Full many a gem of purest ray serene,
The dark unfathom'd caves of ocean bear:
Full many a flow'r is born to blush unseen,
And waste its sweetness on the desert air.

Th. Gray, *"Elegy Written in a Country Churchyard"* (1751), vv.53-56.

Contents

Chapter 1

Introduction

In this chapter we give a brief introduction to the Standard Model of Electroweak Interactions, we describe the Higgs mechanism and the current limits on the mass of the Higgs boson, and finally we explain what are the motivations for a search of $Z \to b\bar{b}$ decays at a $p\bar{p}$ collider. We conclude the chapter with Sec. 1.4, where a quick description of the contents of the present work is given.

1.1 The Current Status of the Standard Model

Our current understanding of the subatomic physical world rests on the fundamental hypothesis that those physical quantities which we expect to be conserved do so in every space-time point, that is *locally*, and not just at a global level. In the framework of Quantum Electrodynamics, for instance, requiring the conservation of electric charge in every space-time point is equivalent to ruling that the free Lagrangian density describing a given system, $\mathcal{L}$, keep the same form under local gauge transformation of the fields:

$$\psi(x) \to e^{ie\alpha(x)}\psi(x) \qquad \Rightarrow \qquad \mathcal{L} \to \mathcal{L}. \tag{1.1}$$

This forces the introduction of a vector field $\mathcal{A}_\mu$, coupling to the electromagnetic current with intensity proportional to the unit charge e. The latter also has the role of generator of the group of gauge transformations, an Abelian group $U(1)$. The vector particle of the field $\mathcal{A}_\mu$ is the photon, and the invariance of $\mathcal{L}$ is warranted only if the photon is devoid of mass. Transformations such as Eq. 1.1 are called local gauge transformations, and generally the theories based on symmetry principles possessed by such operations are called gauge theories.

The first step toward the unification of electromagnetism with the weak interactions was taken by S. Glashow in 1961[1]. He considered a local gauge transformation group $SU(2)_L \times U(1)_Y$ under which the lepton Lagrangian was invariant: assuming that leptons can be written as weak isospin doublets[1],

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}_L, \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}_L, \cdots, \tag{1.2}$$

a local gauge transformation acts on them by 2×2 Hermitian matrices, whose elements are the exact analogue of the $\alpha(x)$ function introduced in Eq. 1.1. Four vector bosons correspond to the matrices, and they carry the four respective gauge fields; with them one can explain both the electromagnetic and the weak interaction. Just as the photon, the $SU(2)_L \times U(1)_Y$ gauge bosons are required to have no mass, in order to maintain $\mathcal{L}$ invariant with respect to the group elements.

The problem of the model described by Glashow was that many experimental results, such as the lepton-nucleon scattering cross section, were not explanable without allowing for a large mass of the vector bosons. This puzzle was given a brilliant explanation a few years later by S. Weinberg and A. Salam[2, 3]. They made the hypothesis of a spontaneous symmetry breaking mechanism (see below, Sec. 1.2), capable of endowing with mass the three bosons responsible for the weak interaction. The price to pay was the introduction of at least one scalar field—the Higgs field[4]—mediated by an appropriate boson H^0. The proof of renormalizability of the theory thus obtained was given by G. 't Hooft in 1971[5], and 12 years later, with the discovery at the CERN laboratories in Geneva of the W and Z bosons[6, 7, 8, 9], the correctness of the electroweak model acquired widespread consensus.

The gauge boson Z^0 is the vector of neutral weak current processes, where no electric charge is exchanged between the interacting particles. The W^+ and W^- bosons mediate instead the charged current weak interactions: they transform a member of a fermionic doublet into its partner. The lepton doublets and singlets listed above can be joined by three quark doublets, plus six right-handed helicity singlets:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L; u_R, d_R, c_R, s_R, t_R, b_R. \tag{1.3}$$

A mixing of the states is at work between the three quark generations. It is governed by the unitary Cabibbo-Kobayashi-Maskawa matrix (CKM)[10]:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{1.4}$$

The amplitude of weak charged currents changing the flavor of quarks are proportional to the elements of the CKM matrix: the nonvanishing value of off-diagonal elements warrants that a mixing between the members of the three quark doublets will occur in weak interactions. This is conventionally attributed to the quarks with weak isospin $T_3 = -1/2$, and is described by the operation $(d's'b')^T = V(dsb)^T$.

[1]Besides the doublets there are also singlets $e_R, \mu_R, \cdots$, which have zero weak isospin and do not vary under the transformations of the $SU(2)_L$ group.

To summarize, the Electroweak Model gathers in the same Lagrangian density the weak and electromagnetic interactions, and classifies fermions into doublets and singlets of weak isospin.

The model now described can be completed by a description of the strong interactions, which are responsible for the stability of hadronic matter. The theory describing these interactions, Quantum Chromodynamics (QCD), is a gauge theory based on the $SU(3)_C$ symmetry group. The charge acting as source of the field, called color, coincides with that attribute of quarks invoked at the beginning of the 60's in the static models of hadrons to avoid having to part with the Pauli exclusion principle: particles such as the Δ^{++} could be really formed by three otherwise identical fermions u only if they possessed some attribute capable of distinguishing them from each other.

According to QCD there are three possible colors per each quark, and just as many anticolors for their antiparticles. Quarks form hadronic matter by joining in pairs $(q\bar{q}')$ or triplets $(q_1q_2q_3, \bar{q}_1\bar{q}_2\bar{q}_3)$ that make up color singlet configurations, *id est* irreducible representations of the $SU(3)_C$ group with zero color charge. The carriers of the color force are eight gluons, the generators of the group. They have no mass, since the $SU(3)_C$ symmetry is not broken as the electroweak symmetry. Gluons carry color charge, which is the conserved quantity in strong interactions.

Combining the electroweak theory with QCD one obtains an invariant gauge theory based on the group $SU(3)_C \times SU(2)_L \times U(1)_Y$: this theory is nowadays commonly named the Standard Model.

1.2 The Higgs Boson

The Higgs mechanism represents the only known method to give mass to the electroweak vector bosons preserving at the same time the renormalizability of the theory. It postulates the existence of a scalar field with a nonvanishing vacuum expectation value. In the Standard Model the scalar field is described by a doublet of complex fields ϕ^+ and ϕ^0 of electrical charge $Q = (+1, 0)$ and hypercharge $Y = +1$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.5}$$

Symmetry breaking occurs when a self-interaction term for the ϕ field is added to the Lagrangian density describing the electroweak interactions,

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2, \tag{1.6}$$

choosing $\mu^2 < 0$ and $\lambda > 0$ such that the potential has a nonvanishing minimum value:

$$\phi^\dagger \phi = -\frac{\mu^2}{\lambda} = \frac{v^2}{2}. \tag{1.7}$$

The most convenient choice for the ground state is the following:

$$\phi_0 = \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.8}$$

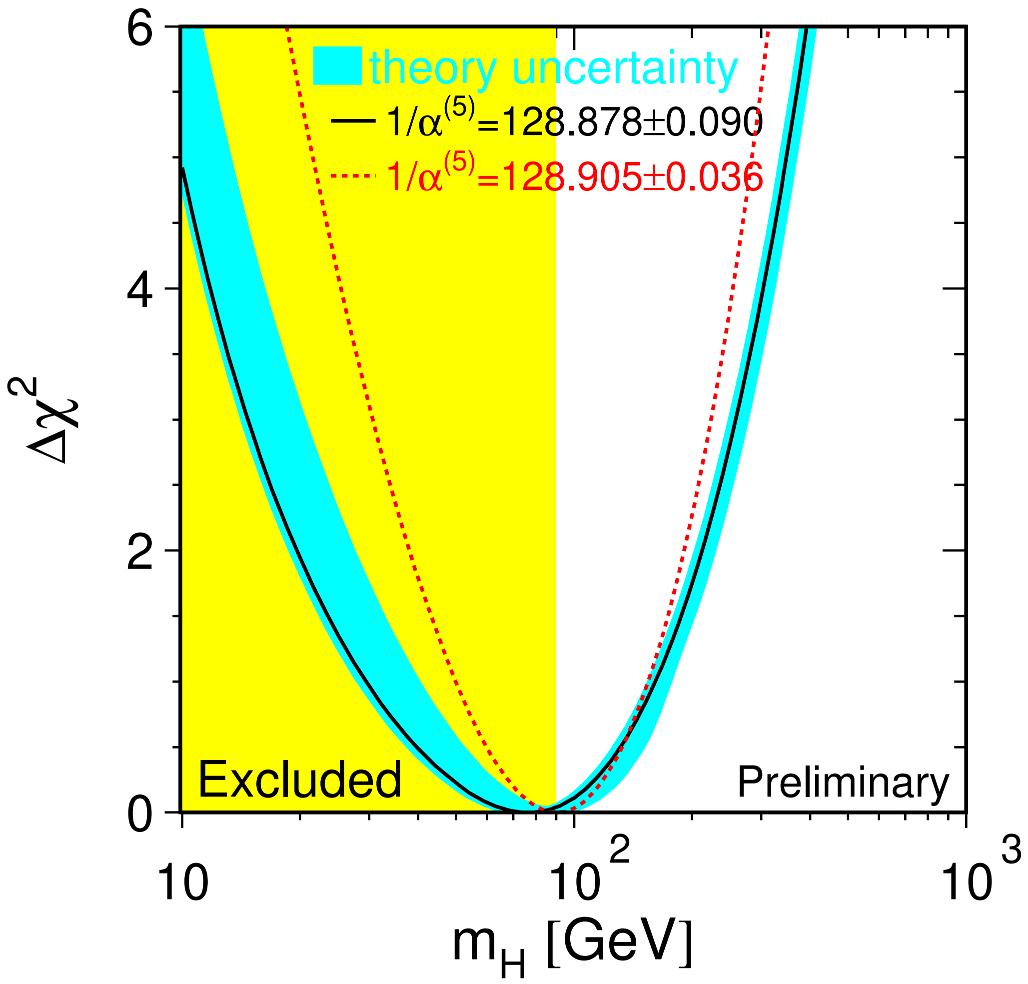

Figure 1.1: *Results of the global likelihood fits of electroweak observables on the Higgs boson mass[11].*

which means defining $\phi_1 = \phi_2 = \phi_3 = 0$ and $\phi_4 = v$. The above definition is based on the fact that an expectation value on the ground state different from zero for the Higgs field automatically breaks the symmetry and endows with mass the gauge bosons corresponding to the lost degrees of freedom; the chosen value of ϕ_0 breaks both the $SU(2)_L$ and the $U(1)_Y$ symmetries of the electroweak Lagrangian density, but—being a neutral field—its expectation value on the $U(1)_Q$ generator is still zero:

$$Q\phi_0 = \left(I_3 + \frac{Y}{2}\right)\phi_0 = 0. \tag{1.9}$$

Because of that, the $U(1)_Q$ symmetry of the Lagrangian is preserved and the photon does not acquire any embarassing mass.

