the table are shot with $p_T = 25$ GeV into the central CDF detector, uniformly distributed in θ and in ϕ. The resulting reconstructed object types are shown at the top of the table, labeling the table columns. Thus the rightmost element of this matrix in the fourth row from the bottom shows $p(\tau^- \to j)$, the number of negatively charged tau leptons (out of 10^5) reconstructed as a jet.

	e^+	e^-	μ^+	μ^-	τ^+	τ^-	γ	j	b
e^+	3737	26	0	0	1	0	71307	24597	0
e^-	24	3834	0	0	0	4	71003	24789	0
μ^+	0	0	10661	0	3	0	1	1061	0
μ^-	0	0	0	10678	0	2	3	1127	0
γ	55	65	0	0	0	0	76374	23064	0
π^0	46	53	0	0	0	1	74111	25522	0
π^+	17	0	16	0	4395	2	554	93462	25
π^-	1	24	0	10	3	4206	673	93570	20
K^+	13	0	59	0	4658	0	421	92807	12
K^-	0	36	0	38	1	4301	834	92958	23
B^+	50	2	102	3	4	0	186	90077	7389
B^-	3	18	2	81	0	13	160	90178	7347
B^0	52	12	96	15	3	8	148	90241	7016
$\bar{B}^0$	10	52	8	107	4	5	126	90149	7095
D^+	32	4	90	0	136	11	738	94326	2148
D^-	2	22	1	57	6	127	817	94367	2100
D^0	9	7	38	2	20	74	671	96983	1027
$\bar{D}^0$	2	15	3	37	74	17	628	96928	1121
K^0_L	1	0	0	0	6	8	1089	97411	6
K^0_S	2	3	0	0	11	39	9532	56689	0
τ^+	339	8	1249	1	341	2	3198	66243	104
τ^-	5	346	0	1226	0	336	3208	66111	108
u	19	12	2	0	73	13	423	99359	47
d	13	17	0	0	15	36	359	99357	60
g	7	11	8	1	19	18	41	98937	426

Table B.11: Plug single particle misidentification matrix. Using a single particle gun, 10^5 particles of each type shown at the left of the table are shot with $p_T = 25$ GeV into the plug CDF detector, uniformly distributed in θ and in ϕ. The resulting reconstructed object types are shown at the top of the table, labeling the table columns. Thus the rightmost element of this matrix in the fourth row from the bottom shows $p(\tau^- \to j)$, the number of negatively charged tau leptons (out of 10^5) reconstructed as a jet.
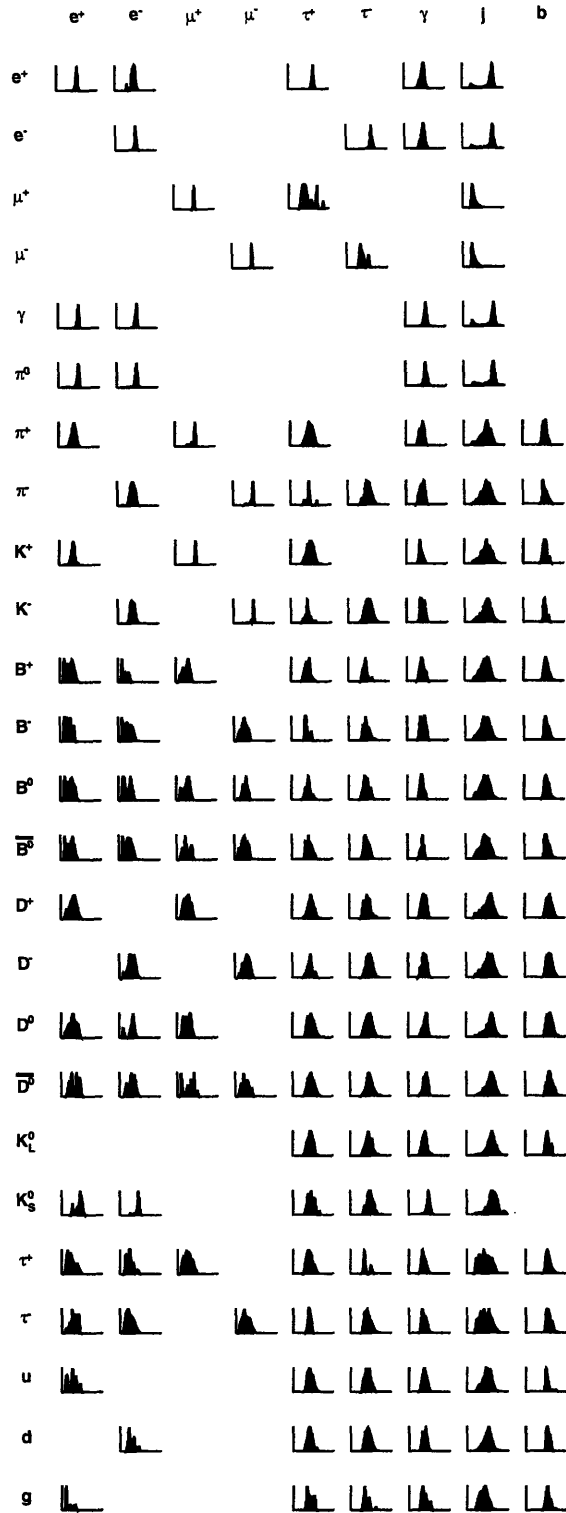
Figure B-1: Transverse momentum distribution of reconstructed objects (labelling columns) arising from single particles (labelling rows) with $p_T = 25$ GeV shot from a single particle gun into the central CDF detector. The area under each histogram is equal to the number of events in the corresponding misidentification matrix element of Table B.10; histograms with fewer than ten events are not shown. The horizontal axis ranges from 0 to 50 GeV, with one tick mark each 5 GeV. The incident single particle distribution is a delta function at the center of each plot, at $p_T = 25$ GeV.
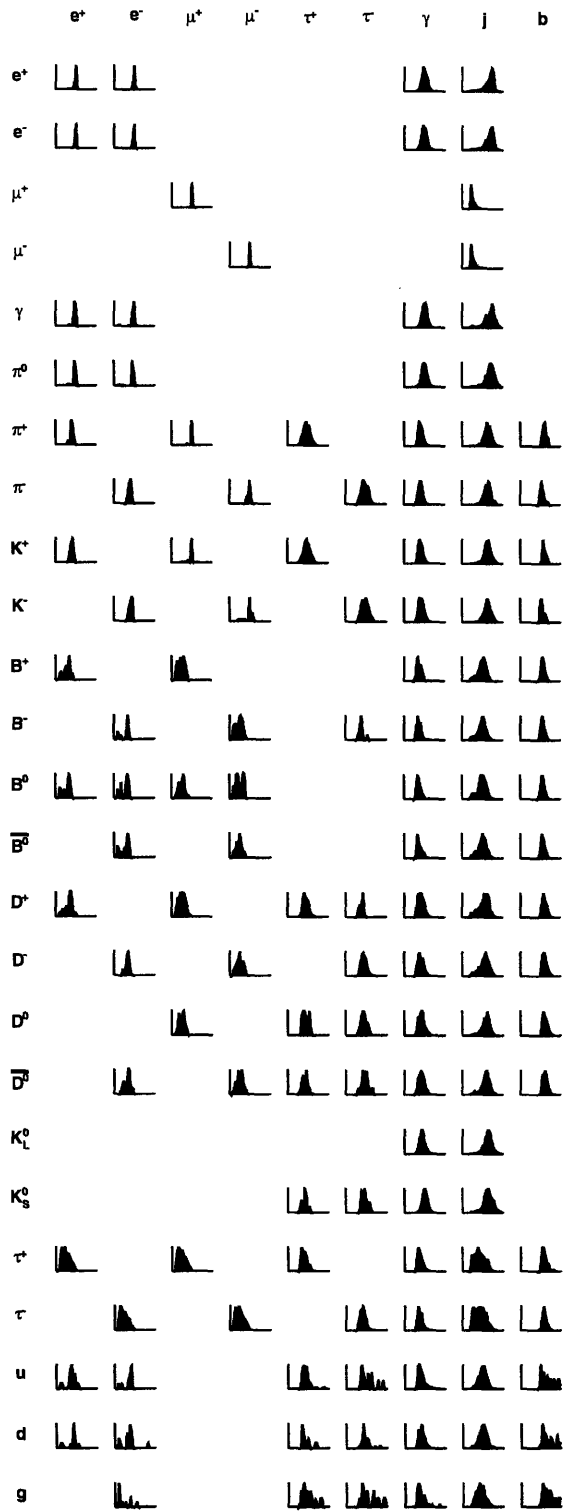
Figure B-2: Transverse momentum distribution of reconstructed objects (labelling columns) arising from single particles (labelling rows) with $p_T = 25$ GeV shot from a single particle gun into the plug CDF detector. The area under each histogram is equal to the number of events in the corresponding misidentification matrix element of Table B.11; histograms with fewer than ten events are not shown. The horizontal axis ranges from 0 to 50 GeV, with one tick mark each 5 GeV. The incident single particle distribution is a delta function at the center of each plot, at $p_T = 25$ GeV.

	e^+	e^-	μ^+	μ^-	τ^+	τ^-	γ	j	b
e^+	6060	2	0	0	38	0	139	3576	1
e^-	0	6103	0	0	0	28	128	3574	0
μ^+	1	0	5217	0	2	0	0	289	0
μ^-	0	1	0	5201	0	2	1	301	0
γ	55	75	0	0	0	0	6554	3118	0
π^0	42	38	0	0	0	0	5751	3991	0
π^+	19	0	9	0	7721	0	2	2089	19
π^-	0	15	0	4	3	7761	2	2100	9
K^+	10	0	20	0	7662	2	2	2109	5
K^-	0	25	0	11	5	7682	3	2119	10
B^+	18	2	6	0	13	1	0	5160	4792
B^-	1	9	0	6	1	10	1	5186	4776
B^0	13	3	5	0	13	9	0	5111	4836
$\bar{B}^0$	0	11	0	3	7	7	0	5095	4868
D^+	41	1	20	0	726	10	14	7861	1241
D^-	2	52	1	11	10	696	10	7898	1247
D^0	9	4	7	2	31	93	25	9036	767
$\bar{D}^0$	3	11	0	9	87	39	24	9081	693
K^0_L	0	0	0	0	7	7	13	9849	16
K^0_S	0	1	0	0	18	69	810	5855	1
τ^+	889	2	711	0	1506	1	63	5008	191
τ^-	3	816	0	672	1	1496	107	5088	188
u	1	1	0	0	46	2	19	9923	31
d	1	0	0	0	3	35	7	9944	24
g	1	1	0	2	3	3	0	9897	174

Table B.12: Central single particle misidentification matrix. Using a single particle gun, 10^4 particles of each type shown at the left of the table are shot with $p_T = 50$ GeV into the central CDF detector, uniformly distributed in θ and in ϕ. The resulting reconstructed object types are shown at the top of the table, labeling the table columns.

B.4 Fake Rates

It would take too many Monte Carlo events to acquire enough statistics of rare fake processes. To overcome this difficulty, we apply our own multiplicative fake rates on reconstructed objects, when they are reconstructed more often than the objects thay may fake. Specifically, we apply fake rates for jets or b-tagged jets faking electrons, muons, photons, τs, jets faking b-tagged jets, and photons faking electrons. Note that other fake processes are not neglected – they are handled by CDFSim. In the interest of simplicity, we try to keep our fake rates as simple as possible. There is generally one overall coefficient for the fake rate, and this value is usually obtained from the VISTA fit to the data. In some cases however, to better model the true fake process, we need to introduce additional modulations as a function of p_T or location within the detector (η or ϕ). This section details all the special modulations applied for VISTA fake rates. Generally, we show a modulating function, which multiplies the appropriate correction factor value to obtain the true fake rate applied. If not shown here, the fake rate is treated as being constant.

Figures B-4 and B-5 show the relative fake rate for jets to fake electrons as a function of η_{det} and ϕ. These functions of η_{det} and ϕ are multiplied by overall correction factors which represent a crude average fake rate over the appropriate region. These shaped functions are meant to model more fine details in fake rates than the overall average can contain. In addition to η_{det} and ϕ dependence, for plug electrons there is a dependance on the p_T, shown in Figure B-3. Figures B-6, B-7, and B-8 show the electron p_T, electron η_{det} and ϕ distribution from data in the 1e+1j final state, where almost all events come from QCD dijet production where one of the jets fakes an electron. This serves as the dominant control region for determining variations in jet to electron fake rate.