In the vicinity of the ground state the symmetry is broken and the scalar field can be written as

$$\phi_0 = \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{1.10}$$

Three of the four lost degrees of freedom are used by the W and Z bosons to get longitudinal components to their polarization vectors,

$$M_W = \frac{gv}{2} \tag{1.11}$$

$$M_Z = \frac{v}{2}\sqrt{g^2 + g'^2} \tag{1.12}$$

and the fourth gives mass to the neutral Higgs field:

$$M_H = \sqrt{2\lambda v^2}. \tag{1.13}$$

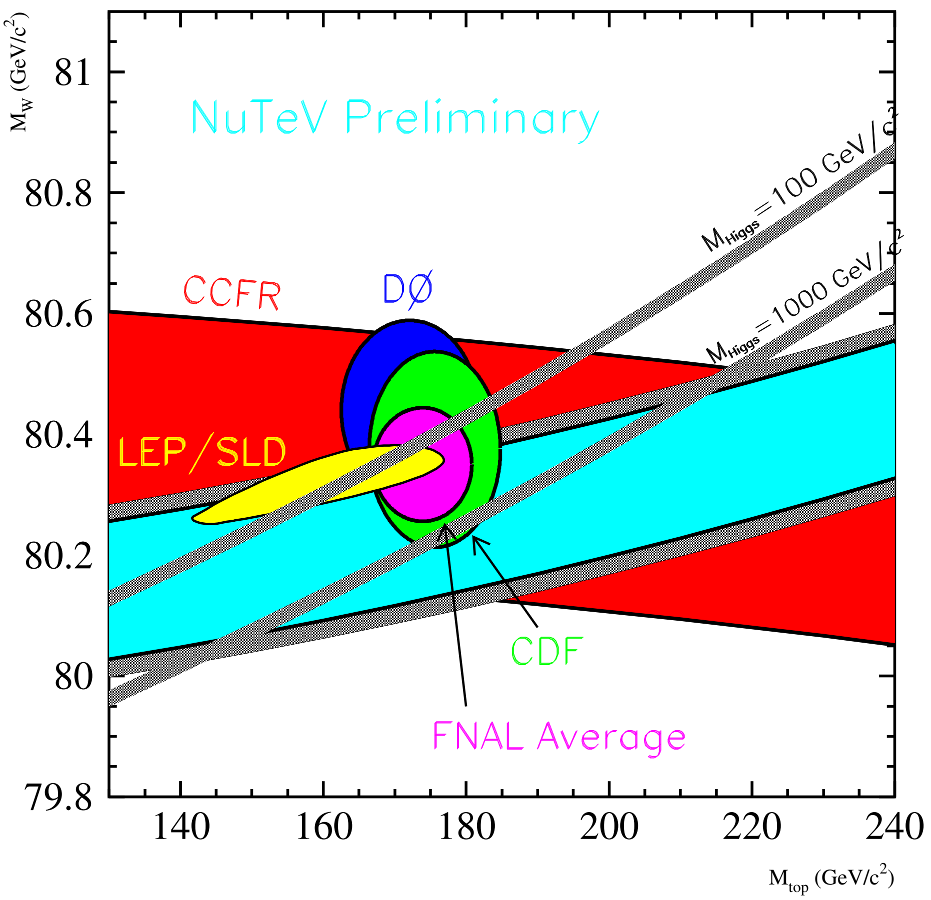

Figure 1.2: *Constraints to M_H from the measurement of M_W and M_t, plus the indirect measurements obtained from the neutrino scattering experiments NuTeV and CCFR[12].*

Experimentally, the Higgs boson has not been found yet, but the recent years have witnessed a dramatic shrinking of the allowed interval for its mass. Direct searches at LEP2 have recently increased the lower limit up to $M_H > 90\ GeV/c^2$ at 95% C.L.[11]; global fits to all electroweak observables measured at the Z pole in e^+e^- machines, in neutrino-nucleon scattering experiments, and in hadron collisions have even allowed the prediction[11] $M_H = 78^{+61}_{-38}\ GeV/c^2$, using the dependence of M_H on the other observables (see Fig. 1.1). The 95% C.L. indirect upper limit on the Higgs boson mass results $M_H < 198\ GeV/c^2$. It must be pointed out that these indirect limits are extremely sensitive to the inclusion of different results in the input to the likelihood fits[13]; nevertheless, a mass in the 100 GeV/c^2 ballpark appears by now likely—and exciting.

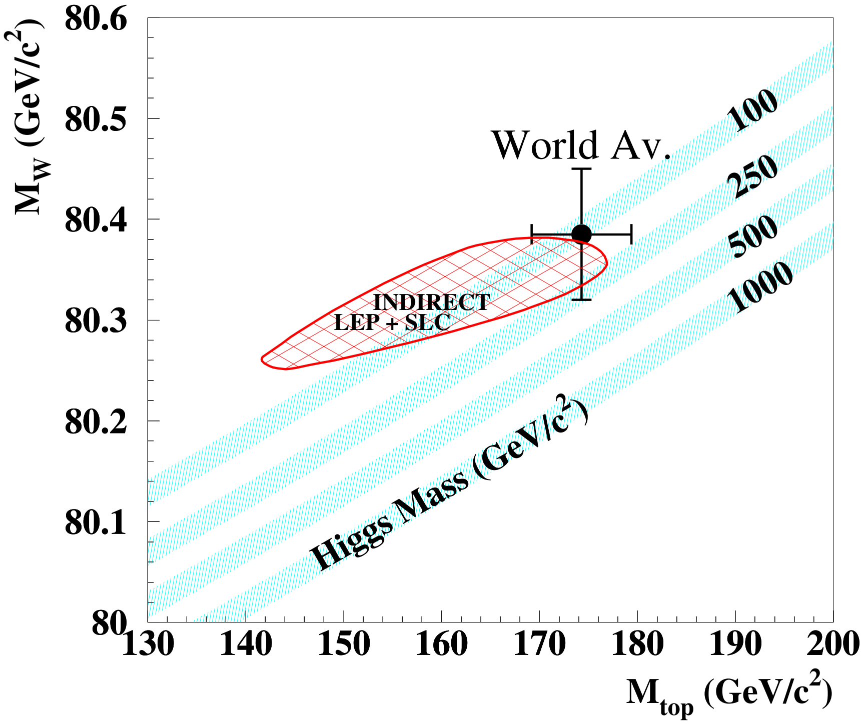

Figure 1.3: *Constraints to M_H from the current world average direct measurements of M_W and M_t[14]. The allowed zone from indirect electroweak measurements at LEP and SLC is also shown.*

After the precision measurements of electroweak observables performed at the Z pole by LEP and SLC, the W boson and top quark masses have become the critical parameters in the prediction of the Higgs boson mass, due to the logarithmic dependence of M_H on these two quantities (see Fig. 1.2, Fig. 1.3). Of course, the very slow dependence makes the game a tough one. The hope is that a direct signal of H^0 production will be obtained—either at LEP2, at the Tevatron run 2, or, if everything else fails, at the LHC—before we indirectly get to know its mass with satisfactory precision. A very clear graphic description of how the precise measurements of the

W boson and the t-quark masses can pinpoint the mass of the Higgs boson is given in Fig. 1.2 and Fig. 1.3.

1.3 Motivations of the Search for $Z \to b\bar{b}$ Decays at the Tevatron

Z decays to b-quark pairs are not exactly an empty page in the book of particle physics. Since 1992 the LEP experiments have detected several millions of them, and more have come from the polarized beams of the SLC. The process is thus perfectly well understood; the Z boson is one of the best known animals in the particle zoo, and there can be no surprise in the thereabouts. At a proton-antiproton collider the $Z \to b\bar{b}$ decay has never been seen before, though. The UA2 Collaboration published in 1987[15] (and later extended[16]) an analysis of jet data where they could extract a combined signal of W and Z decays to jet pairs, but the decay of the Z boson to b quarks was not separated from the other hadronic decays ($W \to c\bar{s}$, $W \to u\bar{d}$, $Z \to u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}$).

The "first ever" argument alone is a poor justification for spending our time on the subject. As a matter of fact, the Z decay to b quarks is the closest observable process to the dominant decay of the Higgs boson[2], and thus—in view of the expected hunt for the Higgs boson in run 2 at the Tevatron—the experimental understanding of the process, the knowledge of the expected mass resolution for a reconstructed decay, and the confidence with the kinematical tools that may help extracting it from the QCD background, are all hot topics. Unfortunately, the expected ratio of cross sections for the direct production of a H and Z boson (factoring in their branching fraction to $b\bar{b}$ pairs) is only $\sim 1.5 \times 10^{-3}$ for a Higgs mass of 120 GeV/c^2[17]. Because of the smallness of this number, it will be very difficult to look for a signal of direct H production in the $b\bar{b}$ final state with run 2 data: even if we were allowed to collect with an unbelievably high 50% efficiency all events having two tagged jets—thanks to, one may trust, the improved SVX II detector and the SVT trigger (see Sec. 9.2)—and collected 20 fb^{-1} of data, we would still end up fighting for a signal of a few thousand events drowned in a background of more than 10 million events[3] of irreducible background.

With more insight, the most important reason for a search of $Z \to b\bar{b}$ decays is to understand if a reasonably clean mass peak can be obtained in our $p\bar{p}$ data, and to learn what are the tools necessary to do it. If we can show that the signal is extractable from the nasty QCD background, we can then think of building a dedicated trigger for these events in run 2 (see Ch. 9). The collection of a high statistics[4] sample of $Z \to b\bar{b}$ decays, even if with a low signal purity, would provide

[2]The Higgs boson decays predominantly to a $b\bar{b}$ pair if its mass is smaller than ~ 130 GeV/c^2.

[3]These numbers are rough first-order estimates based on the acceptances of our search on the $Z \to b\bar{b}$ signal and the QCD background, and should be taken *cum granu salis*; nevertheless, they certainly picture the right orders of magnitude of the problem.

[4]The expected integrated luminosity of run 2 is about 20 times the total luminosity so far collected by the CDF and D0 experiments.

us with a perfect calibration line for the b-quark jets originated in $t\bar{t}$ decays: the energy spectrum of these is very similar to that of Z decay products, and therefore the b-jet energy scale uncertainty on the top quark mass measurement would be drastically reduced.

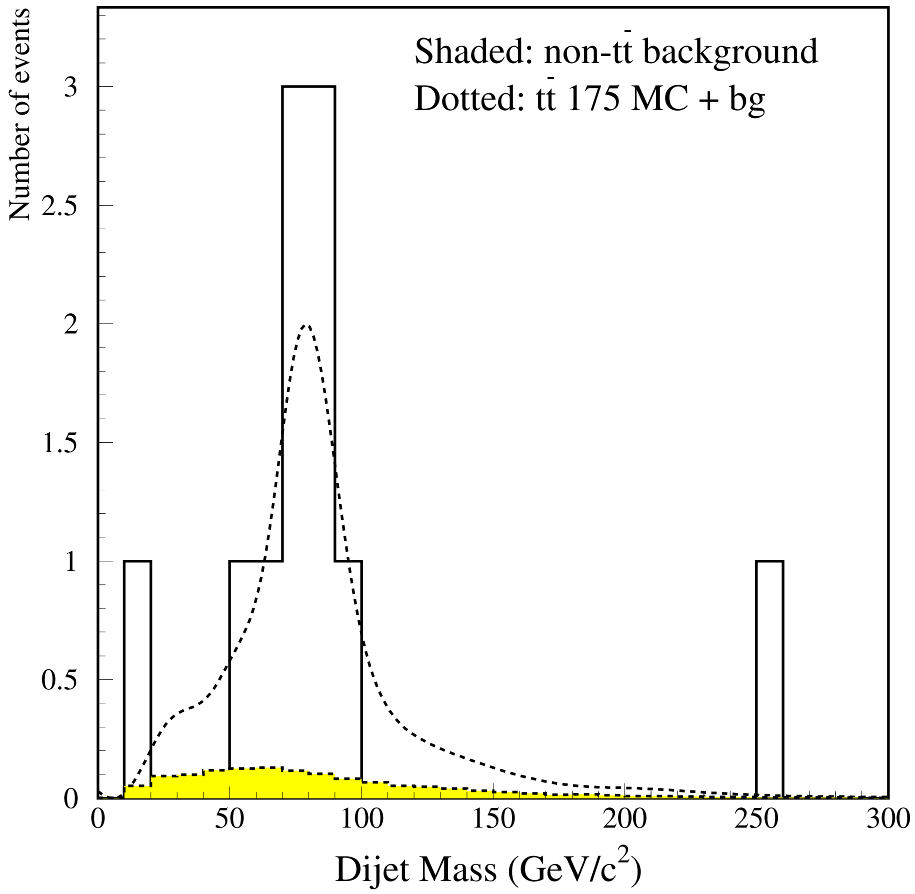

Figure 1.4: *Single lepton $t\bar{t}$ candidates with two SECVTX tagged jets allow a very clean reconstruction of the $W \to jj$ decay with the two untagged jets, thanks to the very high signal purity and the absence of any combinatoric background[18].*

During run 1 the absence of a calibration peak for generic jets as well as heavy flavor jets has caused the absolute jet energy scale be the largest single source of systematic uncertainty in the top quark mass determination[30]: the only available tool to nail down the calorimeter response to hadronic jets has been the transverse momentum balancing of Z bosons or high energy photons to a single hadronic jet (see Fig. 2.15, page 32); but the very low statistics available from these events has not allowed us to do more than estimate the systematic uncertainty. In run 2 this picture is going to change: the estimated $20\times$ gain in statistics means a lot in terms of the possibilities that will open up for the *reduction*, rather than measurement, of the systematic errors on the jet energy scale. Besides collecting 20 times more $Z+j$

and $\gamma + j$ events, a significant sample of $W \to jj$ decays—probably of the order of 300[5]—in double tagged "single lepton" $t\bar{t}$ candidates will provide an independent source of calibration for the light quark jets (compare with Fig. 1.4). Nevertheless, for b-quark jets there is no alternative to using the $Z \to b\bar{b}$ signal.

1.4 Plan of the Present Work

In the present work we will describe a search for $Z \to b\bar{b}$ decays in run 1 data triggered by the presence of a candidate muon. We will indeed show that the $Z \to b\bar{b}$ decay can be extracted by means of a very stringent selection, which allows a reduction of the background by a factor of 10^4, while retaining in the final dataset a handful of signal events.

We introduce the Tevatron accelerator and our experimental apparatus in Ch. 2. After a brief description of the Tevatron, we review the parameters of the detector components and their performance during run 1 data taking; in the second part of Ch. 2 we also describe the trigger system and the principal software algorithms commonly used at CDF to reconstruct jets, leptons, and other important physics objects.

In Ch. 3 we describe the datasets we have used for our search of Z boson decays to b-quark pairs. The selection of a sample rich in heavy flavor decays is discussed, and an estimation of the integrated luminosity it corresponds to is given. We also describe in that section the Monte Carlo datasets used for our studies of the signal characteristics and for a preliminary estimate of the amount of $Z \to b\bar{b}$ events collected in the experimental sample.

In Ch. 4 we discuss our method of selecting a sample of events where the signal fraction becomes visible. We describe the important differences between the searched signal and the experimental data, and study the kinematic variables most useful for our purpose.

A detailed study of the momentum measurement of the b-quark jets originated in the Z boson decay is described in Ch. 5. We devise a method to take advantage of the peculiarities of our selected data—the particular topology required by our kinematic selection and the presence of a semileptonic decay in one of the two jets—for a jet momentum correction. We demonstrate that these improved jet corrections allow us to reconstruct $Z \to b\bar{b}$ decays with a better resolution.

In Ch. 6 we describe a counting experiment performed with our selected dataset in order to demonstrate the presence of a sizable amount of $Z \to b\bar{b}$ decays. The method allows us to estimate the absolute number of events from background processes collected in our dataset, therefore enabling an estimation of the statistical significance of the excess we observe.

In Ch. 7 we describe an alternative method to extract the signal from the selected sample, using an unbinned likelihood fitting technique to interpret the shape of

[5]The increased acceptance of the new SVX II detector will allow a $2\times$ increase[66] in double SECVTX tagging probability with respect to run 1, and the acceptance for triggering leptons is also going to improve (see Sec. 9.2.1).

the mass distribution as the sum of signal and background. A toy Monte Carlo approach allows an independent estimate of the statistical significance of the isolated $Z \to b\bar{b}$ signal.

In Ch. 8 we attempt a preliminary measurement of the cross section for the searched process. We describe a method to estimate the acceptance of the kinematic selection, and use the number of events fitted by the likelihood procedure to estimate the Z cross section.

In Ch. 9 we present our conclusions for the analysis described in the present work. The prospects for the new collider run at the Tevatron are discussed, as well as two proposals for the collection of a $Z \to b\bar{b}$ dataset at CDF in run 2.

Chapter 2

The Experimental Apparatus

In this chapter we describe the Tevatron accelerator, the CDF detector, and the most important hardware and software tools necessary for the identification of the most interesting physics objects. Sec. 2.1 deals with the accelerator, while in Sec. 2.2, Sec. 2.3 and Sec. 2.4 is given a description of all the CDF detector components. In Sec. 2.5 we describe the data acquisition system. The rest of the chapter is devoted to the discussion of the software algorithms by which we are able to identify the signal of hadronic jets, leptons, heavy flavors, and neutrinos in our datasets.

2.1 The Tevatron

The Tevatron is a proton synchrotron built at the Fermi National Accelerator Laboratory in Batavia (IL), 35 miles west of Chicago. It is endowed with superconducting magnets and is capable of accelerating protons and antiprotons in opposite directions and colliding them at a center of mass energy of 1800 GeV[1]. The machine is hosted in a circular tunnel with a radius of 1 kilometer that was originally built for the former synchrotron, the Main Ring, a 400 GeV accelerator. These two machines have coexisted during run 1, when the 400 GeV accelerator has been used as an injector for the Tevatron; in run 2 (starting in the year 2000) a new machine, named Main Injector, will take its place; it has been built beside the Tevatron tunnel.

The accelerating process used in the Tevatron running 1992-1996 is pictured in Fig. 2.1. It begins with negatively charged hydrogen ions being accelerated to a 750 keV energy by a Cockroft-Walton electrostatic accelerator, and then up to 200 MeV inside a linac. Exiting from the linac, the two electrons are stripped off the ions when the latter traverse a thin sheet of carbon; a proton beam is thereby obtained, and can be injected in a small circular accelerator, the Booster. The Booster takes the beam to an energy of 8 GeV and collects the particles into bunches with a number density up to 2.3×10^{11} per bunch. The bunches are then yielded to the Main Ring, that is used to further accelerate them up to 150 GeV.

A proton beam is extracted from the Main Ring at 120 GeV and directed to a tungsten target: the resulting collisions yield a large quantity of antiprotons that

[1] In the next run, expected to start in April 2000, the c.m. energy will be increased to 2 TeV.

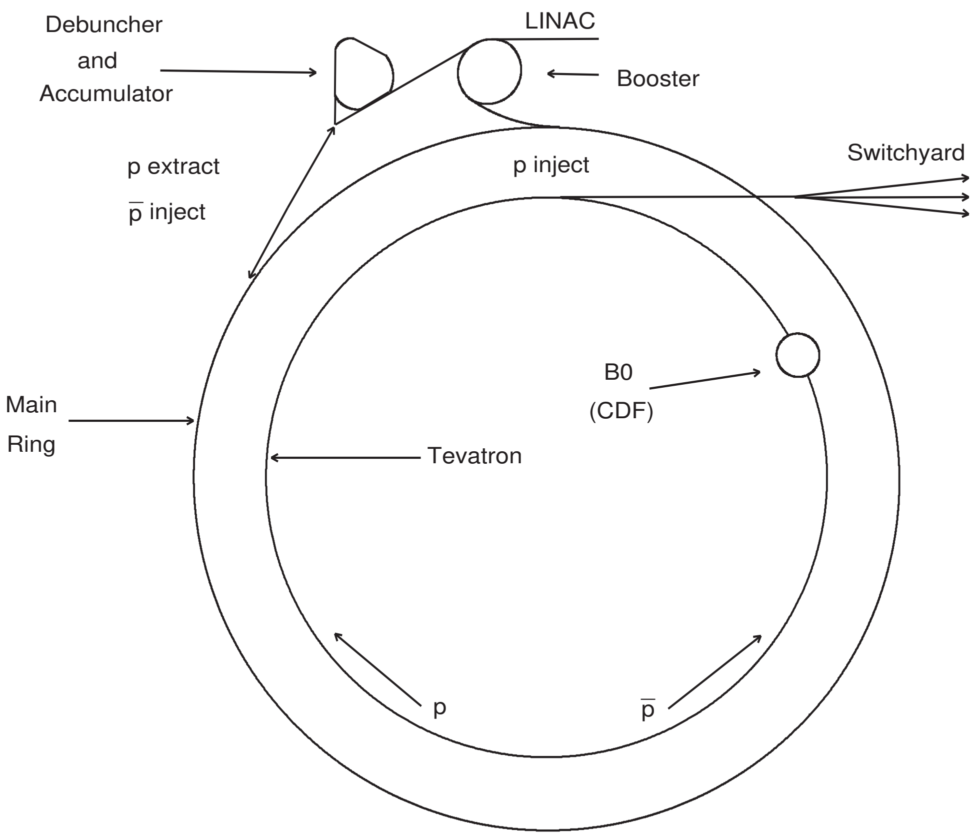

Figure 2.1: *Schematic view of the Tevatron accelerator.*

are collected and focused to a beam by passing them through a lithium lens, and then injected in a Debuncher and Accumulator. In the Debuncher the momentum dispersion of the beam is reduced by stochastic cooling, and the antiprotons can be accumulated for up to 24 hours, allowing the creation of bunches of typically 5×10^{10} antiprotons each.