Figures B-9 and B-10 show the fake rate variation for jets to fake muons as a function of p_T and η_{det}. The fake rate is higher in CMX than in CMU and CMP. The muon p_T, η_{det}, and ϕ distributions in the 1j1mu+ final state are shown in Fig. B-11, B-12, and B-13. These serve as the dominant control regions determining these fake
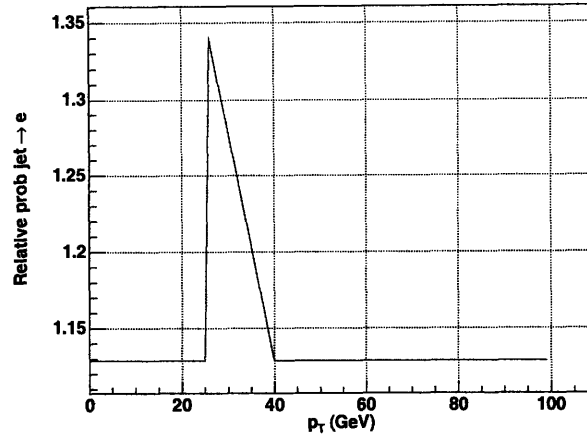
Figure B-3: The relative fake rate for jets to fake electrons in the plug as a function of the p_T of the jet
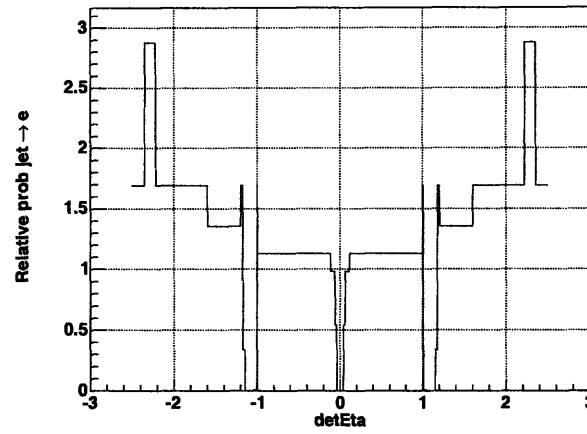
Figure B-4: The relative fake rate for jets to fake electrons as a function of detEta.
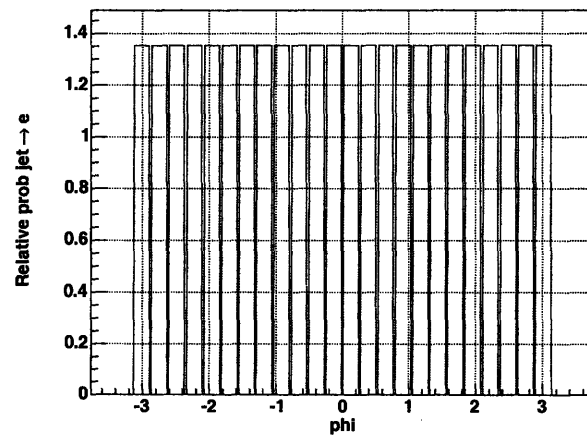
Figure B-5: The relative fake rate for jets to fake electrons as a function of phi.
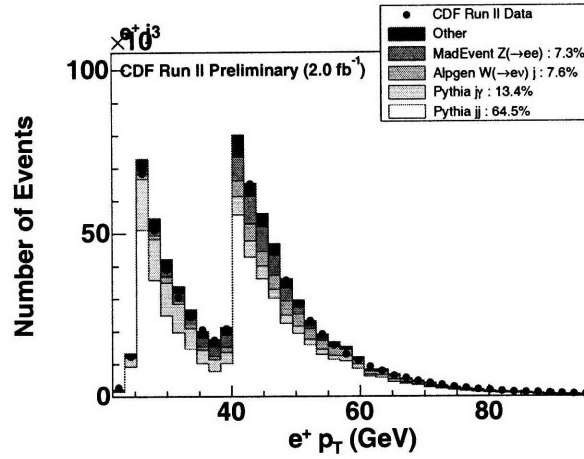
Figure B-6: Electron p_T distribution in the 1e+1j final state.
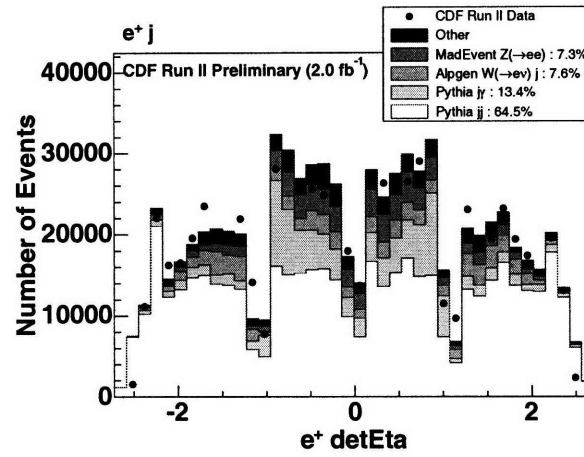
Figure B-7: Electron detector eta distribution in the 1e+1j final state.
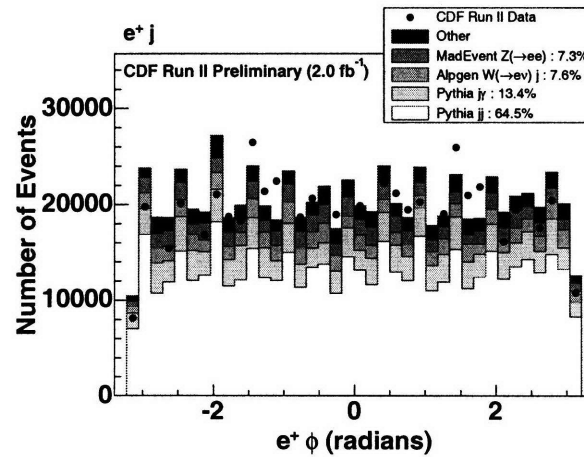
Figure B-8: Electron phi distribution in the 1e+1j final state.
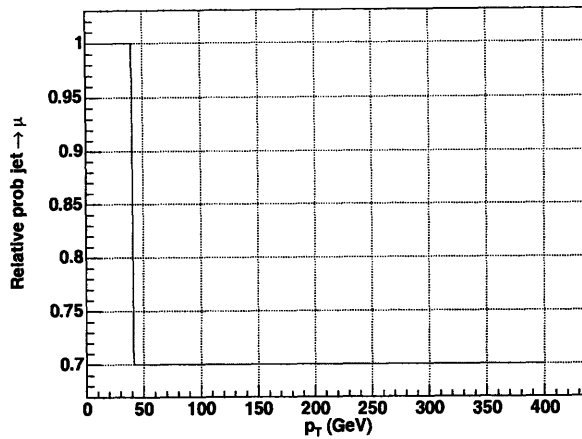
Figure B-9: The relative fake rate for jets to fake muons as a function of p_T.
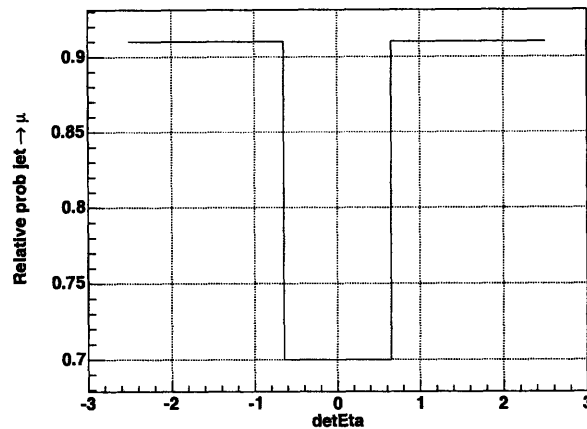
Figure B-10: The relative fake rate for jets to fake muons as a function of η_{det}.

rates.

Figures B-14, B-15, and B-16 show the jet to photon fake rates as functions of p_T, η_{det}, and ϕ. Detector geometry features are analogous to those exhibited in the jet to electron fake rate. The photon p_T, η_{det}, and ϕ distributions in the 1j1ph final state are shown in Fig. B-17, B-18, and B-19. This is one of the dominant control regions determining the jet to photon fake rates. Unlike the previous two cases, this final state is dominated by real γ+jet production, rather than the fake process, which contributes about 35% to this final state.

The variation in jet faking b-jet rate is shown in B-20, as a function of p_T. This shape is consistent with the one measured by the b-tagging group. Before comparing absolute values, however, it should be noted that this VISTA fake rate includes contributions from charm quarks to fake b, which is not usually included in the b-tagging
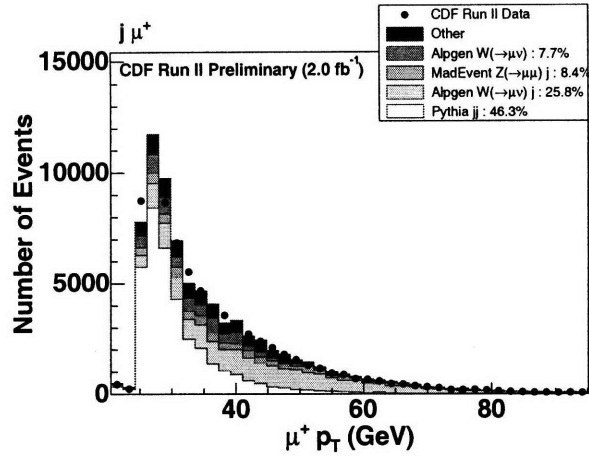
Figure B-11: Muon p_T distribution in the 1j1mu+ final state.
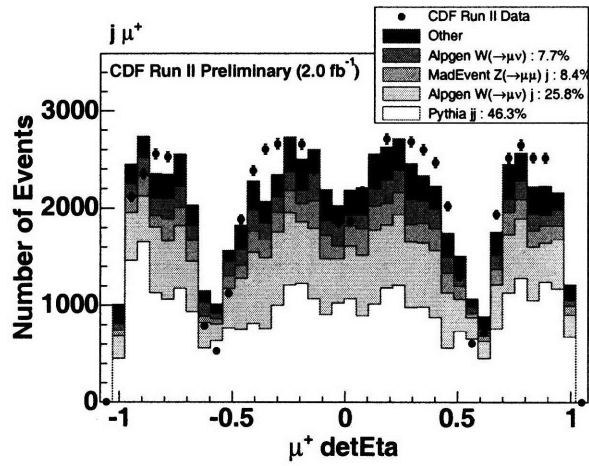
Figure B-12: Muon η_{det} distribution in the 1j1mu+ final state.
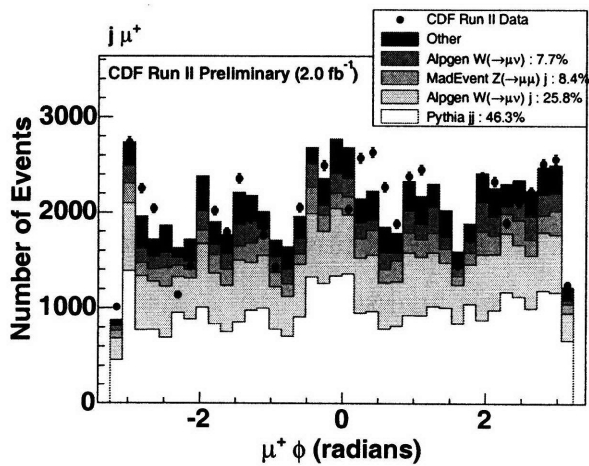
Figure B-13: Muon ϕ distribution in the 1j1mu+ final state.
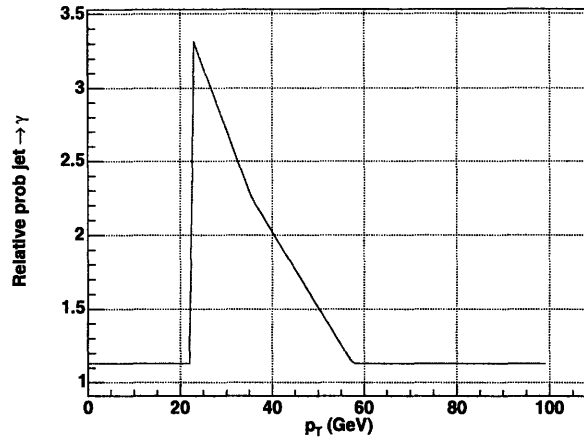
Figure B-14: The relative fake rate for jets to fake photons as a function of p_T.
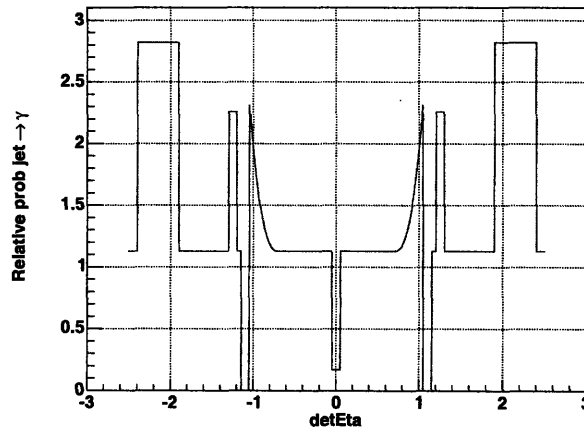
Figure B-15: The relative fake rate for jets to fake photons as a function of η_{det}.
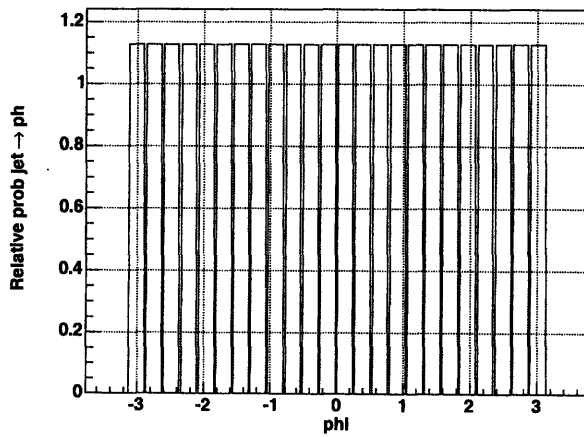
Figure B-16: The relative fake rate for jets to fake photons as a function of ϕ.