The preparation for the colliding phase ends with the preacceleration of antiprotons in the Main Ring and the injection of the proton and antiproton beams in the Tevatron ring, where they are orbited in opposing directions and accelerated from 150 GeV to the maximum 900 GeV. The beams intersect each other in two points along the beam, where the two detectors CDF and D0 sit; each intersection point is surrounded by quadrupole magnets that focus the beams immediately before the crossing point, to maximize the interaction probability. The latter is in fact a function of the instantaneous luminosity, defined by

$$\mathcal{L} = \frac{N_p \times N_{\bar{p}} \times B \times f}{4\pi\sigma^2} \tag{2.1}$$

where N_p and $N_{\bar{p}}$ are the number of protons and antiprotons in each bunch, B is the bunch multiplicity per beam (equal to six during run 1), f is the revolution frequency (47 kHz), and σ the cross section of the bunches in the plane transverse to the beam. The latter depends on the phase space area occupied by the beam, the emittance $\epsilon = \int dx dp_x$.

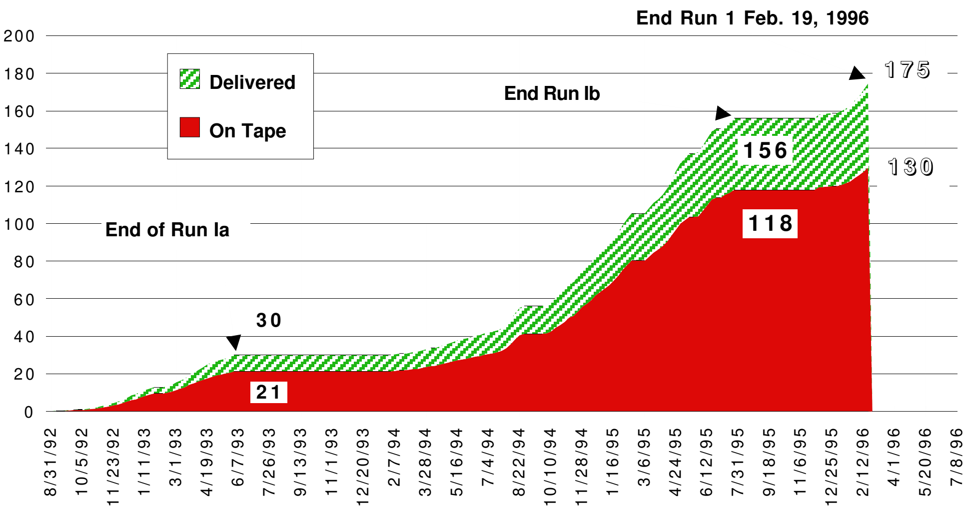

Figure 2.2: *Integrated luminosity delivered by the accelerator (upper curve) and collected by the CDF detector (lower curve) as a function of time during run 1.*

The focusing low-beta quadrupoles force a rotation of the beam in the phase space, such that the position of each particle along the orbit becomes a function of a phase ϕ and of an amplitude function $\beta(s)$. The latter is a function of the focusing action, and by its minimization in the interaction point the beam transverse area is also reduced, because

$$\sigma^2 = \frac{\epsilon\beta}{\pi}. \tag{2.2}$$

The emittance increases rapidly during the first few hours of data taking, due to the collisions and to the "beam-gas interactions", *id est* the collisions between particles in the beam and the residual gas in the vacuum chamber. The beams are restored after about 20 hours.

During run 0, in 1988–89, the instantaneous luminosity reached the peak value of $2 \times 10^{30} cm^{-2}s^{-1}$; during run 1a the luminosity reached the peak value of 9.0×10^{30}, thanks to the improved performance in the collection of antiprotons, which allowed to gather up to 7.2×10^{10} $\bar{p}$/bunch. Further improvements during run 1b have allowed us to surpass even this record: the peak instantaneous luminosity has reached $2.7 \times 10^{31} cm^{-2}s^{-1}$.

For run 2 many improvements are planned in order to increase the luminosity and therefore extend the physics reach of the CDF and D0 detectors. A new accelerator, the Main Injector, doted with superconducting magnets, will be used during the accelerating phase instead of the Main Ring. It will have a much larger phase space acceptance than the older machine, and it will allow us to inject into the Tevatron beams of up to 108 bunches each, thereby allowing to further increase the luminosity without the need to pay the price of a higher average number of collisions per bunch

C.m. energy	1800 GeV
Number of bunches	6
Typical number of protons/bunch	$7 \cdot 10^{10}$
Typical number of antiprotons/bunch	$7.2 \cdot 10^{10}$
Antiproton collection speed	$4 \cdot 10^{10}/hr$
Proton emittance, ϵ_p	$15\pi\ mm \cdot mr$
Antiproton emittance, $\epsilon_{\bar{p}}$	$18\pi\ mm \cdot mr$
β(CDF)	50 cm
Maximum luminosity reached	$2.7 \cdot 10^{31}\ cm^{-2}s^{-1}$

Table 2.1: *Tevatron performance during run 1.*

crossing[2].

2.2 The CDF Detector: an Overview

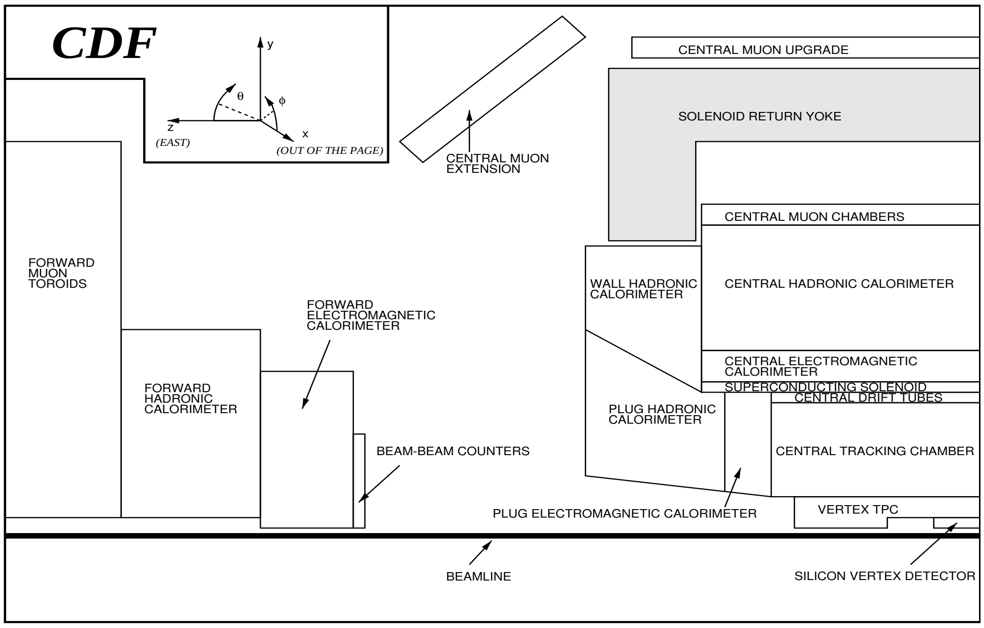

Figure 2.3: *Cutaway view of a quarter of the CDF detector. CDF is symmetric around the interaction vertex and the beamline.*

CDF[19] is a 5,000-ton magnetic detector, designed to study a wide range of physics in high energy $p\bar{p}$ collisions. It is composed of a large central structure surrounded by two spectrometers for the detection of particles emitted at small angle with respect to the beam axis. The design allows a good coverage of the solid angle surrounding the interaction vertex. Fig. 2.3 and Fig. 2.4 show a cross section and a tridimensional view of the detector, respectively.

[2]See Sec. 9.2.1.

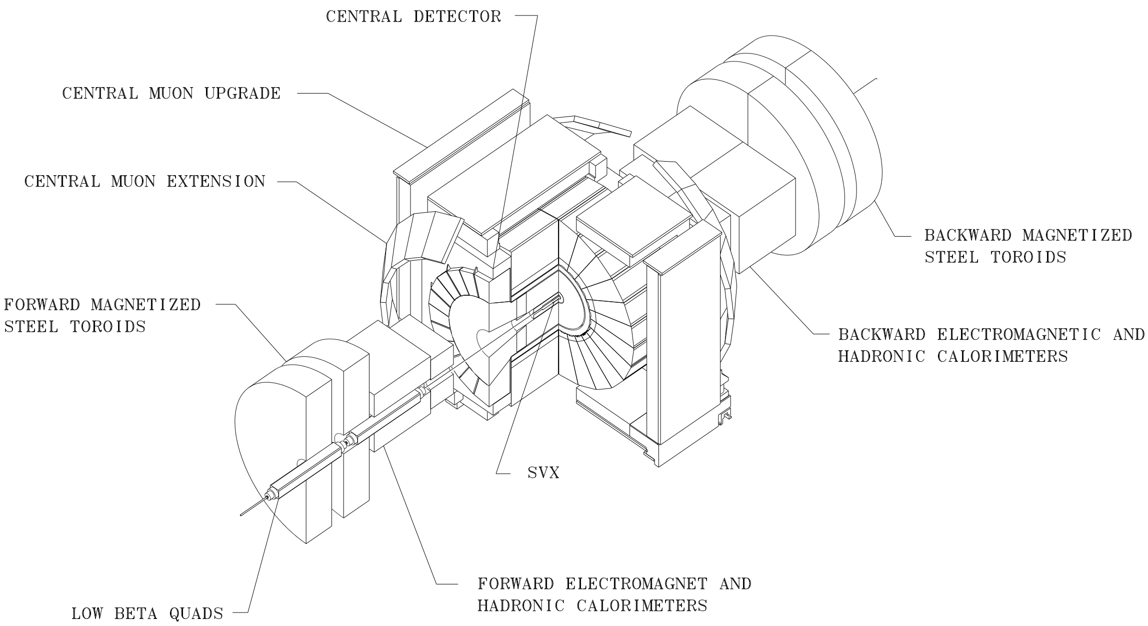

Figure 2.4: *Isometric view of the CDF detector.*

The central structure includes several detector components, the principal ones being calorimeters and tracking chambers. The latter are immersed in an axial magnetic field of 1.4 Tesla, generated by a cylindrical magnet of 2.6 m in diameter by 3.2 m in length that allows a precise measurement of the transverse momentum of charged particles.