209
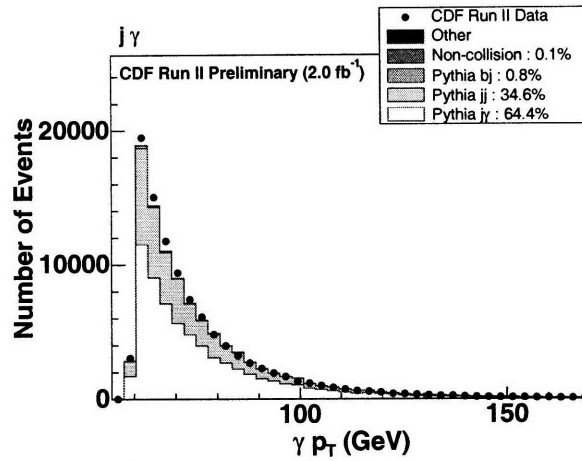
Figure B-17: Photon p_T distribution in the 1j1ph final state.
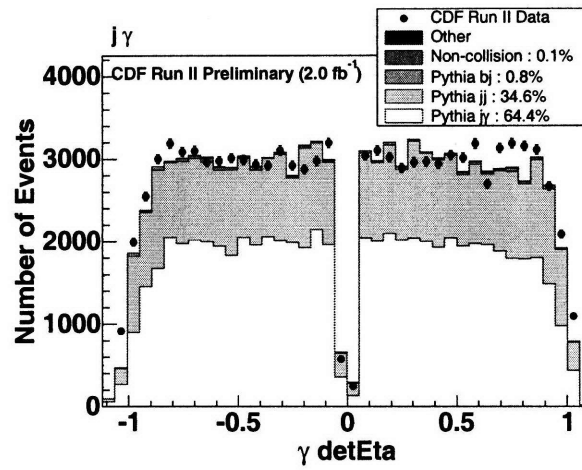
Figure B-18: Photon η_{det} distribution in the 1j1ph final state.
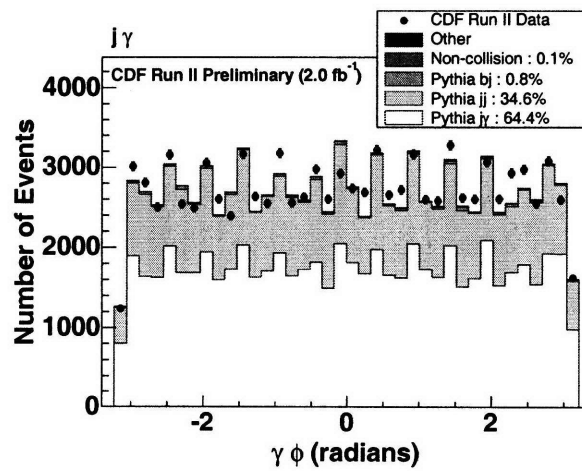
Figure B-19: Photon ϕ distribution in the 1j1ph final state.
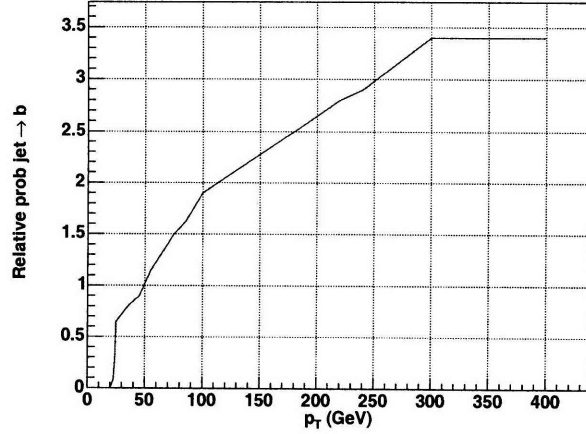
Figure B-20: The relative fake rate for jets to fake $b - tagged$ jets as a function of p_T. It is essentially the mistag rate.
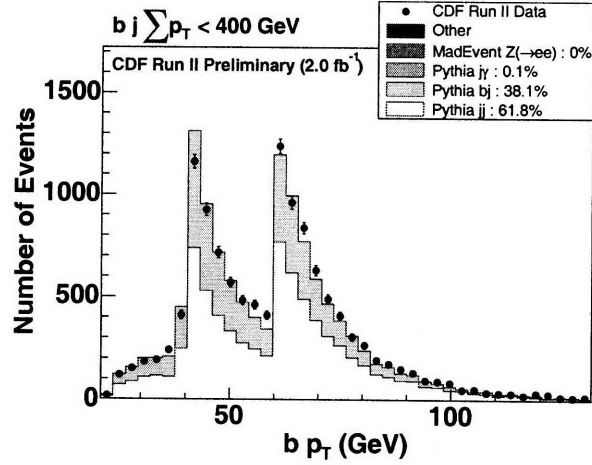
Figure B-21: The b-jet p_T distribution in the 1b1j low $\sum p_T$ final state.

mistag rate. When we accounted for the expected relative contribution of charmed quarks in our 'denominator jets', we found values consistent with the mistag rates. The b jet p_T distribution is shown in Fig. B-21 and B-22, for the 1b1j high $\sum p_T$ and 1b1j low $\sum p_T$ final states. These are the dominant control regions determining the mistag rates.

The jet to τ relative fake rate is given in Fig. B-23. This shape is then multiplied by the function $\exp(-\text{GeneratedSumPt}/350\ \text{GeV})$ and the jet to τ fake rate correction factor to obtain the final fake rate. The shape is consistent with previous studies of the jet to τ fake rate. The τ p_T distributions in the 1j1tau+ low-$\sum p_T$, 1j1tau+ high-$\sum p_T$, and 1tau+1tau- final states are shown in Fig. B-24, B-25, and B-26. These serve as the dominant control regions determining the jet to τ fake rate.
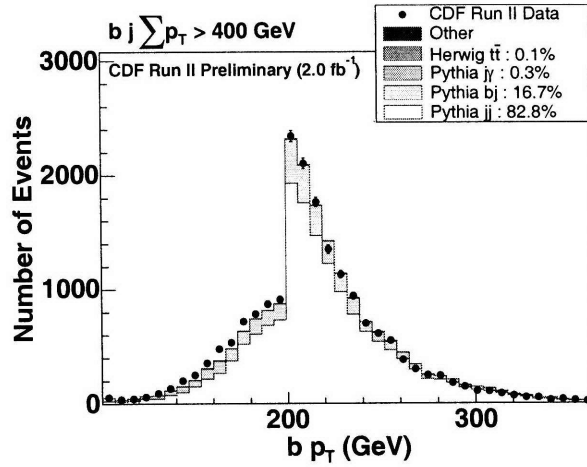
Figure B-22: The b-jet p_T distribution in the 1b1j high-$\sum p_T$ final state.
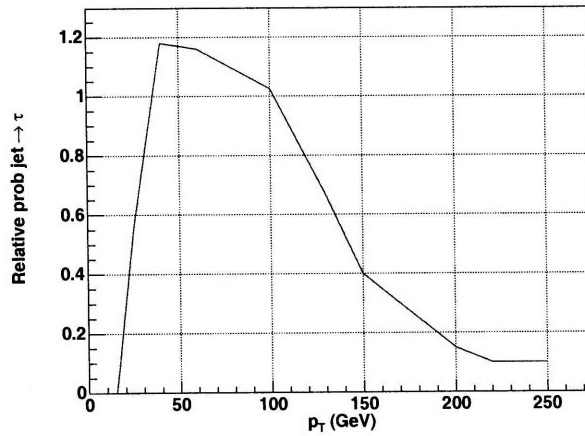
Figure B-23: The relative fake rate for jets to fake τs as a function of p_T.
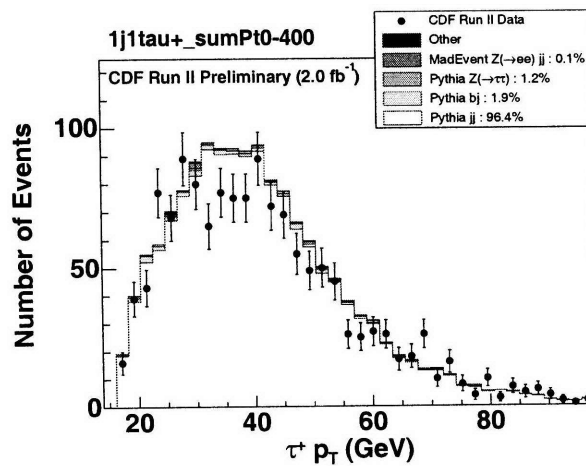
Figure B-24: The τ p_T distribution in the 1j1tau+ low-$\sum p_T$ final state.
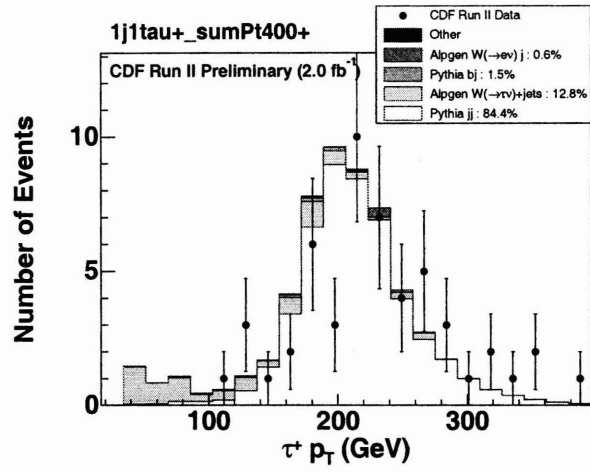
Figure B-25: The τ p_T distribution in the `1j1tau+` high-$\sum p_T$ final state.
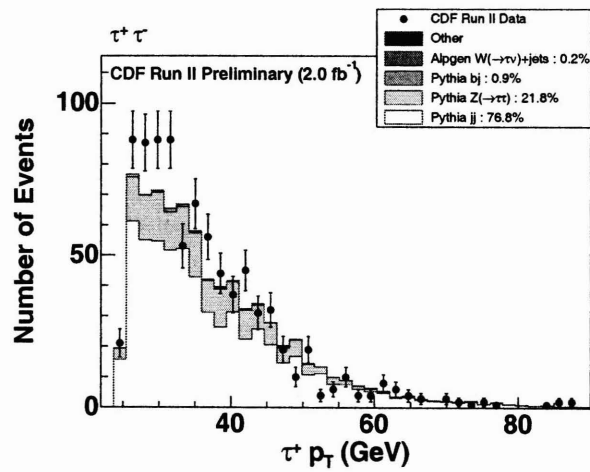
Figure B-26: The τ p_T distribution in the `1tau+1tau-` final state.
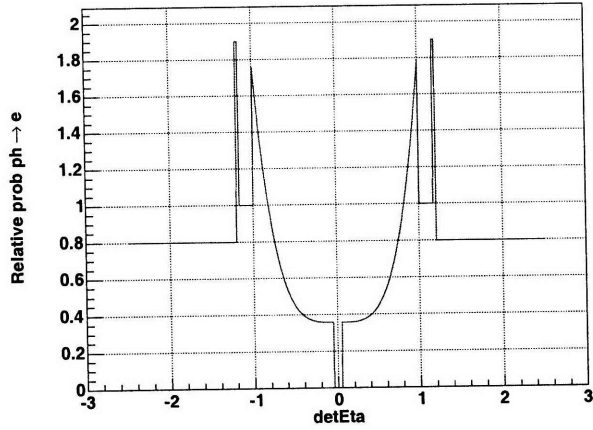
Figure B-27: The relative fake rate for jets to fake τs as a function of p_T.
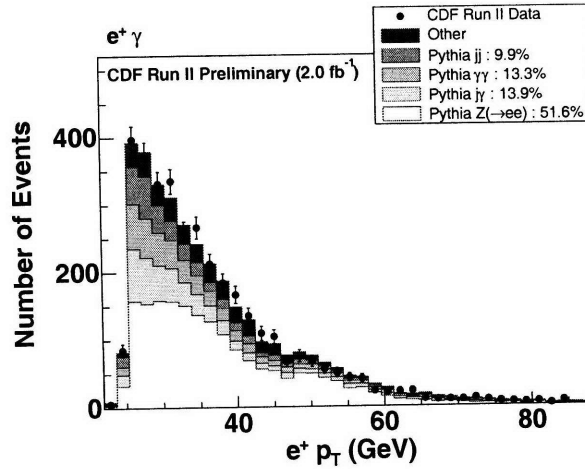
Figure B-28: The electron p_T distribution in the 1e+1ph final state.

Figure B-27 shows the relative fake rate for photons to fake electrons as a function of $\eta_{\rm det}$. Fig. B-28 and B-29 show the electron p_T and $\eta_{\rm det}$ distributions in the 1e+1ph final state. This final state is the dominant control region determining the photon to electron fake rate. However, this underlying process does not contribute very much to the background in this final state and, as a result, the photon to electron fake rate is not as well constrained as other fake rates. Fig. B-30 and B-31 show the photon p_T and $\eta_{\rm det}$ distributions in this same final state. As a general comment, this final state is a particularly good example of how well-modelled our fake backgrounds are, since the background contributing to this final state is a mixture of various different fake processes.

214
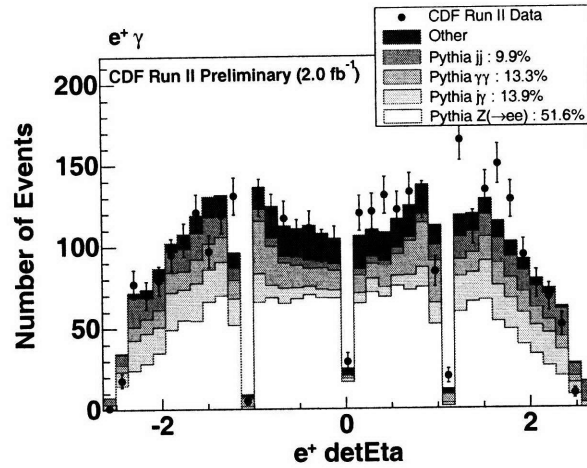
Figure B-29: The electron η_{det} distribution in the 1e+1ph final state.
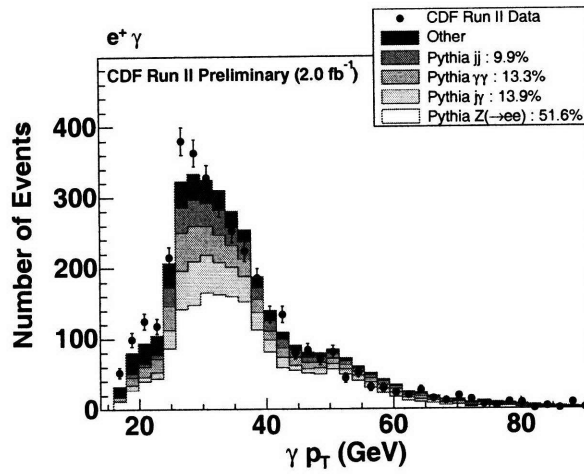
Figure B-30: The photon p_T distribution in the 1e+1ph final state.
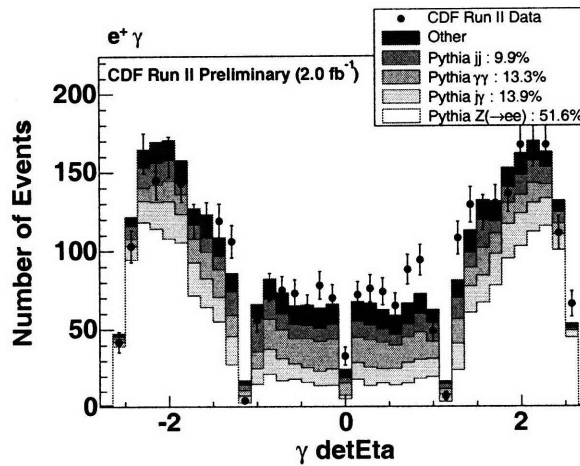
Figure B-31: The photon η_{det} distribution in the 1e+1ph final state.