The two forward and backward structures are composed of gas calorimeters and muon chambers. Several scintillator planes close to the beam axis, called BBC (*Beam-Beam Counters*), are used in coincidence for luminosity measurements.

Before starting a detailed description of the various detector components, it is useful to define for them the two principal coordinate systems normally used, a Cartesian and a polar system. The Cartesian system has the z axis coincident with the beamline, its positive direction being that of motion of the protons. The xy plane contains the nominal interaction point, which is the geometric center of the CDF detector; the x axis is oriented horizontally towards the outer side of the accelerator ring, and the y axis is fixed by the requirement that the xyz system be right-handed. The polar system also has its origin in the nominal interaction point; the azimuthal angle ϕ is measured in the xy plane anticlockwise, starting from the x axis, and the polar angle θ is measured with respect to the positive direction of the z axis. r is defined as the distance from the z axis in the transverse plane. The pseudorapidity η is a useful quantity at a $p\bar{p}$ collider: it is defined by the formula $\eta = -\log(\tan(\theta/2))$.

2.3 Calorimetry

The CDF calorimetry system is composed of inner electromagnetic and outer hadronic sections, segmented in cells projecting towards the geometric center of the detector. The segmentation is uniform in pseudorapidity: since η transforms linearly for Lorentz boosts along the z axis, the uniformity allows a homogeneous definition of physics objects (hadronic jets, for instance) throughout the detector.

Electromagnetic Calorimeters

	Central	End Plug	Forward
$\|\eta\|$ coverage	$0-1.1$	$1.1-2.4$	$2.2-4.2$
Cell dimensions $\Delta\eta \times \Delta\phi$	$\sim 0.1 \times 15^o$	0.09×5^o	0.1×5^o
Active medium	*Scintillator*	*Proportional tubes*	
Dimensions of active medium	$0.5\ cm$	$0.7 \times 0.7\ cm^2$	$1.0 \times 0.7\ cm^2$
Absorber	Pb	Pb	$94\%Pb, 6\%Sb$
Absorber thickness	$0.32\ cm$	$0.27\ cm$	$0.48\ cm$
Energy resolution for 50 GeV electrons	2%	4%	4%

Hadronic Calorimeters

	Central	End Wall	End Plug	Forward
$\|\eta\|$ coverage	$0-0.9$	$0.7-1.3$	$1.3-2.4$	$2.3-4.2$
Cell dim. $\Delta\eta \times \Delta\phi$	$\sim 0.1 \times 15^o$	$\sim 0.1 \times 15^o$	0.09×5^o	0.1×5^o
Active medium	*Scintillator*		*Proportional tubes*	
Dimensions	$1.0\ cm$	$1.0\ cm$	$1.4 \times 0.8\ cm^2$	$1.5 \times 1.0\ cm^2$
Absorber	Fe	Fe	Fe	Fe
Thickness	$2.5\ cm$	$5.1\ cm$	$5.1\ cm$	$5.1\ cm$
Energy res. for 50 GeV π^-	11%	14%	20%	20%

Table 2.2: *Specifications of the e.m. and hadronic calorimeters.*

The CDF calorimeters are divided in three zones, named *Central*, *End-Plug*, and *Forward*, according to the pseudorapidity coverage they provide. The central zone, covering the interval $|\eta| < 1.1$, includes the Central and Endwall calorimeters, that are segmented in towers of $\Delta\eta = 0.1$ by $\Delta\phi = 15°$. The cells are built with lead sheets sandwiched between plastic scintillators in the e.m. section[3], for a total thickness of about 20 radiation lengths; in the hadronic section they are built with a iron-scintillator sandwich, for a total thickness of about five interaction lengths.

[3] The Endwall calorimeter completes the coverage at high pseudorapidity for the hadronic sections.

Figure 2.5 shows the structure of one of the 48 cell stacks that make up the central calorimeter: the wavelength shifters system and the light guides carrying the scintillator output to the photomultiplier tubes are clearly visible, as well as the proportional tubes making up the Central Strip Chambers, that provide a precise measurement of the e.m. showers. The wire chambers are located at a depth of six radiation lengths in the e.m. section, *id est* roughly where a shower produced by a 10 to 50 GeV electron has its maximum development: the wires provide $r\phi$ readout of the location of the shower.

The Plug calorimeters are inserted at the sides of the central tracking chamber, and provide pseudorapidity coverage up to $|\eta| < 2.4$ around the nominal interaction point. Like the Central calorimeter, they are divided in cells uniformly segmented in rapidity, with an inner e.m. sampling section with lead slabs as absorber and an outer hadronic section where the absorber is iron. The active medium is provided by a 50% mixture of argon-ethane inside proportional tubes: these allow a finer ϕ segmentation in 5° degrees cells[4]. Finally, the Forward calorimeters are located in two large structures physically separated from the central part of the detector, at a distance of 6.5 m from the interaction vertex; this space is needed by the CMX arcs (Sec. 2.4.4) and to allow the servicing of the whole structure. These devices provide coverage up to $|\eta| = 4.2$[5]; they feature the same geometry, active medium, and absorbers as those of the Plug calorimeters. A sketch of the $\eta\phi$ segmentation and coverage of the CDF calorimeter system is shown in Fig. 2.6. The technical characteristics of the calorimeters are summarized in Table 2.2. The energy resolution has been computed in test-beam studies using electron and charged pion beams with energy in the range $10 \div 50$ GeV.

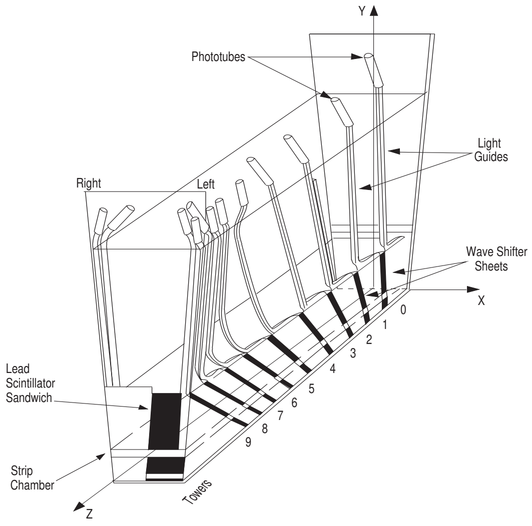

Figure 2.5: *Sketch of one of the 24 wedges of the central calorimeter.*

2.4 Tracking Devices

Charged particle detection and momentum measurement is performed at CDF using three separate detectors immersed in the axial magnetic field provided by the large solenoid. Coming out of the interaction vertex, the produced particles cross successively the SVX, or *Silicon Vertex Detector*, the VTX, or *Vertex Drift Chamber*, and finally the CTC, or *Central Tracking Chamber*, a large drift chamber that fills most of the volume inside the magnet coils.

[4]The η segmentation is still 0.1, but the physical extension of the cells is much smaller than that of the Central Calorimeter, due to the definition of pseudorapidity as $\eta = -\log(\tan(\theta/2))$.

[5]As one can notice in Fig. 2.6, part of the polar angle coverage in these devices extends only up to $|\eta| = 3.6$, to allow space for the low β quadrupoles. Because of that, in the computation of missing transverse energy, $\not\!\!E_T$, only calorimeter towers with $|\eta| < 3.6$ are used. The $\not\!\!E_T$ is discussed in Sec. 2.6.3.

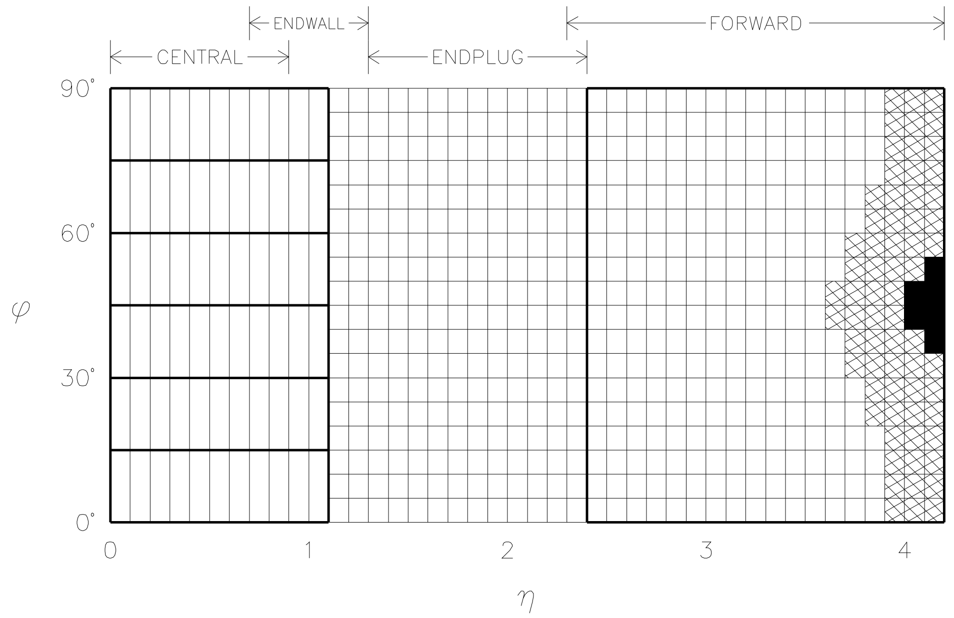

Figure 2.6: *Map of the $\eta\phi$ coverage provided by the CDF calorimeter system. In the grey area the coverage is only partial; the black area is not instrumented.*

2.4.1 The Silicon Vertex Detector

The Silicon Vertex Detector, first installed in 1992, is located immediately outside the vacuum chamber containing the beams. The vacuum chamber has a diameter of 5 cm and is built with beryllium with a thickness of only 0.05 cm, to minimize the amount of matter traversed by the particles before any measurement of position and momentum can take place. The SVX is 51 cm long, and consists of two barrels that join at $z = 0$[6]. They are composed of four concentric layers located at a distance between 3 and 7.9 cm from the z axis (see Fig. 2.7). Each layer is equipped with silicon microstrips with a pitch of 60 μm (55 μm for the inner layer) arranged in ladders as shown in Fig. 2.8. Each ladder is made up of three silicon wafers, with microstrips arranged longitudinally to provide a precise measurement of the tracks in the $r\phi$ plane. The silicon strips are microbonded together at the wafer interface and to the integrated circuits that record the signals.