Code	Category	Explanation	Before	After	deviation (σ)	Change(%)
5001	luminosity	CDF integrated luminosity	0.927 ± 0.02	2 ± 0.0608	53.4	115.3
5102	k-factor	cosmic_ph	0.69 ± 0.05	0.81 ± 0.05	2.5	18.4
5103	k-factor	cosmic_j	0.45 ± 0.014	0.19 ± 0.014	-18.2	-57.1
5121	k-factor	$1\gamma1j$ photon+jet(s)	0.95 ± 0.04	0.92 ± 0.04	-0.7	-2.9
5122	k-factor	$1\gamma2j$	1.2 ± 0.05	1.3 ± 0.05	1.3	5.3
5123	k-factor	$1\gamma3j$	1.5 ± 0.07	1.6 ± 0.07	1.5	7.0
5124	k-factor	$1\gamma4j+$	2 ± 0.16	1.9 ± 0.14	-0.5	-3.9
5130	k-factor	$2\gamma0j$ diphoton(+jets)	1.8 ± 0.08	1.6 ± 0.07	-2.4	-10.5
5131	k-factor	$2\gamma1j$	3.4 ± 0.24	3 ± 0.17	-1.8	-12.8
5132	k-factor	$2\gamma2j+$	1.3 ± 0.16	1.2 ± 0.09	-0.6	-7.7
5141	k-factor	W0j W (+jets)	1.5 ± 0.027	1.4 ± 0.04	-2.8	-5.2
5142	k-factor	W1j	1.1 ± 0.03	1.3 ± 0.04	9.1	25.7
5143	k-factor	W2j	1 ± 0.03	2 ± 0.06	32.0	94.1
5144	k-factor	W3j+	0.76 ± 0.05	2.1 ± 0.07	26.9	177.4
5151	k-factor	Z0j Z (+jets)	1.4 ± 0.024	1.4 ± 0.03	-1.3	-2.2
5152	k-factor	Z1j	1.2 ± 0.04	1.2 ± 0.04	1.3	4.5
5153	k-factor	Z2j+	1 ± 0.05	1 ± 0.04	-0.3	-1.5
5161	k-factor	2j $\hat{p}_T$<150 dijet	0.96 ± 0.022	1 ± 0.031	1.9	4.4
5162	k-factor	2j 150<$\hat{p}_T$	1.3 ± 0.028	1.3 ± 0.04	2.9	6.5
5164	k-factor	3j $\hat{p}_T$<150 multijet	0.92 ± 0.021	0.94 ± 0.03	1.0	2.3
5165	k-factor	3j 150<$\hat{p}_T$	1.4 ± 0.032	1.5 ± 0.05	3.7	8.7
5167	k-factor	4j $\hat{p}_T$<150	0.99 ± 0.025	1.1 ± 0.04	3.0	7.7
5168	k-factor	4j 150<$\hat{p}_T$	1.7 ± 0.04	1.9 ± 0.07	5.5	12.8
5169	k-factor	5j low	1.3 ± 0.05	1.3 ± 0.06	1.7	6.8
5170	k-factor	1b2j 150<$\hat{p}_T$ heavyflavor multijet	NA ± NA	2.2 ± 0.12	NA	NA
5171	k-factor	1b3j 150<$\hat{p}_T$	NA ± NA	3 ± 0.16	NA	NA
5211	misId	p(e→e) central	0.99 ± 0.006	0.98 ± 0.007	-1.5	-0.9
5212	misId	p(e→e) plug	0.93 ± 0.009	0.97 ± 0.007	3.6	3.5
5213	misId	p(μ→μ) CMUP+CMX	0.85 ± 0.008	0.89 ± 0.007	5.3	5.0
5216	misId	p(γ→γ) central	0.97 ± 0.018	0.95 ± 0.013	-1.6	-2.9
5217	misId	p(γ→γ) plug	0.91 ± 0.018	0.85 ± 0.007	-3.2	-6.4
5219	misId	p(b→b) central	1 ± 0.04	0.97 ± 0.02	-0.8	-3.2
5246	misId	p(γ→e) plug	NA ± NA	0.062 ± 0.0021	NA	NA
5256	misId	p(q→e) central	9.71×10^{-5} ± 1.9×10^{-6}	7.077×10^{-5} ± 1×10^{-6}	-13.9	-27.1
5257	misId	p(q→e) plug	0.0008761 ± 1.8×10^{-5}	0.0007611 ± 5×10^{-6}	-6.4	-13.1
5261	misId	p(q→μ)	1.157×10^{-5} ± 2.7×10^{-7}	1.235×10^{-5} ± 5×10^{-7}	2.9	6.7
5266	misId	p(b→μ)	NA ± NA	3.522×10^{-5} ± 1.1×10^{-5}	NA	NA
5273	misId	p(j→b) 25<p_T	0.0168 ± 0.00027	0.0183 ± 0.00018	5.4	8.7
5285	misId	p(q→τ)	0.0034 ± 0.00012	0.0052 ± 8×10^{-5}	14.9	52.5
5292	misId	p(q→γ) central	0.0002651 ± 1.5×10^{-5}	0.0002611 ± 1.2×10^{-5}	-0.3	-1.5
5293	misId	p(q→γ) plug	0.00159 ± 0.00013	0.000478 ± 4×10^{-5}	-8.6	-70.0
5402	trigger	p(e→trig) plug, p_T>25	0.83 ± 0.015	0.86 ± 7×10^{-5}	1.8	3.2
5403	trigger	p(μ→trig) CMUP+CMX, p_T>25	0.917 ± 0.007	0.918 ± 0.004	0.2	0.1

Table B.13: Comparison of correction factors that were used also in the first 0.927 fb^{-1}. The Luminosity is in units of fb^{-1}.

B.5 Correction Factors

B.5.1 Comparison with first round

The correction factor values obtained in the second round (v02) (corresponding to 2 fb^{-1}) are here compared with the correction factor values obtained in the first round (v01) (corresponding to 927 pbb^{-1}). The numerical values can be found in Table B.13; analysis of the changes is provided below.

5001. The integrated luminosity of the sample has of course increased with respect to v01. The present integrated luminosity obtained from the fit is again consistent with the luminosity obtained from the CLC measurement.
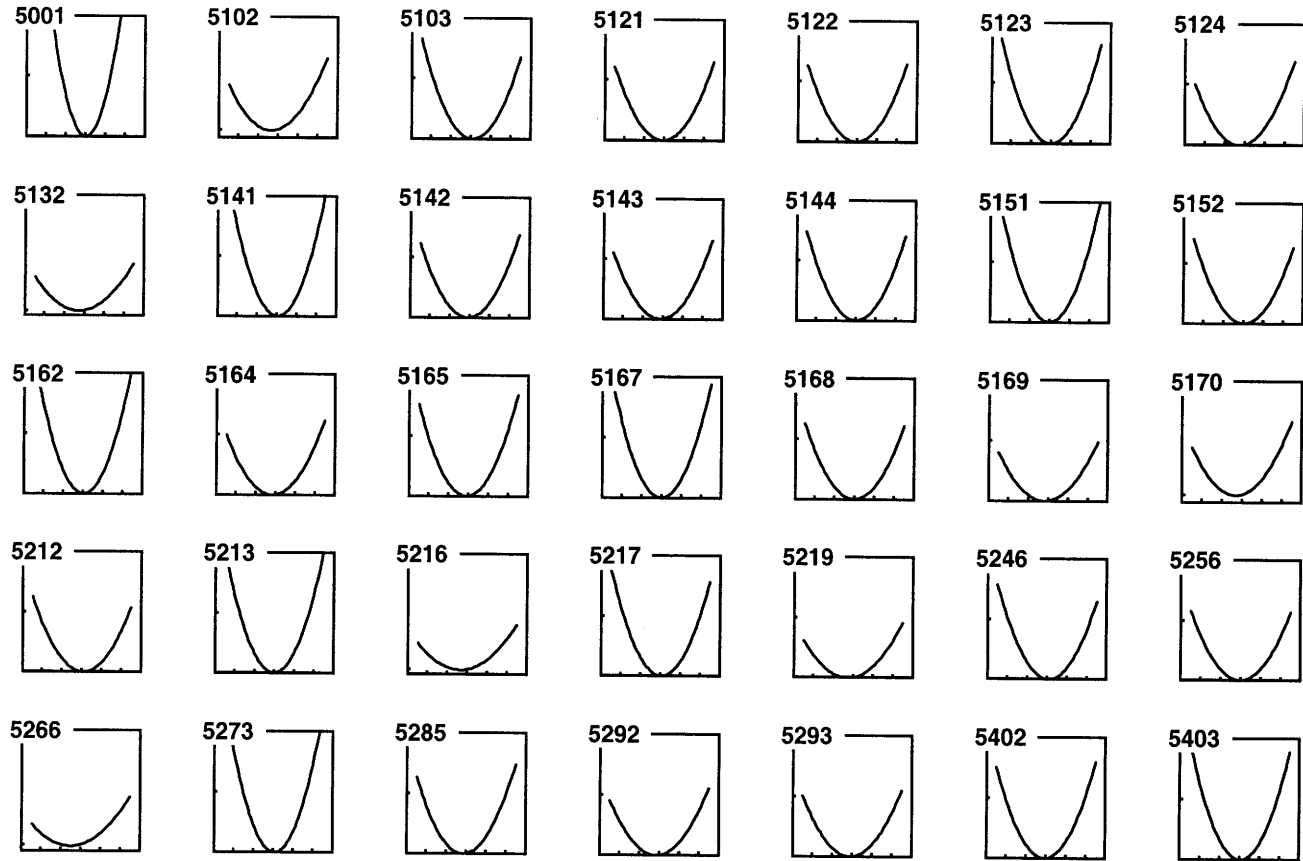
Figure B-32: Profiles of the χ^2 function at its minimum along each correction factor axis. This ar debugging tool to validate the parabolic form of the minimum and the calculation of the error matri: χ^2 as a function of integrated luminosity (correction factor code 0001), holding all remaining correcti panes show χ^2 profiles along each of the other correction factors. One tick on the horizontal axis of t δs_i, the obtained error on the correction factor value. One tick on the vertical axis corresponds to o

Code	Pull Apart	Contributions
0042	10.3	(e6pmiss = -3.2 , e6jjpmiss = 2 , e6jj = 1.3 , e6j = -1.1)
0037	10.2	(jtau2 = -1 , ph0tau = 0.9 , jtau1 = -0.9 , ph6tau = 0.8)
0036	9.2	(bj5 = 1.5 , b5j = -1.3 , be0 = 0.8 , b5jj = -0.5)
0013	8.9	(e6jjpmiss = 2.9 , e0jjpmiss = -1.1 , jjmu0pmiss = -0.7 , jjmu0 = -0.5)
0031	8.8	(bbj = -0.9 , bj5 = 0.7 , be0 = 0.5 , bbjj = -0.5)
0034	8.5	(e6ph6 = -2.6 , e6jj = 0.8 , e6ph0 = 0.8 , e6j = -0.7)
0001	8.4	(e0j = -0.5 , e6pmiss = -0.4 , e0pmiss = 0.4 , e6jjpmiss = 0.4)
0033	8.2	(e0j = -4.7 , e0jj = 1.3 , be0 = 1 , e0jjj = 0.4)
0018	7.7	(jj = 1.4 , e0j = -1.3 , e6j = -0.9 , bj = 0.6)
0034	7.2	(e6jj = 2.3 , e6j = -1.6 , be6 = -0.9 , bej = 0.4)
0014	6.9	(e0jjjpmiss = -1 , jjjmu0 = -1 , e6jjjpmiss = 1 , jjjmu0pmiss = -0.6)
0020	6.2	(jjj = -2.1 , e6jj = 1.3 , e0jj = 0.5 , bej = 0.3)
0004	6	(jph0 = 2.2 , e0j = -1.4 , bph0 = -0.6 , be0 = 0.5)
0026	5.7	(e6pmiss = -1.3 , e6e6 = 1.3 , e6jjpmiss = 1 , e6e6j = -0.6)
0012	5.3	(e6jjpmiss = 0.7 , jmu0pmiss = 0.6 , be0pmiss = -0.6 , e0jpmiss = 0.5)
0016	5.1	(e6e6j = -1.5 , jmu0mu0 = 0.5 , constraints = 0.4 , e0jj = 0.27)
0030	5	(constraints = 1.4 , e6ph6 = -1 , mu0ph6pmiss = -0.4 , ph6tau = 0.27)
0017	4.9	(e6e6jj = 0.5 , jjjmu0 = -0.4 , e6e6jjj = -0.27 , e6e6j = -0.24)
0029	4.8	(jph0 = 0.8 , constraints = 0.5 , jjph0 = -0.5 , bph0 = -0.4)
0005	4.5	(jjph0 = -2.3 , e0jj = 0.7 , bjph0 = -0.3 , e6jj = 0.24)
0040	4	(ph6tau = 1.3 , e6ph6 = -0.9 , ph0ph6 = -0.3 , j5ph6 = -0.28)
0038	4	(bmu0 = 0.9 , jjmu0 = -0.8 , jjjmu0 = -0.7 , bjjmu = -0.5)
0039	3.7	(jph0 = 1.5 , jjph0 = -0.5 , bph0 = -0.4 , e6ph0 = 0.22)
0025	3.5	(e0pmiss = 0.9 , e0e0 = -0.8 , e0j = -0.2 , e0jjpmiss = -0.19)
0015	3.3	(e0e0 = -0.7 , e6e6 = 0.6 , e0e6 = -0.4 , constraints = 0.3)
0006	3.1	(jjjjph0 = -1.9 , e0jjj = 0.6 , bjjph0 = 0.16)
0007	3.1	(jjjjph0 = -2.1 , e0jjjj = 0.6 , e6jjjj = -0.13)
0022	3	(bjjj = 0.6 , e6jjj = -0.4 , e0jjj = 0.3 , jjjmu0 = -0.3)
0019	2.9	(bj5 = 1.3 , b5j = -1 , bb5 = -0.14 , jj5 = -0.13)
0035	2.9	(jmu0 = 1.1 , jjmu0 = -0.6 , jjjmu0 = -0.5 , bmu0 = 0.4)
0010	2.6	(jjjph ph = -0.9 , jjph ph = 0.4 , e0jjph6 = 0.23 , e6jjjph6 = 0.17)
0021	2.2	(jjj5 = 1.2 , b5jj = -0.6 , jjj5ph0 = -0.12)
0024	2.1	(e6jjjj = -0.5 , e0jjjj = 0.5 , jjjjj = -0.4 , bjjjj = 0.24)
0011	2	(e0pmiss = 0.6 , e6pmiss = -0.6 , mu0pmiss = -0.25 , constraints = -0.19)
0027	2	(mu0pmiss = -0.5 , jmu0pmiss = 0.16 , jjmu0pmiss = -0.13 , jjjmu0 = -0.12)
0043	1.8	(mu0pmiss = -0.7 , constraints = 0.17 , jmu0pmiss = 0.16 , jjjmu0 = -0.16)
0025	1.7	(b5jj = -0.7 , bjj5 = 0.31 , bb5j = 0.29 , jjj5 = 0.15)
0008	1.6	(ph0ph6 = -0.5 , e6ph0 = 0.3 , constraints = 0.24 , ph0ph0 = -0.22)
0026	1.2	(bbjj5 = -0.5 , b5jjj = 0.29 , bb5jj = 0.25)
0023	1.1	(jjjj5 = -0.6 , b5jjj = 0.3)
0002	0.7	(j5ph0 = -0.11)
0009	0.6	(e6jph0 = 0.19 , constraints = 0.16 , e0jph0 = -0.11)