A charged track traversing a silicon wafer yields a signal in a small cluster of strips (typically two or three). The crossing point, transverse to the strip direction, is found from the weighted average of the charge recorded in the strips. The resulting accuracy of the position measurement in each plane of silicon crossed by the track is about 12 μm. This allows the reconstruction of the impact parameter of the tracks[7]

[6]Due to the longitudinal dimensions of the proton and antiproton bunches, the interaction vertex shows a Gaussian spread in z with a width of about 30 cm around the nominal interaction point. The resulting acceptance of the SVX amounts to about 60%.

[7]The impact parameter is defined as the distance of closest approach between the track helix and the location of the interaction vertex. It is a fundamental parameter for the identification of secondary vertices from long-lived particle decays: see Sec. 2.6.4 for a description of the use of impact parameter in a determination of secondary vertices.

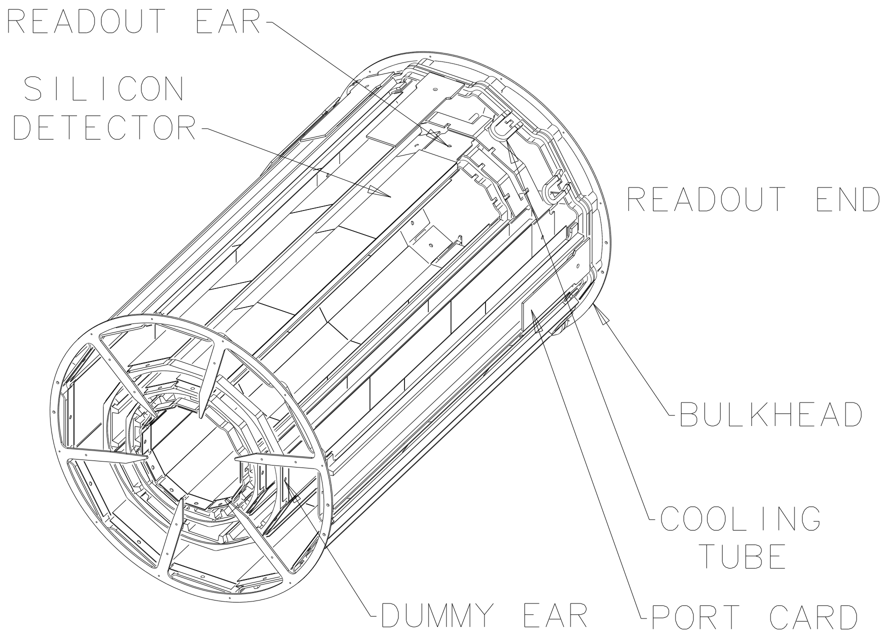

Figure 2.7: *Isometric view of a SVX barrel.*

with a precision of $(13 + 40/P_T)$ μm (with P_T in GeV/c) from a combined fit of CTC and SVX hits.

2.4.2 The Vertex Tracking Chamber

After traversing up to four layers of silicon, charged tracks produced in a $p\bar{p}$ collision in the center of CDF cross the VTX, a drift chamber that provides rz measurement of the tracks out to a radius of 22 cm for a maximum track rapidity $\eta < 3.25$ (for particles produced in the center of the detector), thus enabling the reconstruction of the primary vertex along the beam axis with a resolution of about 0.2 cm. The VTX is made of trapezoidal wedges arranged in a very lightweight carbon-fiber octagonal structure: like the silicon vertex, this chamber is designed to minimize the total number of radiation lengths traversed by particles coming from the interaction vertex, to reduce multiple scattering; moreover, a typical $p\bar{p}$ collision at 1.8 TeV produces on average 30 photons, whose conversion to e^+e^- pairs may increase the local track density and make pattern recognition in the chamber very difficult. The average number of radiation lengths traversed by a track passing through the chamber is 0.025.

The VTX is similar in design to the Vertex Time Projection Chamber[19], which it replaced after run 0, but has shorter drift regions (56 rather than 16) and a larger drift field to allow for operation at higher instantaneous luminosity. Each of the eight octagonal modules is equipped with high voltage grids that divide it into seven 4 cm-long drift regions. Electrons produced by the ionizing track drift away from the grid and are collected by the sense wires; the arrival time of the signals provide a measurement of the rz coordinate of the track. A track in the rz plane

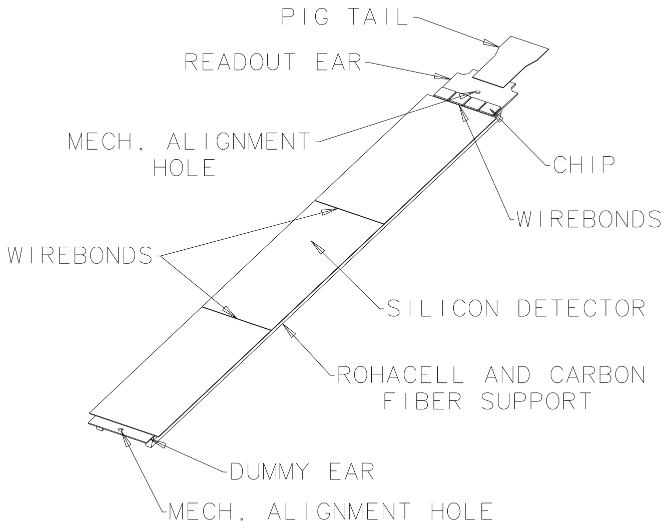

Figure 2.8: *Isometric view of a SVX ladder.*

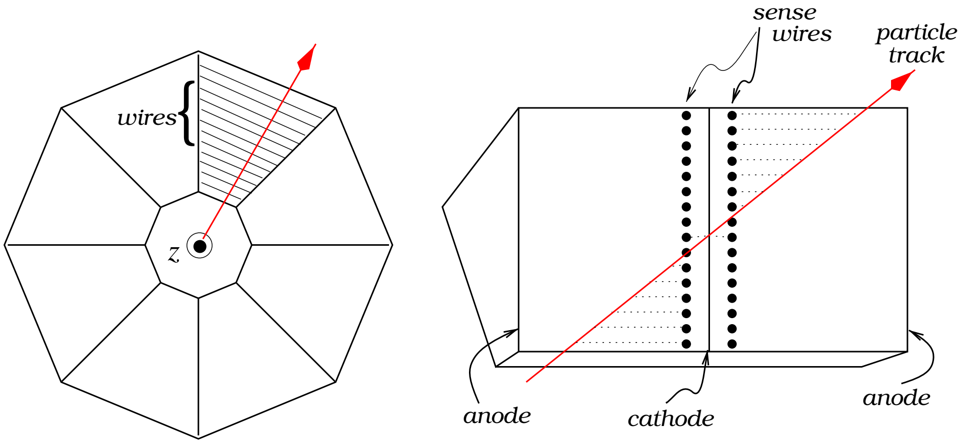

Figure 2.9: *Schematic view of a VTX wedge.*

appears to a good approximation as a straight line: the tracks are fit to reconstruct the primary vertex they belong to. The z coordinate of the primary vertex is mainly used for the computation of the transverse energy of the jets from the total energy measured in the calorimeter.

2.4.3 The Central Tracking Chamber

The Central Tracking Chamber is a cylindrical multi-wire drift chamber with a length of 3.214 m and an outer radius of 1.32 m, filled with a mixture of argon-ethane-ethanol and immersed in the 1.4 T axial magnetic field provided by the surrounding solenoid. Layers of tungsten sense wires are arranged in nine superlayers that provide 84 position measurements in the $r\phi$ coordinates; the superlayers alternate between axial alignment, parallel to the beamline, and stereo alignment, tilted by 3° to provide a measurement of the z coordinate of the hits. A view of a

Length	3.2 m
Electric field	1.6 kV/cm
Gain	$0.5 \cdot 10^4$
Number of modules	28
Number of channels	8412
Coverage	$\|\eta\| \leq 3.5$
Maximum drift distance	4 cm
Stereo angle	11.3^o
Primary vertex resolution (z)	2 mm

Table 2.3: *Technical characteristics of the VTX chamber.*

CTC endplate is shown in Fig. 2.10; Table 2.4 lists the principal characteristics of the CTC design.

Length	3.214 m
Inner radius	0.287 m
Outer radius	1.32 m
Number of superlayers	9
Number of layers	84
Total sense wires	$30,504$
Maximum drift distance	4 cm
Stereo angle	$\pm 3^o$
Spatial resolution ($r\phi$)	$<$ 200 μm *per wire*
Spatial resolution (z)	$<$ 0.200 $mm/\sin 3^o = 4$ mm
Momentum resolution	$\delta P_T/P_T^2 \leq 0.002\ (GeV/c)^{-1}$

Table 2.4: *Characteristics of the CTC.*

Charged particles crossing the CTC volume leave a trail of ionization electrons that drift in the crossed electric and magnetic fields for a maximum of 4 cm along the ϕ direction towards the sense wires. Near the wires the electrons accelerate in the $1/r$ field and produce secondary ionization; the wires collect the charge and carry it to one end where the signal is amplified. The position in ϕ of the track at the wire radius is inferred from the arrival time of the signal.

The single hit resolution in the wires is 200 μm in $r\phi$ and 4 mm in z (stereo layers only). The transverse momentum resolution for a track in the fiducial volume is $\delta P_T/P_T = 0.002\ P_T$ (CTC only); together with the SVX, the CTC can measure charged particle transverse momenta with a precision $\sigma_{P_T} \sim \sqrt{0.0066^2 + (0.0009 P_T^2)^2}$ (with P_T in units of GeV/c). The pulse heigths from the sense wires in superlayers 4, 6 and 8 can be used to obtain a measurement of the ionization energy loss undergone by the track: since that quantity is a function of $\beta\gamma$, a measurement of the particle mass can be attained by combining the ionization measurement with the knowledge of the track momentum. The resolution in mass is very scarce, but a statistical discrimination of electrons from hadrons is possible.