Table B.14: Correction factor pull apart table, intended to show which correction factors are being pulled in different directions. Letting χ_k^2 denote the kth term in the χ^2 sum, and s_i the ith correction factor, the *pull* of the kth bin on the ith correction factor is denoted pull_{ki}. Intuitively, bin k "pulls" on the ith correction factor with a strength of pull_{ki}. More precisely, the value obtained by the ith correction factor is pull_{ki} standard deviations away from where it would be in the absence of the kth bin. If bin k pulls the ith correction factor toward larger values, pull_{ki} is positive; if bin k favors smaller values of the ith correction factor, pull_{ki} is negative. The units of pull_{ki} are units of χ^2. The correction factors are sorted in order of decreasing *pull apart*, where the pull apart of the ith correction factor is defined as $\text{pullApart}_i \equiv \sum_k |\text{pull}_{ki}|$, provided in the second column. Intuitively, a correction factor has large pull apart if some bins strongly favor a larger value, and some bins strongly favor a smaller value. In the third column between parentheses are the bins k that contribute most to the pull apart of each correction factor, along with each individual contribution pull_{ki}. In each line, only the four largest contributions with pull ≥ 0.1 are listed. In the bin labels, a 0 following a particle specifies its centrality; a 4 following a particle indicates it has $p_T > 200$ GeV; a 5 following mu indicates it is a CMX muon in the region $0.6 < |\eta| < 1.0$; a 10 following a particle indicates it lies in the plug region $1 < |\eta| < 2.5$; constraints specifies the contribution from $\chi^2_{\text{constraints}}$.

Bin	Total Influence	IndividualInfluence
e0j	8.9	(0033 = -4.7 , 0004 = -1.4 , 0018 = -1.3 , 0001 = -0.5)
e6jjpmiss	8.1	(0013 = 2.9 , 0042 = 2 , 0026 = 1 , 0012 = 0.7)
e6jj	6.6	(0034 = 2.3 , 0042 = 1.3 , 0020 = 1.3 , 0034 = 0.8)
e6pmiss	6.1	(0042 = -3.2 , 0026 = -1.3 , 0011 = -0.6 , 0034 = -0.5)
e6ph6	5.5	(0034 = -2.6 , 0030 = -1 , 0040 = -0.9 , 0034 = -0.21)
jph0	5.2	(0004 = 2.2 , 0039 = 1.5 , 0029 = 0.8 , 0018 = 0.4)
e6j	5	(0034 = -1.6 , 0042 = -1.1 , 0018 = -0.9 , 0034 = -0.7)
constraints	4.7	(0030 = 1.4 , 0029 = 0.5 , 0016 = 0.4 , 0015 = 0.3)
bj5	3.9	(0036 = 1.5 , 0019 = 1.3 , 0031 = 0.7 , 0001 = 0.28)
be0	3.8	(0033 = 1 , 0036 = 0.8 , 0031 = 0.5 , 0004 = 0.5)
jjph0	3.7	(0005 = -2.3 , 0039 = -0.5 , 0029 = -0.5 , 0020 = -0.2)
jjjmu0	3.7	(0014 = -1 , 0038 = -0.7 , 0035 = -0.5 , 0017 = -0.4)
e0jj	3.6	(0033 = 1.3 , 0005 = 0.7 , 0020 = 0.5 , 0016 = 0.27)
be6	3.3	(0034 = -0.9 , 0042 = -0.6 , 0018 = -0.5 , 0036 = -0.4)
jjmu0	2.9	(0038 = -0.8 , 0035 = -0.6 , 0013 = -0.5 , 0020 = -0.2)
b5j	2.8	(0036 = -1.3 , 0019 = -1 , 0031 = -0.3 , 0001 = -0.2)
ph6tau	2.8	(0040 = 1.3 , 0037 = 0.8 , 0030 = 0.27 , 0042 = 0.2)
e6e6j	2.5	(0016 = -1.5 , 0026 = -0.6 , 0017 = -0.24 , 0001 = -0.1)
jjjph0	2.5	(0006 = -1.9 , 0029 = -0.2 , 0039 = -0.17 , 0022 = -0.12)
bph0	2.4	(0004 = -0.6 , 0036 = -0.5 , 0039 = -0.4 , 0029 = -0.4)
jjjjph0	2.4	(0007 = -2.1)
jjj	2.4	(0020 = -2.1 , 0001 = -0.24)
jmu0	2.3	(0035 = 1.1 , 0038 = 0.5 , 0012 = 0.2 , 0018 = 0.17)
e6jjjpmiss	2.3	(0014 = 1 , 0042 = 0.4 , 0013 = 0.4 , 0026 = 0.21)
bej	2.3	(0034 = 0.4 , 0020 = 0.3 , 0031 = 0.32 , 0036 = 0.3)
e0jjjj	2.2	(0007 = 0.6 , 0024 = 0.5 , 0014 = 0.29 , 0033 = 0.28)
bmu0	2.2	(0038 = 0.9 , 0035 = 0.4 , 0036 = 0.23 , 0031 = 0.16)
b5jj	2.2	(0025 = -0.7 , 0021 = -0.6 , 0036 = -0.5 , 0031 = -0.21)
e6e6	2.1	(0026 = 1.3 , 0015 = 0.6 , 0001 = 0.18)
e6ph0	2	(0034 = 0.8 , 0008 = 0.3 , 0039 = 0.22 , 0029 = 0.21)
e0jjj	2	(0006 = 0.6 , 0033 = 0.4 , 0022 = 0.3 , 0014 = 0.21)
e0pmiss	2	(0025 = 0.9 , 0011 = 0.6 , 0001 = 0.4)
e0jjpmiss	1.9	(0013 = -1.1 , 0012 = -0.25 , 0025 = -0.19 , 0014 = -0.12)
be0pmiss	1.8	(0012 = -0.6 , 0036 = -0.5 , 0013 = -0.25 , 0031 = -0.19)
e0e0	1.8	(0025 = -0.8 , 0015 = -0.7 , 0001 = -0.21)
e0jjjpmiss	1.7	(0014 = -1 , 0013 = -0.4 , 0025 = -0.12)
ph0tau	1.7	(0037 = 0.9 , 0004 = 0.31 , 0029 = 0.2 , 0039 = 0.15)
mu0pmiss	1.7	(0043 = -0.7 , 0027 = -0.5 , 0011 = -0.25 , 0001 = -0.2)
jjmu0pmiss	1.6	(0013 = -0.7 , 0017 = -0.2 , 0012 = -0.17 , 0043 = -0.14)
bj	1.6	(0018 = 0.6 , 0031 = 0.5 , 0036 = 0.4)

Table B.15: Correction factor influence table. Letting χ_k^2 denote the kth term in the χ^2 sum and s_i the ith correction factor, the *pull* of the ith bin on the kth correction factor is denoted pull_{ki}. The *total influence* of a bin k is defined as $\text{totalInfluence}_k \equiv \sum_i |\text{pull}_{ki}|$. Intuitively, bins with large total influence are "important" in influencing the position of the χ^2 minimum. Bins with large total influence tend to be big (containing many data events), pull on many correction factors, and prefer correction factors values significantly different from the values they would otherwise assume. Bins in this table are sorted in order of decreasing total influence, provided in the second column. In the third column between parentheses are the correction factors s_i that are most influenced by the bin. The extent to which these correction factors are influenced is also shown in the third column, with an entry such as 0001 = -0.65 indicating correction factor code 0001 feels a pull of -0.65. In each line, only the four largest contributions with pull ≥ 0.1 are listed. Due to the multiplicative nature of the correction factors, the pull on each correction factor from bin k is typically negative if the Standard Model prediction exceeds the number of data events in bin k, and positive if the Standard Model prediction falls short of the data in bin k.

	5001	5102	5103	5121	5122	5123	5124	5130	5131	5132	5141	5142	5143	5144	5151	5152	5153	5161	5162	5164	5165	5167	5168	5169	5170	5171	5211	5212	5213	5216	5217	5219	5246	5256	5257	5261	!
01	1	-.45	-.84	-.6	-.6	-.53	-.34	-.47	-.4	-.33	-.97	-.9	-.79	-.64	-.88	-.68	-.48	-.96	-.99	-.96	-.94	-.92	-.86	-.75	-.49	-.53	-.63	-.61	-.58	-.07	-.17	-.07	-.15	-.08	+.01	0	·
02	-.45	1	+.37	+.46	+.49	+.44	+.28	+.48	+.4	+.3	+.43	+.4	+.36	+.25	+.39	+.31	+.2	+.43	+.44	+.43	+.42	+.41	+.39	+.34	+.22	+.24	+.3	+.28	+.28	-.36	+.02	+.04	+.09	-.18	-.06	-.01	·
03	-.84	+.37	1	+.5	+.5	+.44	+.28	+.39	+.34	+.27	+.81	+.75	+.66	+.53	+.73	+.57	+.4	+.8	+.82	+.8	+.78	+.77	+.72	+.63	+.41	+.44	+.53	+.51	+.49	+.06	+.14	+.06	+.13	+.07	-.01	0	-
21	-.6	+.46	+.5	1	+.92	+.8	+.54	+.75	+.57	+.43	+.58	+.53	+.47	+.38	+.52	+.41	+.28	+.56	+.59	+.58	+.57	+.55	+.52	+.46	+.29	+.31	+.39	+.38	+.37	-.4	+.02	+.07	+.12	-.6	-.13	+.01	(
22	-.6	+.49	+.5	+.92	1	+.79	+.54	+.77	+.58	+.44	+.58	+.53	+.47	+.38	+.52	+.41	+.28	+.58	+.59	+.56	+.57	+.55	+.52	+.46	+.3	+.32	+.39	+.38	+.37	-.46	+.01	+.06	+.13	-.57	-.13	-.01	·
23	-.53	+.44	+.44	+.8	+.79	1	+.48	+.69	+.53	+.37	+.51	+.47	+.41	+.32	+.46	+.36	+.24	+.51	+.52	+.51	+.49	+.44	+.46	+.4	+.26	+.28	+.35	+.33	+.33	-.43	+.01	+.04	+.11	-.5	-.12	-.02	·
24	-.34	+.28	+.28	+.54	+.54	+.48	1	+.47	+.36	+.26	+.32	+.3	+.27	+.2	+.29	+.24	+.14	+.33	+.33	+.33	+.32	+.3	+.29	+.15	+.17	+.18	+.22	+.21	+.21	-.31	0	+.04	+.07	-.35	-.08	-.01	·
30	-.47	+.48	+.39	+.75	+.77	+.69	+.47	1	+.63	+.49	+.44	+.41	+.36	+.28	+.4	+.32	+.21	+.45	+.46	+.45	+.44	+.43	+.4	+.35	+.23	+.25	+.32	+.29	+.3	-.63	-.14	+.04	+.18	-.48	-.17	-.01	·
31	-.4	+.4	+.34	+.57	+.58	+.53	+.36	+.63	1	+.3	+.39	+.36	+.31	+.25	+.35	+.27	+.19	+.39	+.4	+.39	+.38	+.37	+.35	+.3	+.2	+.21	+.26	+.24	+.25	-.51	-.13	+.04	+.15	-.33	-.12	-.01	·
32	-.33	+.3	+.27	+.43	+.44	+.37	+.26	+.49	+.3	1	+.31	+.29	+.25	+.2	+.28	+.22	+.13	+.31	+.32	+.31	+.31	+.3	+.28	+.24	+.16	+.17	+.21	+.2	+.2	-.4	-.16	+.03	+.13	-.23	-.09	-.01	·
41	-.97	+.43	+.81	+.58	+.58	+.51	+.32	+.44	+.39	+.31	1	+.92	+.81	+.65	+.95	+.72	+.52	+.93	+.96	+.94	+.91	+.89	+.84	+.73	+.48	+.52	+.43	+.47	+.43	+.08	+.16	+.07	+.11	+.08	+.04	+.01	·
42	-.9	+.4	+.75	+.53	+.53	+.47	+.3	+.41	+.36	+.29	+.92	1	+.68	+.62	+.88	+.64	+.5	+.86	+.89	+.87	+.84	+.83	+.78	+.68	+.45	+.48	+.4	+.43	+.39	+.07	+.15	+.08	+.11	+.06	+.02	-.07	·
43	-.79	+.36	+.66	+.47	+.47	+.41	+.27	+.36	+.31	+.25	+.81	+.68	1	+.37	+.77	+.59	+.38	+.76	+.78	+.75	+.74	+.73	+.68	+.6	+.39	+.42	+.35	+.39	+.35	+.06	+.13	+.06	+.09	+.05	+.07	+.07	·
44	-.64	+.25	+.53	+.38	+.38	+.32	+.2	+.28	+.25	+.2	+.65	+.62	+.37	1	+.62	+.51	+.23	+.62	+.63	+.62	+.6	+.57	+.55	+.45	+.32	+.35	+.29	+.32	+.3	+.06	+.11	+.03	+.08	+.04	+.01	+.05	·
51	-.88	+.39	+.73	+.52	+.52	+.46	+.29	+.4	+.35	+.28	+.95	+.88	+.77	+.62	1	+.71	+.54	+.85	+.87	+.85	+.83	+.81	+.76	+.66	+.43	+.47	+.23	+.22	+.21	+.07	+.13	+.07	+.08	+.06	0	-.01	·
52	-.68	+.31	+.57	+.41	+.41	+.36	+.24	+.32	+.27	+.22	+.72	+.64	+.59	+.51	+.71	1	+.17	+.66	+.67	+.66	+.64	+.63	+.59	+.52	+.34	+.36	+.23	+.23	+.23	+.05	+.1	+.05	+.07	+.01	+.01	-.04	·
53	-.48	+.2	+.4	+.28	+.28	+.24	+.14	+.21	+.19	+.13	+.52	+.5	+.38	+.23	+.54	+.17	1	+.46	+.47	+.46	+.45	+.43	+.42	+.35	+.23	+.25	+.11	+.11	+.11	+.05	+.07	+.04	+.04	+.02	0	+.03	·
61	-.96	+.43	+.8	+.56	+.58	+.51	+.33	+.45	+.39	+.31	+.93	+.86	+.76	+.62	+.85	+.66	+.46	1	+.95	+.94	+.9	+.89	+.82	+.73	+.52	+.55	+.6	+.58	+.56	+.06	+.16	+.01	+.15	+.04	-.06	-.07	·
62	-.99	+.44	+.82	+.59	+.59	+.52	+.33	+.46	+.4	+.32	+.96	+.89	+.78	+.63	+.87	+.67	+.47	+.95	1	+.95	+.93	+.91	+.85	+.74	+.5	+.53	+.62	+.6	+.57	+.07	+.17	+.09	+.15	+.08	-.01	-.01	·
64	-.96	+.43	+.8	+.58	+.56	+.51	+.33	+.45	+.39	+.31	+.94	+.87	+.75	+.62	+.85	+.66	+.46	+.94	+.95	1	+.89	+.9	+.83	+.73	+.5	+.54	+.6	+.59	+.56	+.07	+.16	+.03	+.15	+.06	-.08	-.01	·
65	-.94	+.42	+.78	+.57	+.57	+.49	+.32	+.44	+.38	+.31	+.91	+.84	+.74	+.6	+.83	+.64	+.45	+.9	+.93	+.89	1	+.86	+.82	+.7	+.29	+.47	+.59	+.57	+.55	+.06	+.16	+.14	+.14	+.08	0		·
67	-.92	+.41	+.77	+.55	+.55	+.44	+.3	+.43	+.37	+.3	+.89	+.83	+.73	+.57	+.81	+.63	+.43	+.89	+.91	+.9	+.86	1	+.76	+.7	+.48	+.51	+.58	+.56	+.54	+.07	+.15	+.03	+.15	+.06	-.05	-.01	·
68	-.86	+.39	+.72	+.52	+.52	+.46	+.29	+.4	+.35	+.28	+.84	+.78	+.68	+.55	+.76	+.59	+.42	+.82	+.85	+.83	+.82	+.76	1	+.64	+.4	+.24	+.54	+.52	+.5	+.06	+.14	+.16	+.13	+.08	0		·
69	-.75	+.34	+.63	+.46	+.46	+.4	+.15	+.35	+.3	+.24	+.73	+.68	+.6	+.45	+.66	+.52	+.35	+.73	+.74	+.73	+.7	+.7	+.64	1	+.39	+.42	+.47	+.46	+.44	+.05	+.13	+.01	+.12	+.05	-.03	-.02	·
70	-.49	+.22	+.41	+.29	+.3	+.26	+.17	+.23	+.2	+.16	+.48	+.45	+.39	+.32	+.43	+.34	+.23	+.52	+.5	+.5	+.29	+.48	+.4	+.39	1	+.43	+.31	+.3	+.29	+.04	+.08	-.23	+.06	+.02	-.01	-.08	·
71	-.53	+.24	+.44	+.31	+.32	+.28	+.18	+.25	+.21	+.17	+.52	+.48	+.42	+.35	+.47	+.36	+.25	+.55	+.53	+.54	+.47	+.51	+.24	+.42	+.43	1	+.34	+.33	+.31	+.04	+.09	-.35	+.07	+.03	-.01	-.07	·
11	-.63	+.3	+.53	+.39	+.39	+.35	+.22	+.32	+.26	+.21	+.43	+.4	+.35	+.29	+.23	+.23	+.11	+.6	+.62	+.6	+.59	+.58	+.54	+.47	+.31	+.34	1	+.77	+.81	+.02	+.12	+.03	+.21	+.06	-.13	-.03	·
12	-.61	+.28	+.51	+.38	+.38	+.33	+.21	+.29	+.24	+.2	+.47	+.43	+.39	+.32	+.22	+.23	+.11	+.58	+.6	+.59	+.57	+.56	+.52	+.46	+.3	+.33	+.77	1	+.77	+.03	+.14	+.04	+.1	+.08	+.11	+.01	(
13	-.58	+.28	+.49	+.37	+.37	+.33	+.21	+.3	+.25	+.2	+.43	+.39	+.35	+.3	+.21	+.23	+.11	+.56	+.57	+.56	+.55	+.54	+.5	+.44	+.29	+.31	+.81	+.77	1	+.01	+.12	+.03	+.17	+.06	-.04	+.07	·
16	-.07	-.36	+.06	-.4	-.46	-.43	-.31	-.63	-.51	-.4	+.08	+.07	+.06	+.06	+.07	+.05	+.05	+.06	+.07	+.07	+.06	+.07	+.06	+.05	+.04	+.04	+.02	+.03	+.01	1	+.16	0	-.04	+.51	+.13	+.01	·
17	-.17	+.02	+.14	+.02	+.01	+.01	0	-.14	-.13	-.16	+.16	+.15	+.13	+.11	+.13	+.1	+.07	+.16	+.17	+.16	+.16	+.15	+.14	+.13	+.08	+.09	+.12	+.14	+.12	+.16	1	+.01	-.24	+.1	+.08	+.01	(
19	-.07	+.04	+.06	+.07	+.06	+.04	+.04	+.04	+.04	+.03	+.07	+.08	+.06	+.03	+.07	+.05	+.04	+.01	+.09	+.03	+.14	+.03	+.16	+.01	-.23	-.35	+.03	+.04	+.03	0	+.01	1	+.02	+.03	+.02	+.09	
46	-.15	+.09	+.13	+.12	+.13	+.11	+.07	+.18	+.15	+.13	+.11	+.11	+.09	+.08	+.08	+.07	+.04	+.15	+.15	+.15	+.14	+.15	+.13	+.12	+.06	+.07	+.21	+.1	+.17	-.04	-.24	+.02	1	-.02	-.43	-.01	(
56	-.08	-.18	+.07	-.6	-.57	-.5	-.35	-.48	-.33	-.23	+.08	+.06	+.05	+.04	+.06	+.01	+.02	+.04	+.08	+.06	+.08	+.06	+.08	+.05	+.02	+.03	+.06	+.08	+.06	+.51	+.1	+.03	-.02	1	+.18	+.03	
57	+.01	-.06	-.01	-.13	-.13	-.12	-.08	-.17	-.12	-.09	+.04	+.02	+.07	+.01	0	+.01	0	-.06	-.01	-.08	0	-.05	0	-.03	-.01	-.01	-.13	+.11	-.04	+.13	+.08	+.02	-.43	+.18	1	+.07	
61	0	-.01	0	+.01	-.01	-.02	-.01	-.01	-.01	-.01	+.01	-.07	+.07	+.05	-.01	-.04	+.03	-.07	-.01	-.01	0	-.01	0	-.02	-.08	-.07	-.03	+.01	+.07	+.01	+.01	+.09	-.01	+.03	+.07	1	
66	-.02	+.02	+.01	0	+.02	+.02	+.01	+.02	+.01	+.01	+.01	+.01	+.05	-.09	-.06	+.01	+.01	+.03	-.05	+.02	+.01	+.01	+.01	+.01	+.01	+.02	+.07	+.06	+.03	0	-.03	-.01	0	-.07	0		-.02	-.89
73	+.03	-.01	-.02	-.05	-.04	-.03	-.02	-.02	-.02	-.01	-.03	-.04	-.03	0	-.03	-.03	-.01	-.04	-.08	-.03	-.06	-.02	-.09	-.01	-.07	+.12	-.01	-.01	-.01	0	0	-.68	+.01	+.01	-.02	+.05	
85	-.04	+.02	+.03	-.05	-.04	-.03	-.02	+.05	+.04	+.02	+.04	+.03	+.03	+.02	+.02	+.02	+.01	0	+.04	+.02	+.04	+.03	+.04	+.02	-.01	0	+.04	+.06	+.04	-.01	+.12	+.04	+.07	+.1	+.04	+.04	
92	+.01	-.04	-.01	-.61	-.51	-.43	-.3	-.27	-.13	-.06	-.01	0	-.01	-.01	-.01	0	+.01	0	-.01	-.01	-.01	-.02	0	-.02	0	-.02	-.02	-.01	-.01	-.02	-.03	+.03	0	-.03	+.63	+.13	0
93	+.01	-.12	-.01	-.29	-.28	-.24	-.17	-.38	-.24	-.12	0	0	0	0	-.01	-.01	-.01	0	-.01	-.01	-.01	-.01	-.01	-.01	0	0	-.03	+.01	-.02	+.27	-.47	-.02	-.13	+.3	+.16	0	
02	+.02	0	-.01	-.02	-.02	-.02	-.01	+.01	0	0	-.08	-.05	-.12	-.05	-.01	-.04	0	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.01	-.01	+.13	-.2	0	-.01	+.05	0	+.02	+.01	-.61	-.06	
03	-.03	+.03	+.03	+.02	+.02	+.02	+.01	+.02	+.02	+.02	-.05	-.02	-.04	-.04	+.01	-.02	-.01	+.03	+.03	+.03	+.03	+.03	+.03	+.03	+.01	+.01	+.21	-.01	-.33	0	0	+.01	+.06	0	-.16	-.21	

Table B.16: Correction factor correlation matrix. The topmost row and leftmost column show correctio element of the matrix shows the correlation between the correction factor labeling the element's colum factor labeling the element's row. Each matrix element is dimensionless; the elements along the diagonal is symmetric; positive elements indicate positive correlation, and negative elements anti-correlation.

5102. This cosmic photon "k-factor" has increased due to requiring that this background satisfies the same good run list that are required for the data and by requiring that these events contain at least one reconstructed photon. As a result the number of events in this background has been decreased prompting this k-factor to increase accordingly.

5103. This cosmic jet "k-factor" has decreased due to the cut on the second jet in jet final states, as described in Sec. 4.2.3. The cut removes events in which the leading jet is due to a cosmic ray, and the other jets are due to the underlying event. As a result of this removal, the kfactor for this background has been reduced.

5121--5132. The k-factors for photon + jet production and diphoton production is consistent with values obtained in v01.

5151--5153. The k-factors for Z + jet production is consistent with values obtained in v01.

5141--5144. Motivated by a mistake in the modelling of the inoperational period of the keystone and miniskirt portions of the muon detector, we switched from the MadEvent W+jets Monte Carlo sample to the standard Top Group Alpgen W+jets sample. These k-factors were changed to correspond to Alpgen cross sections.

5161--5169. In v01 of this analysis we used $p(j \to j) = 1$, despite the fact that $p(j \to b) \gtrsim 0.01$. It is logically more consistent to chose $p(j \to j) = 1 - p(j \to b)$, so this is what is done in v02. The result of this modification is that k-factors for processes with one or more jets have increased.

5170,5171. These two k-factors for heavy flavor multijet production have been introduced.

5211,5212. The central electron identification efficiency is consistent with value obtained in v01. The phoenix electron identification efficiency scale factor has changed reflecting our change to the phoenix electron identification criteria.

5213. The muon identification efficiency scale factor has changed due to our change to the muon identification criteria, and the correction to the modelling of the inoperational period of the keystone and miniskirt portions of the muon detector.

5216,5217,5219. The identification efficiencies $p(\gamma \rightarrow \gamma)$ in the central and plug regions, and $p(b \rightarrow b)$ in the central region are consistent with values obtained in v01.

5245. The fake rate $p(e \rightarrow \gamma)$ has been removed after the change to the plug electron and photon identification. It was found to be unnecessary. This vanished correction factor is not listed in Table B.13.

5246. The fake rate $p(\gamma \rightarrow e)$ in the plug has been promoted to a correction factor from a fixed value of 0.005. This value increased significantly due to a redefinition of plug photons into electrons in the 1e+1ph final state. This was motivated by the fact that this plug photon was much more likely to have been an electron. We have removed this renaming procedure for the current version of the analysis.

5256,5257. The fake rates $p(q \rightarrow e)$ in the central and plug regions have decreased by roughly 13% and 6%, respectively, due to our improved conversion removal. In v01 we required a candidate conversion track to have $p_T > 2$ GeV; in v02 we make no transverse momentum requirement on the candidate converstion track. The change to the fake rate in the plug region is also affected by our change to the phoenix electron identification.

5261. The fake rate $p(q \rightarrow \mu)$ is consistent with the value obtained in v01.

5273. The fake rate $p(j \rightarrow b)$ is consistent with the value obtained in v01.

5285. A different p_T dependence has been imposed for the fake rate $p(q \rightarrow \tau)$ in v02 than applied in v01 and a dependence on the generated sumPt has also been applied. As a result of not being careful about proper normalizations of those functions, this number is not directly comparable to the one from v01.

5292. The value obtained for the fake rate $p(q \rightarrow \gamma)$ in the central region is consistent with the value obtained in v01.

5293. The fake rate $p(q \rightarrow \gamma)$ in the plug has decreased to due our correction to the plug photon identification criteria.