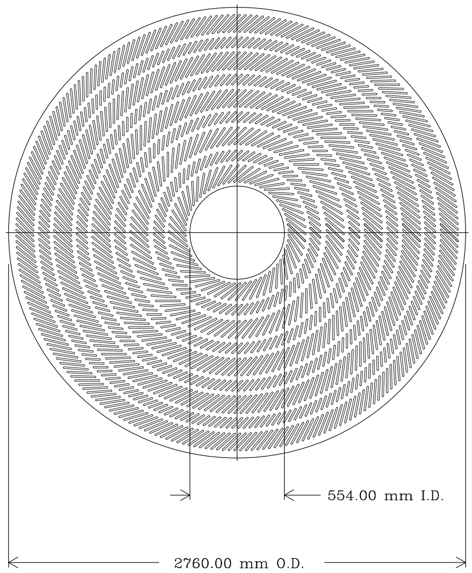

Figure 2.10: *Schematic view of a CTC endplate. The structure of superlayers is visible.*

2.4.4 The Muon System

Muon detection at CDF is provided in the central region by three subsystems: the Central Muon Detector (CMU), the Central Muon Upgrade (CMP), and the Central Muon Extension (CMX). The CMU and CMP provide coverage up to $|\eta| = 0.6$, the CMX extends the coverage to a rapidity of ± 1.1. In addition, two large iron toroids, the Forward Muon System (FMU), provide additional coverage for muons emitted at small angle ($3°$ to $16°$) from the beam.

The CMU system is composed of four layers of drift tubes filled with a $50\% - 50\%$ mixture of argon and ethane, operated in limited streamer mode. The chambers are located on top of the hadronic calorimeter wedges, at a distance of 3.47 m from the beamline and a total depth of 5.4 pion interaction lengths. They are segmented in ϕ into 24 wedges $12.6°$ wide, composed of three $4.2°$ modules each; this arrangement leaves a gap of $2.4°$ between each wedge. Each module consists of four layers of rectangular drift cells, 2.26 m long and with a cross section of 63.5×26.8 mm^2. Muons with transverse momentum greater than 1.5 GeV/c are able to traverse the whole calorimeter and leave a signal in the CMU cells. Their position can be completely determined, since the ϕ coordinate is measured by the drift time to the anode wire with a resolution of 250 μm, and the z coordinate is obtained from charge division in the wire with a resolution of 1.2 mm.

Outside the CMU system, muons with momentum greater than 2.7 GeV/c manage to pass through a layer of 60 cm of steel (equivalent to three additional pion interaction lengths) and reach the CMP system. The CMP is composed of four additional layers of drift chambers, arranged in a box structure around the central detector; the mismatch between the outer box and the inner cylinder creates a curved acceptance boundary, as shown in Fig. 2.11. The chambers have a cross section of $2.5 \times 15\ cm^2$ and a length of $6.4\ m$; they are run in proportional mode, with a maximum drift time of $1.4\ \mu s$. The combined signal from the CMU and CMP chambers provides a powerful reduction of the background.

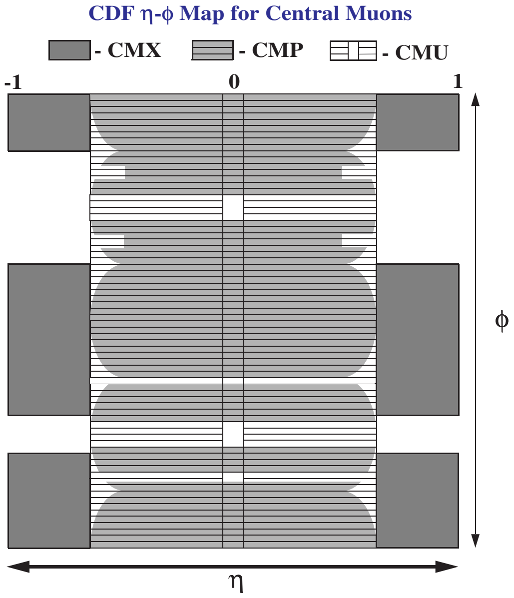

Figure 2.11: *$\eta\phi$ Coverage of the central muon system.*

The CMX system is composed of a standalone structure located between the Central and Forward detectors. It is made of four conical sections of drift tubes (CMX) sandwiched between two layers of scintillation counters (CSX) and extending in polar angle from 42° to 55°. The ϕ coverage is partial, due to a 30° uninstrumented region at the top of the detector and a 90° gap where the arches meet the floor of the collision hall. The drift tubes are identical in shape to those of the CMP system, but are only 180 cm long. At their radial distance from the interaction vertex the tubes cover 2.5° in ϕ each. They are arranged into eight layers and are half-cell staggered to eliminate tracking ambiguities. A front and side view of a CMX arch is shown in Fig. 2.12.

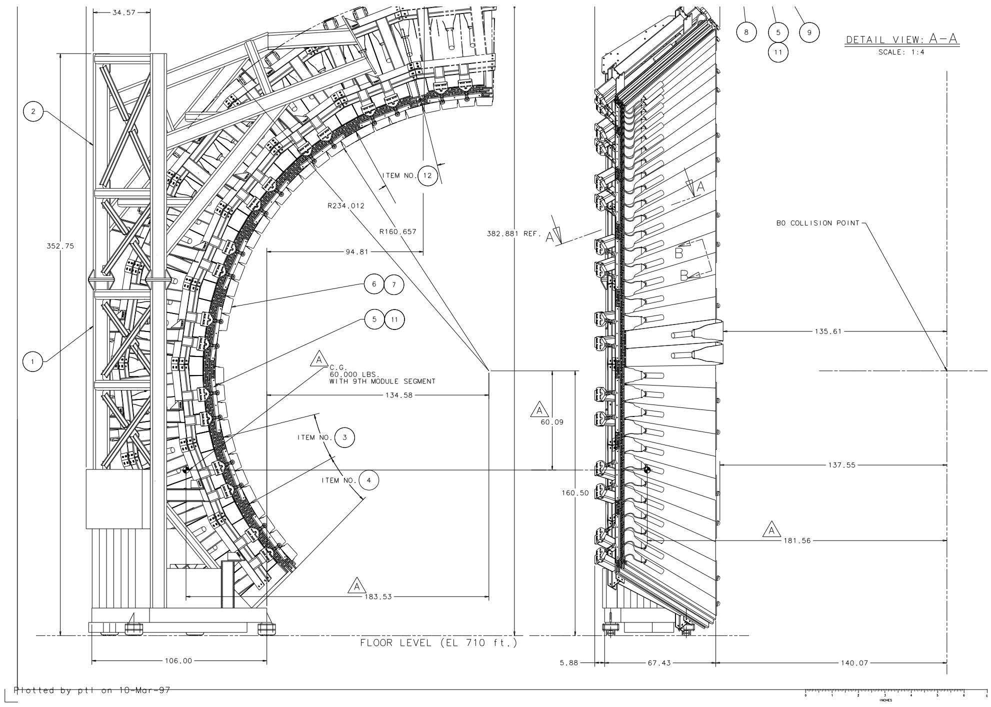

Figure 2.12: *Front and side view of a CMX arch.*

The region at high rapidity is instrumented with two large iron toroids of magnetized iron ($B \sim 1.6 \div 2.0\ T$) with a width of 2 m. Each toroid is covered with three layers of drift chambers alternated to two scintillator planes, and provides a measurement of the track position in the $r\phi$ plane: the resolution attainable on track momentum is about 13%, while the position resolution is of about 200 μm in the radial direction and of 5° in the ϕ angle. During run 1 these chambers have suffered high backgrounds from beam-gas interactions and have been seldom used for data analysis.

2.5 The Trigger System

The trigger system at CDF is a collection of hardware modules and speed-optimized software programs, designed to select the most interesting events for physics analysis among many hundreds of thousands of $p\bar{p}$ collisions taking place every second in the core of the detector. At a center-of-mass energy of 1.8 TeV the total cross section for $p\bar{p}$ scattering is about 80 mb[20], which means, at a typical instantaneous luminosity of $5 \times 10^{30}\ cm^{-2}s^{-1}$, that about 400,000 interactions are produced every second. On the other hand, the typical cross section scale for processes of high interest at a hadron collider like the Tevatron can be set, for the sake of comparison, by the inclusive production of W bosons, which corresponds, at the same luminosity and energy, to 0.1 events per second[52].

The rarity of interesting and exotic processes alone would not be a reason for discarding the most common events. The number of electronic channels that have to be read and stored per event is about 10^5: this number, once interpreted by the trigger algorithms, is converted into about 80 kilobytes of information. The storage speed allowed the recording of a maximum of ten events per second during run 1: a very selective and fast system is thus needed to reduce the rate by more than four orders of magnitude; on the other hand, the huge reduction in rate that the trigger system must achieve demands obviously a very sophisticated and careful choice of the selection criteria.

Data selection is operated by three trigger levels in series, each feeding the next one; they are physically located in the first, second, and third floor of the CDF building, above the collision hall: good events thus "percolate" up to the third level, where the final storage occurs. The first and second level triggers are managed by FASTBUS electronic modules, while the third level trigger is made up with an array of 60 Motorola 68020 processors that allow a complete reconstruction of the events by software algorithms in FORTRAN language. The idea behind a multi-level trigger system is to operate a very conservative selection at the lower levels, with the aim of reducing the rate to a value that allows a more sophisticate processing at the following level without the introduction of significant deadtime. The slowest process in data acquisition at CDF is the complete readout of the various devices, which necessitates about a millisecond to be carried out completely. The readout begins after a Level 2 accept signal has been issued; this must happen at a rate of no more than 100 Hz if the deadtime due to detector readout is to be kept below

10%[21].

At Level 1 the FASTBUS modules have about 300 ns to decide if the event being analyzed should be passed to the following level[8]. To achieve that, a very approximative readout of the detector is performed. Calorimeters—the fastest devices in the whole detector—are read out by adding the energy measured in pairs of towers neighboring in rapidity (trigger towers), and scaled by a factor $sin\theta$ (θ being the angle between tower and beamline, viewed by the detector center) to obtain an estimate of their transverse energy[9]. The event is passed at the second level if one or more trigger towers are above a predetermined threshold which depends on the type (e.m. or hadronic) and location of the tower: that allows us to collect events with high energy jets. Events with muon candidates are passed to the second level if a signal in one of the muon chambers is consistent with a muon track of moderate transverse momentum ($\sim 6\ GeV/c$ in the CMU/CMP chambers, or $\sim 10\ GeV/c$ in the CMX chambers). Timing information in different layers of muon chambers is used to get the azimuthal crossing angle of the track, which can be compared to the expected angle for a straight line originating from the center of the detector to give a rough measurement of track momentum: the crossing angle is due to the bending inside the solenoidal field before the particle reaches the chambers. Possible electron or photon candidates are collected by accepting events with a low E_T deposit in a trigger tower of the electromagnetic calorimeter.