5401. The central electron trigger efficiency has been found to increase to unity in the current version of the analysis, because we now allow an event to pass on any online trigger. As a consequence, it is no longer appropriate to constrain this trigger efficiency to the Joint Physics value for the CEM trigger. We now simply fix the central electron trigger efficiency to 1.0 and it is no longer a correction factor. This vanished correction factor is not listed in Table B.13.

5402. The plug electron trigger efficiency is consistent with the value from v01.

5403. We have combined the CMUP and CMX trigger efficiencies due to the fact that they were very close to each other from v01 of the analysis. The value in v02 of the analysis is consistent with the values from v01.

Appendix C

Risk of Being *Ad Hoc*

C.1 Introduction

Here follows a general discussion, not so much about the actual SM implementation in this analysis, but about concerns such as bias, not being "blind", and how these factors affect the meaning of a null result.

In a search for new physics, especially a model-independent one, it is necessary to construct the Standard Model (SM) prediction. Then, one can test whether the data (D) are consistent with it.

By definition, the data follow the true law of nature. Denote the true theory by T. If there is physics beyond the SM, then $T \neq SM$. If new physics is to be observed, the p.d.f. of at least one observable quantity needs to differ adequately from that predicted by the SM.

Having the data events distributed according to T, one has the freedom to test their consistency with any conceivable theory. However, what is really interesting, is how well the data agrees with the SM, rather than some arbitrary model, not necessarily well motivated. We could, for example, construct a model agreeing bin by bin with the data. Imagine for instance having a dedicated k-factor[1] per final state;

---

[1] k-factors are corrections to the cross sections of processes. Typically, cross sections are calculated to leading-order, or next-to-leading-order, and rarely to an even higher order. k-factors are meant to correct such approximate calculations to the infinite-order cross section, which is incalculable, therefore k-factors are inferred from the data.
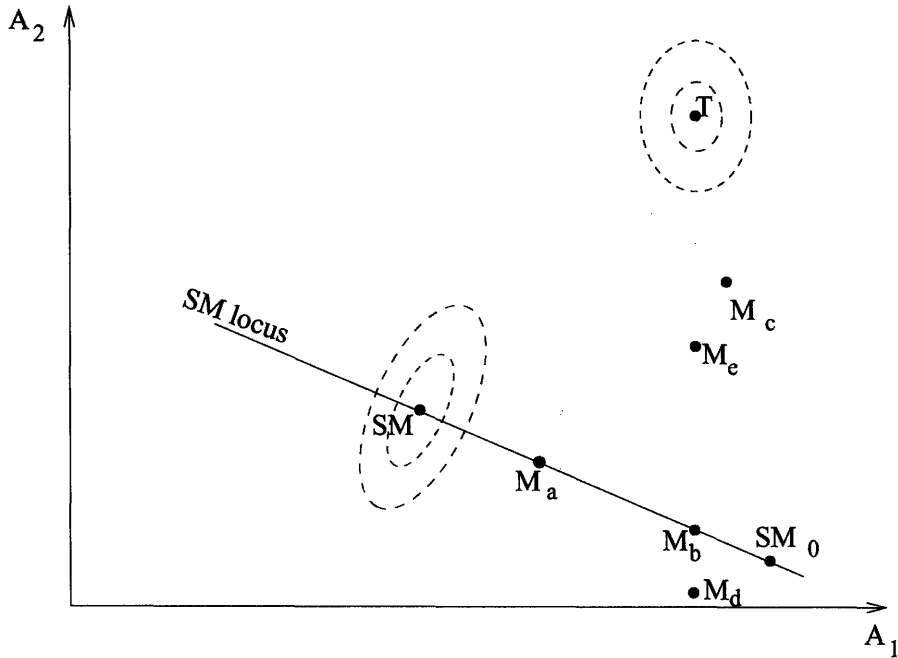
Figure C-1: Simplified picture of the p.d.f.s of the true theory and several possibilities for the SM implementation.

then we would be able to adjust this elastic pseudo-theory to match any combination of populations across final states. That data-obeying model would be consistent with T. Then, by construction, testing the quality of the fit would confirm the null hypothesis, namely that data agree with the constructed model. The hypothesis test itself would be perfectly legitimate, and its outcome would be correct, yet completely uninteresting, since nobody is interested in that absurd model anyway.

The problem then begins with the realization that the truly interesting hypothesis, the SM, is itself not known exactly; one needs information about correction factors, such as fake rates, k-factors, efficiencies etc. Different values of such parameters result in different "SM" predictions.

Let's assume there are only two observable quantities, A_1 and A_2 (Fig. C-1). For example, A_1 and A_2 could be the populations of events in two final states. Depending on the values of some correction factors (like k-factors etc.), the prediction of the SM implementation can be centered anywhere in some locus. In this case, the allowed locus is represented by a one-dimensional solid line; in general, the locus may be

226

higher-dimensional.

The correction factors have some true values, which may be unknown. The true Standard Model prediction is located at the "SM" point, which corresponds to the true values of the correction factors. Ideally, *that* is the SM we would like to compare to the data.

When the work to construct the SM prediction begins, one has no adjustments made yet, which results in some prediction centered on, say, point SM_0. One sees then the data[2], which are by definition near point T, and notices the discrepancies in A_1 and A_2. Since he has applied no corrections yet, he can not be confident that the current prediction is the real SM. The SM has been successful so far, therefore to rule it out one needs convincing evidence. To be convincing, he needs to be conservative; he must exploit any source of systematic uncertainty that he can identify in order to correct the prediction in a direction that brings it closer to the data. Unfortunately, there is no prescription how to do that correctly.

There are some obvious sources of uncertainty: k-factors reflecting the fact that it is not possible to calculate the infinite-order cross section of SM processes, uncertainties in the exact probability by which a particle may be misidentified, uncertainty in the integrated luminosity etc. For specific discrepancies that are not accounted for by such obvious uncertainties, one needs to become more imaginative to identify what may be causing them, but it is important to not invent false corrections. It requires judgment to make well motivated adjustments instead of *ad hoc* corrections that hide the signal of potentially new physics. The locus, represented by the solid line in Fig. C-1, is meant to represent the possible predictions that can be derived by making well motivated corrections, whereas points out of the locus represent the results of poorly motivated corrections.

Suppose that throughout the process one makes well motivated corrections. Then his prediction should drift along the locus from point SM_0 to point M_a, which gives the best agreement with the observed data in A_1 and A_2 simultaneously. Even though $M_a \neq SM$, he will need to stop at M_a and not proceed towards the actual SM point.

---

[2]Whether he sees all, or part, or only some aspect of them will be discussed later.

That is because he has no way to know if he has reached the actual SM to stop there; his only guidance is the data and his judgment. To be conservative, he would have to bring the prediction as close to point T as allowed, but not closer – that is point M_a. The wrong thing to do would be to introduce extraneous, poorly motivated corrections that would drive one from SM_0 to a prediction like M_c, namely out of the locus. That would be the result of *ad hoc* treatment of discrepancies, which in its extreme limit would result in a model as uninteresting as the data-obeying model mentioned earlier.

What can safeguard one from constructing the prediction of some poorly motivated model? Only prudence and an over-constrained system that limits systematic uncertainties, making it harder to deviate from the SM locus. The risk of implementing an *ad hoc* model remains, unless all systematic uncertainties shrunk to zero, in which ideal case the locus would shrink into just the true SM point. However, there are some "blind" approaches that, as will be argued, create the illusion of safety against erring, or the sensation that information is generated out of nothing, by using the data in "clever" ways, i.e. by not seeing all of them at the same time.

C.2 Blind to signal region

In some cases (not in this analysis) one may presume that the new physics will be affecting A_2 but not A_1. That is clearly an assumption, which in many cases can be motivated. A_2 is then treated as "signal region", and A_1 as "control region". Adjusting the correction model to achieve maximal agreement with the data in A_1 is legitimate, since the premise is that the SM should distribute A_1 as T does. That leads (if everything is done correctly) to a SM implementation with p.d.f. centered on M_b.

There is nothing wrong in defining control and signal regions. Clearly, when interpreting the result of the comparison of the data with M_b one needs to remember that M_b is not the globally best fitting model (that would be M_a). Furthermore, M_b is not necessarily the true SM, but is the model that best fits the control region.

Indicative of the value of such results is the fact what is "signal" region in one analysis can be "control" in another. Depending on what one defines as "signal" and "control", the result may vary from agreement to disagreement with the data. Although these results can be valid, they are convincing only if the initial premise is accepted.

Unfortunately, staying "blind" in A_2 does not guarantee that the final model will not be an absurd and *ad hoc* one. For example, a human error may lead one from SM_0 to M_d or M_e. Apart from a human error that may occur during the development of the correction model, opening the box (e.g. looking at the measured A_2) often makes people question the correctness of their implemented model, especially in the event of a discrepancy with the data. In that phase of reconsideration, one may even accidentally change his background model from M_b to M_e, so the notion of "blindness" is questionable, unless no discrepancy is seen. Therefore, as in the non-blind analysis case, prudence and an over-constrained system that limits systematic uncertainties, making it harder to deviate from the SM locus, can prevent testing the goodness of a worthless model (like M_e, M_c or M_d).

C.3 Blind to part of the data

Another approach considered "blind" is to split the whole data set (D) in two parts (D_{control}, D_{signal}), assigning for example every third event to D_{control} and the rest to D_{signal}. Then, D_{control} can be used to develop the correction model, and D_{signal} is only revealed in the end, to check how well it is fitted by the derived background model.

The supposed advantage of this approach is that D_{signal} is independent from D_{control}. So, if agreement is observed between D_{signal} and the background model, that supposedly can not be due to a biased model, as the background model was developed knowing nothing about D_{signal}. Though psychologically reassuring, this impression of safety is false.

Obviously, all data come from the same distribution T, therefore there is no reason why D_{signal} would be distributed any differently than D_{control}, apart from random statistical fluctuations, which actually become bigger when D_{control} and D_{signal} have

smaller populations.

If one makes wrong judgments in the way he uses $D_{control}$, then there are two possibilities: If one observes agreement between the background model and D_{signal}, that only means that D_{signal} didn't fluctuate too differently than $D_{control}$. On the other hand, if one observes disagreement, that would only be due to (rare) statistical fluctuation of D_{signal} with respect to $D_{control}$. In other words, if one makes the wrong use of $D_{control}$ the result is as uninformative as it would be if he had used the whole D in a wrong way.

Furthermore, even if one is very prudent and has an over-constrained system with small systematic uncertainties, still splitting the data makes the situation worse. Having less data in $D_{control}$ to constrain the correction factors makes the locus where SM could be larger, therefore it is more likely to end up with a correction model farther away from the actual SM, simply due to larger systematic uncertainties. Furthermore, having a smaller number of data in D_{signal} reduces statistical power, making it harder to observe a real effect that may appear in the measured A_1 and A_2.

In summary, splitting D in two does not secure one from implementing wrongly his theoretical prediction. If one can make proper use of $D_{control}$, then he can also make proper use of the whole D, which would offer the advantage of smaller uncertainties.

C.4 Summary

To summarize, there is no way to be sure that the null hypothesis compared to the data is the SM, rather than some other uninteresting one. However, there is reason to hope that what was tested in this analysis is the agreement of the data with a model that at least is possible to be the SM, namely belongs to the SM locus determined by well motivated systematic uncertainties. Certainly, the tested model is biased to agree with the data more than the SM may actually agree[3], since the best fitting choice of correction parameters was made, but that is inevitable, since the SM is assumed correct until proof of the contrary. The hope that the implemented background

---

[3]Think of the analogy given by points "SM" and M_a in Fig. C-1.

model is not far from the actual SM is based on the fact that the correction model is significantly over-constrained by examining not just a couple of observables, but thousands. After all, human errors are always possible, but the best effort was made to eliminate them. Well motivated corrections usually fix several problems at once, while mistaken adjustments tend to fix one problem but cause other. Our global approach allowed us to distinguish the former from the latter, by monitoring simultaneously so many observables before and after the adjustments.

Appendix D

Nomenclature

BMU Barrel Muon system. Often synonymous to IMU

CDF Collider Detector at Fermilab

CEI Charge Exchange Injection

CEM Central Electromagnetic calorimeter

CERN Conseil Européen pour la Recherche Nucléaire

CES Central Electromagnetic Showermax detector

CHA Central Hadronic calorimeter

CKM Cabibbo Kobayashi Maskawa

CLC Cerenkov Luminosity Counter

CMP Central Muon Upgrade

CMUP A muon that has both CMU and a CMP hits

CMU Central Muon Detector

CMX Central Muon Extension

COT Central Outer Tracker

CPR Central Preshower detector

CPU Central Processor Unit

CP Charge Parity

CSL Consumer Server Logger

DAQ Data Acquisition

DIS Deep Inelastic Scattering

EM Electromagnetic

EVB Event Builder

EWK Electroweak

EWSB Electroweak Symmetry Breaking

FCC Feynman Computing Center

FIFO First In First Out

FNAL Fermi National Accelerator Laboratory

GMSB Gauge Mediated Supersymmetry Breaking

GUT Grand Unification Theory

ID Identification

IMU Intermediate Muon system

ISL Intermediate Silicon Layer

KS Kolmogorov Smirnov

L00 Layer 0 of the Silicon Detector

LHC Large Hadron Collider

LO leading order

MC Monte Carlo

MET Missing Transverse Energy

MIP Minimum Ionizing Particle

MI Main Injector

PDF Parton Distribution Function

p.d.f. Probability Density Function

PEM Plug Electromagnetic calorimeter

PES Plug Electromagnetic Showermax detector

PHA Plug Hadronic calorimeter

PHX "Phoenix", referring to forward tracks reconstructed from silicon hits

PMNS Pontecorvo Maki Nakagawa Sakata

PMT Photomultiplier

QCD Quantum Chromodynamics

RF Radio Frequency

SCPU	Scanner CPU
SM	Standard Model of elementary particles
SUGRA	Supergravity
SUSY	Supersymmetry
SVT	Silicon Vertex system
SVX	Silicon Vertex Detector
TSI	Trigger Supervisor
UV	Ultraviolet
VME	Virtual Machine Environment, a standard mainframe operating system
VRB	VME Readout Buffer (or Board)
WHA	Endwall Hadronic calorimeter
WLS	Wavelength Shifting (optic fiber)
XCES	Extrapolation to Central Electromagnetic Showermax
XFT	Extremely Fast Tracker
XTRP	Extrapolation Unit

Bibliography

[1] John F. Donoghue, Eugene Golowich, and Barry R. Holstein. *Dynamics of the Standard Model*, chapter I,II,III. Cambridge University Press, 1994.

[2] Esteban Roulet. Beyond the Standard Model. 2001, hep-ph/0112348.

[3] Review of Particle Physics. *Physics Letters B*, 592:1+, 2004.

[4] Keith A. Olive. Dark matter. 2003, astro-ph/0301505.

[5] Chris Quigg. Spontaneous symmetry breaking as a basis of particle mass. *Reports on Progress in Physics*, 70(7):1019–1053, 2007.

[6] Jogesh C. Pati and Abdus Salam. Unified Lepton-Hadron Symmetry and a Gauge Theory of the Basic Interactions. *Phys. Rev.*, D8:1240–1251, 1973.

[7] H. Georgi and S. L. Glashow. Unity of All Elementary Particle Forces. *Phys. Rev. Lett.*, 32:438–441, 1974.

[8] Stephen P. Martin. A supersymmetry primer. 1997, hep-ph/9709356.

[9] Ted Thomas Derek Raine. *An introduction to the science of Cosmology*, chapter 3.12.3. Institute of Physics Publishing, 2001.

[10] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. The hierarchy problem and new dimensions at a millimeter. *Phys. Lett.*, B429:263–272, 1998, hep-ph/9803315.

[11] Thomas Appelquist, Hsin-Chia Cheng, and Bogdan A. Dobrescu. Bounds on universal extra dimensions. *Phys. Rev.*, D64:035002, 2001, hep-ph/0012100.

[12] Lisa Randall and Raman Sundrum. A large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999, hep-ph/9905221.

[13] R. Sekhar Chivukula. Technicolor and compositeness. 2000, hep-ph/0011264.

[14] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995, hep-ex/9503002.

[15] Fermilab's chain of accelerators: The proton source. http://www-bd.fnal.gov/public/proton.html.

[16] Fermilab linac upgrade conceptual design revision 4a. FERMILAB-LU-CONCEPTUAL-DESIGN.

[17] Booster rookie book. http://www-bdnew.fnal.gov/operations/rookie_books/Booster_V3_1.pdf.

[18] C. Hojvat et al. The multiturn charge exchange injection system for the fermilab booster accelerator. (talk). *IEEE Trans. Nucl. Sci.*, 26:3149–3151, 1979.

[19] Main injector performance goals. 1998. http://www-fmi.fnal.gov/Preform%20Goals/Chapter_5.pdf.

[20] The antiproton source rookie book. 1999. http://www-bdnew.fnal.gov/pbar/documents/PBAR_Rookie_Book.pdf.

[21] B. F. Bayanov et al. The proton beam lithium lens for the fermilab anti-proton source. FERMILAB-TM-1000.

[22] Gerry Jackson. The fermilab recycler ring technical design report. rev. 1.2. FERMILAB-TM-1991.

[23] Dave McGinnis. Run ii handbook. 2001. http://www-bd.fnal.gov/runII/index.html.

[24] Krzysztof Genser and Paul Lebrun. Tevatron Transverse Beam Emittance and Luminosity at CDF and D0 for stores 3179-3293. 2004. Beams-doc-1075.

[25] Vaia Papadimitriou. Updated history of CDF/D0 luminosity ratio and some correlations. 2006.
http://www-bd.fnal.gov/SDA_Viewer/VaiaLuminosity/vaia_lum_16Aug06_web.ppt.

[26] Patrick T. Lukens. The CDF IIb detector: Technical design report. FERMILAB-TM-2198.

[27] Bart Zeghbroeck. *Principles of Semiconductor Devices*. 2004.
http://ece-www.colorado.edu/~bart/book/book/.

[28] A. Boveia. Status and performance of the CDF Run II silicon detector. *PoS*, HEP2005:377, 2006.

[29] Joe Incandela, Jeff Spalding, Paul Shepard, Tim Nelson, David Stuart, Maurice Garcia-Sciveres, Igor Volobouev, and Steve Worm. The CDF Run II Silicon Tracking System. *Nucl. Instrum. Meth.*, A447:1–8, 2000.

[30] Open-cell chamber to replace the CTC. *CDF note 3648*, 1996.

[31] CDF Central Outer Tracker. *CDF note 6227 and Nucl.Instru.Meth.*, A 526, 2004.

[32] Hans Wenzel. Tracking In The SVX. *CDF note 1790*, 1992.

[33] Paolo Gatti. Performance of the new tracking system at CDF II. *CDF note 5561*, 2001.

[34] W.M. Yao and K. Bloom. Outside-in Silicon Tracking at CDF. *CDF note 5991*, 2002.

[35] Yimei Huang, Chris Hays, and Ashutosh Kotwal. Inside-Out Tracking. *CDF note 6707*, 2003.

[36] F. Abe et al. The CDF detector: an overview. *Nucl. Instr. Meth.*, A271:387–403, 1988.

[37] D. Acosta et al. The cdf cherenkov luminosity monitor. *Nucl. Instrum. Meth.*, A461:540–544, 2001.

[38] Georgios Choudalakis, Conor Henderson, Khaldoun Makhoul, and Markus Klute. CDF Event Builder for the Shift Crew. *CDF Note 7846*, 2005.

[39] Markus Klute. CDF RunIIb Event Builder: Data Integrity Checks performed by the SCPU. *CDF note 8021*, 2006.

[40] R. Brun and F. Rademakers. Root: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[41] T. Aaltonen et al. Model-Independent and Quasi-Model-Independent Search for New Physics at CDF. 2007, arXiv:0712.1311 [hep-ex].

[42] CDF Collaboration. Model-Independent Global Search for New High-pT Physics at CDF. 2007, arXiv:0712.2534 [hep-ex].

[43] Georgios Choudalakis. Vista, results of a model-independent search for new physics in 927 pb^{-1} at CDF. 2007, arXiv:0710.2372 [hep-ex].

[44] Georgios Choudalakis. Sleuth at CDF, a quasi-model-independent search for new electroweak scale physics. 2007, arXiv:0710.2378 [hep-ex].

[45] F. Krauss. Matrix elements and parton showers in hadronic interactions. *JHEP*, 08:015, 2002, hep-ph/0205283.

[46] CDF Collaboration. Measurements of inclusive w and z cross sections in p-pbar collisions at sqrts =1.96 tev. *J. Phys. G*, 34:2457, 2007.

[47] CDF Collaboration. 2007. *Phys. Rev. D* **75** 092004.

[48] CDF Collaboration. Measurement of the cross section for prompt diphoton production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ tev. 2005. *Phys. Rev. Lett.* **95** 022003.

[49] Robert Craig Group. Measurement of the inclusive jet cross section using the midpoint algorithm in run ii at the collider detector at fermilab (cdf). FERMILAB-THESIS-2006-29.

[50] A. Bhatti et al. Determination of the jet energy scale at the collider detector at fermilab. 2006. *Nucl. Instrum. Meth.* **A566** 375.

[51] Christopher Neu. CDF b-tagging: Measuring efficiency and false positive rate. 2006. Presented at TOP 2006: International Workshop on Top Quark Physics, Coimbra, Portugal.

[52] A.V. Kotwal, H.K. Gerberich, and C. Hays. Identification of cosmic rays using drift chamber hit timing. *Nucl. Instrum. Meth.*, A506:110, 2003.

[53] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *JHEP*, 05:026, 2006, hep-ph/0603175.

[54] Fabio Maltoni and Tim Stelzer. Madevent: Automatic event generation with madgraph. *JHEP*, 02:027, 2003, hep-ph/0208156.

[55] G. Corcella et al. Herwig 6.5 release note. 2002, hep-ph/0210213.

[56] CTEQ Collaboration. Global QCD analysis of parton structure of the nucleon: Cteq5 parton distributions. 2000. *Eur. Phys. J.* **C 12** 375.

[57] Rick Field. Cdf run 2 monte-carlo tunes. 2005. Proceedings of TeV4LHC Workshop, Fermilab, Batavia, IL. FERMILAB-CONF-06-408-E.

[58] Stephen Mrenna and Peter Richardson. Matching matrix elements and parton showers with herwig and pythia. *JHEP*, 05:040, 2004, hep-ph/0312274.

[59] Nikolaos Kidonakis and Ramona Vogt. Next-to-next-to-leading order soft-gluon corrections in top quark hadroproduction. *Phys. Rev.*, D68:114014, 2003, hep-ph/0308222.

[60] T. Stelzer and W. F. Long. Automatic generation of tree level helicity amplitudes. 1994. *Comput. Phys. Commun.* **81** 357.

[61] E. Gerchtein and M. Paulini. Cdf detector simulation framework and performance. 2003, physics/0306031.

[62] Guenter Grindhammer, M. Rudowicz, and S. Peters. The fast simulation of electromagnetic and hadronic showers. *Nucl. Instrum. Meth.*, A290:469, 1990.

[63] D. Acosta et al. The performance of the cdf luminosity monitor. *Nucl. Instrum. Meth.*, A494:57–62, 2002.

[64] CDF Collaboration. Further properties of high-mass multijet events at the fermilab proton-antiproton collider. 1996. *Phys. Rev. D* **54** 4221.

[65] CDF Collaboration. Properties of six-jet events with large six-jet mass at the fermilab $p\bar{p}$ collider. 1997. *Phys. Rev. D* **56** 2532.

[66] D0 Collaboration. Transverse energy distributions within jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ tev. 1995. *Phys. Lett. B* **357** 500.

[67] J. Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. 2007, arXiv:0706.2569 [hep-ph].

[68] B. Knuteson. PhD thesis, University of California, Berkeley, 2000.

[69] DØ Collaboration. Search for new physics in $e\mu X$ data at DØ using SLEUTH: A quasi model independent search strategy for new physics. 2000. *Phys. Rev. D* **62** 092004.

[70] DØ Collaboration. A quasi-model-independent search for new physics at large transverse momentum. 2001. *Phys. Rev. D* **64** 012004.

[71] DØ Collaboration. A quasi-model-independent search for new high p_T physics at DØ. 2001. *Phys. Rev. Lett.* **86** 3712.

[72] H1 Collaboration. A general search for new phenomena in e p scattering at hera. 2004. *Phys. Lett. B* **602** 14.

[73] CDF Collaboration. Observation of top quark production in $\bar{p}p$ collisions. 1995. *Phys. Rev. Lett.* **74** 2626.

[74] DØ Collaboration. Observation of the top quark. 1995. *Phys. Rev. Lett.* **74** 2632.

[75] CDF Collaboration. Measurement of the top-quark mass in all-hadronic decays in $p\bar{p}$ collisions at cdf ii. 2007. *Phys. Rev. Lett.* **98** 142001.

[76] CDF Collaboration and DØ Collaboration. A combination of CDF and DØ results on the mass of the top quark. 2007. hep-ex/0703034.

[77] CDF Collaboration. Search for anomalous production of diphoton events with missing transverse energy at cdf and limits on gauge- mediated supersymmetry-breaking models. 2005. *Phys. Rev. D* **71** 031104.

[78] A. Abulencia et al. Search for $z' \rightarrow e^{+}e^{-}$ using dielectron mass and angular distribution. 2006. *Phys. Rev. Lett.* **96** 211801.

[79] CDF Collaboration. Search for resonant ttbar production in ppbar collisions at sqrts=1.96 tev. 2007, arXiv:0709.0705 [hep-ex].

[80] CDF Collaboration. Search for large extra dimensions in the production of jets and missing transverse energy in p anti-p collisions at s**(1/2) = 1.96-tev. 2006. *Phys. Rev. Lett.* **97** 171802.

[81] CDF Collaboration. Inclusive Search for New Physics with Like-Sign Dileptons in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. 2006. CDF-8643.

[82] D. Hare, E. Halkiadakis, T. Spreitzer. Electron ID Efficiency and Scale Factors for Winter 2007 Analyses. 2006. CDF-8614.

[83] Sarah Budd, Matthias Buhler, Catalin Ciobanu, Peter Dong, Richard Hughes, Thomas Junk, Kevin Lannon, Jan Lueck, Thomas Muller, Svenja Richter, Jason Slaunwhite, Bernd Stelzer, Jeannine Wagner, Wolfgang Wagner, Rainer Wallny, Brian Winer. Event detection efficiency for single-top events and MC based background estimate for Summer 2006. 2006. CDF-8286.

[84] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56–74, 1998, hep-ph/9704448.

[85] Zoltan Nagy. Next-to-leading order calculation of three-jet observables in hadron hadron collision. 2003. *Phys. Rev. D* **68** 094002.

[86] Zoltan Nagy and Zoltan Trocsanyi. Multi-jet cross sections in deep inelastic scattering at next-to-leading order. 2001. *Phys. Rev. Lett.* **87** 082001.

[87] T. Binoth, J. P. Guillet, E. Pilon, and M. Werlen. A full next to leading order study of direct photon pair production in hadronic collisions. 2000. *Eur. Phys. J. C* **16** 311.

[88] P. J. Sutton, Alan D. Martin, R. G. Roberts, and W. James Stirling. Parton distributions for the pion extracted from drell-yan and prompt photon experiments. 1992. *Phys. Rev. D* **45** 2349.

[89] S. Alekhin. Parton distribution functions from the precise nnlo qcd fit. *JETP Lett.*, 82:628–631, 2005.

[90] Daniel Stump et al. Inclusive jet production, parton distributions, and the search for new physics. 2003. *J. High Energy Phys.* **10** 046.

[91] CDF Collaboration. Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 tev using lepton + jets events with secondary vertex b–tagging. 2005. *Phys. Rev. D* **71** 052003.

[92] Andrea Messina. Measurement of the w + jet cross section at cdf. 2007, arXiv:0708.1380 [hep-ex].

[93] T. Spreitzer, C. Mills, J. Incandela. Electron Identification in Offline Release 6.1.2. 2006. CDF-7950.