When none of the requirements of Level 1 are met, the readout memory is cleared, to allow for the storage of information from the following bunch crossing. Otherwise, data acquisition is temporarily disabled, and the existing information—together with the map of CTC hits—is passed to Level 2, where the FASTBUS modules perform two main tasks: the reconstruction of energy clusters from the calorimeter towers, and the pattern recognition of CTC signals to identify charged tracks and estimate their momenta.

Hardware reconstruction of jets is performed by a dedicated algorithm that sorts out the highest energy towers and searches for additional E_T depositions in the neighboring ones; the search is performed in the $\eta\phi$ lattice iteratively, until no more towers over a predefined E_T threshold can be added to a cluster. The algorithm thus obtains a list of energy clusters and a fair estimate of their transverse energy, so that stringent requirements can be applied to their number and E_T .

Pattern recognition is performed by the CFT, or *Central Fast Tracker*, a fast algorithm that searches for hits in the CTC layers aligned as expected for charged particles having P_T greater or equal to one of five different values, ranging from 2 to 12 GeV/c. These tracks are then associated to candidate electron signals in the calorimeters and to candidate muon hits in the outer chambers. For electrons, at Level 2 the signals in the CES chambers are used as a further discrimination, and

[8]Most of the time interval between two successive bunch crossings—3.5 μs—is spent for the readout of calorimeters and the transfer of data to the control room, where the hardware modules are located.

[9]Of course, due to the variable energy of the hard subprocess, which depends on the Bjorken x of the interacting partons, *transverse* energy is the critical variable for the selection of interesting processes at a $p\bar{p}$ collider.

for photon candidates a veto on charged tracks in the vicinity of the CES deposit is applied.

A large number of different triggers are then fed with the information from these algorithms, and the accepted events are passed to the third level trigger, where a complete reconstruction of the event is performed by software algorithms optimized for speed. The Level 3 triggers validate the decisions operated by their corresponding Level 2 counterparts, in most cases tightening the requests on variables that can now be known with higher precision, to reduce the rate by another order of magnitude and thus allow the data to be stored.

The events are stored and then divided in different data streams according to the triggers that accepted them. A data stream is in most instances a collection of a number of different triggers with similar characteristics: accordingly, all muons with intermediate momentum (P_T thresholds at 6 to 12 GeV/c) are collected in a "Inclusive Muon Dataset" regardless of the particular requirements they fulfilled to be accepted.

The general criterion for the design of the different triggers is the versatility of the collected data; in fact, besides its primary purpose—the recording of interesting physics processes—many a trigger is designed to also provide a calibration tool or a monitor of other triggers. An example of the former is given by the low E_T jet triggers, whose rate is totally dominated by QCD $2 \rightarrow 2$ processes: their main use is the determination of the inclusive jet cross section[22], but they are also very useful in the determination of the relative scale in the calorimeter energy (see below, Sec. 2.6.1); an example of the latter is given by those high E_T "electron" triggers that only require an e.m. deposit in the calorimeter (with no track requirement): they efficiently collect photon events, but are also essential to provide a clear understanding of track reconstruction efficiency for ordinary electron triggers that do require a matching track for the calorimetric deposition.

Of course, the high rates due to low thresholds and loose requirements have to be reckoned with: the so-called *prescaling factors* take care of this. A fixed ("static") prescaling factor is used to define how often an event passing a high-rate trigger can be accepted: if the prescaling factor is x, then on average only one every x events will be picked among those fulfilling the trigger requirements. Static prescaling therefore prevents the data acquisition system to get drowned by high rate triggers. As a benchmark, the inclusive jet production process discussed above gives cross sections as high as hundreds of microbarns for the lower E_T thresholds[10]; thus, since a 10 Hz budget at Level 3 implies that at the typical instantaneous luminosity of $10^{31} cm^{-2} s^{-1}$ the cross section of all triggers must not exceed one microbarn, these triggers need static prescaling factors as high as $x = 1000$.

Prescaling is also used to protect data acquisition from a different pitfall: since the rate of any given process is luminosity dependent ($N[Hz] = \sigma[cm^2]\mathcal{L}[cm^{-2}s^{-1}]$), while the total budget is fixed, one needs to adjust the trigger accept rates accordingly. If this were not done, the system would suffer considerable deadtimes at high luminosity, making the accelerator improvements totally pointless. The less inter-

[10] CDF collects jet data with four inclusive jet triggers: their corresponding thresholds are 20, 50, 70, and 100 GeV.

esting triggers are therefore provided with a variable ("dynamic") prescaling factor, that is modified during data taking according to the instantaneous luminosity: this allows us to keep the deadtime at a level not exceeding $10 \div 15\%$. A typical shape of the variation of these dynamic prescaling factors with instantaneous luminosity is shown in Fig. 2.13.

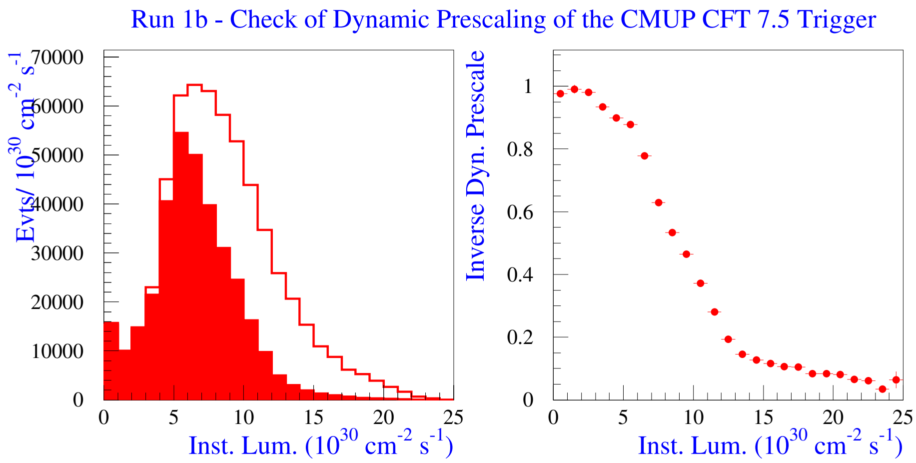

Figure 2.13: *In order to study the dynamic prescaling of the* 7.5 *GeV/c muon trigger (active during run 1b), the luminosity distribution of events accepted by the* 12 *GeV/c trigger (left plot, empty histogram) and of events accepted by both the* 7.5 *and* 12 *GeV/c triggers (full histogram) can be compared. Since a muon accepted by the higher* P_T *trigger passes by definition also the lower* P_T *one, the ratio of the two histograms gives the inverse of the dynamic prescaling and its average luminosity dependence (shown on the right).*

2.6 Reconstruction of Physics Objects

2.6.1 Jet Reconstruction

We have described in Sec. 2.5 how energy clusters are reconstructed by the hardware cluster finder at Level 2. When all the information from the various detector components is available, and processing time is no longer a constraint, the reconstruction of jets can be made more precise. CDF uses for that purpose a cone algorithm[11], called JETCLU[23], to reconstruct hadronic jets from the energy deposits in the calorimeter towers and thus allow the measurement of energy and direction of the

[11]Technically, rather than of a cone one should speak of a circle drawn in the $\eta\phi$ plane. Pseudorapidity transforms linearly for Lorentz boosts, so jets are expected to maintain a circular section in these two variables, while in the normal xyz space they betray a variable elliptical shape.

originating final state quarks and gluons. The radius of the cone in $\eta\phi$ space inside which the energy is clustered can be chosen from the three values 0.4, 0.7, and 1.0. A large radius allows one to include in the computation of jet characteristics those particles emitted at large angle from the outgoing parton, but has the disadvantage of a lower resolving power for pairs of jets close in angle. When one needs to reconstruct a high jet multiplicity final state, as in the case of $t\bar{t}$ decay, the 0.4 cone is the usual choice; for most other analyses the radius $R = 0.7$ is chosen, as recommended at the 1992 Snowmass Workshop[24].

JETCLU associates to every calorimeter tower a tridimensional vector in $R\eta\phi$, with origin in the event vertex and pointing towards the baricenter of the energy deposition in the calorimeter tower[12], and with length proportional to the transverse energy of the tower. Vectors of length greater than 1 GeV (seed towers) are inserted in a list in descending order of E_T; the list is scanned downwards, grouping together all vectors as follows:

1. a circle of radius R is drawn around the first tower in the list;
2. all vectors pointing inside that circle are associated to it and removed from the list;
3. a new circle is drawn around the next vector in the list;
4. steps 2 and 3 are repeated until all vectors are removed from the list of seed towers.

The above procedure is called *preclustering*. The real cone algorithm phase starts thereafter: to each circle are associated all the vectors with energy larger than 0.1 GeV falling inside it, and all vectors belonging to the same circle are added to compute the resultant vector. New circles are then drawn around the resultant vectors, thus including new tower vectors and losing some; new resultant vectors are then computed. If two circles are intersecting and share some vectors, the latter are associated to the closest resultant if they total less than 75% of the energy of the smaller of the two resultants; otherwise, the algorithm replaces them with a single circle, centered around the total resultant. The whole procedure is iterated until a stable configuration is reached for the circles: that is, until they are reconstructed in the same position for two consecutive iterations. This happens typically after three iterations.

The output of the algorithm is thus a list of energy clusters, their energy, and direction. These clusters are commonly agreed to be called *jets* if their energy is high enough to allow their unambiguous interpretation in terms of the result of quark or gluon fragmentation; the commonly used minimum value for the transverse energy

[12]Each tower has an electromagnetic and a hadronic section; to evaluate the radial depth of the energy deposition the assumption is made that the energy has been released at a depth of six interaction lengths I° in the e.m. section, and at a depth of ten radiation lengths λ_0 in the hadronic section: those quoted are the average values expected for hadronic jets. The radial depth of the energy deposition is used when the true z vertex coordinate is used to reconstruct clusters: for $z \neq 0$, the effective rapidity of the energy deposition is a function of its assumed radial depth.

of a jet is 10 to 15 GeV. Below that value, the correlation between the measured quantities (energy, azimuth, and rapidity) and the actual figures of the originating final state partons start to vanish, due to finite resolution and calorimeter noise.

The Standard Jet Energy Corrections

The measured energy of a jet is nothing but the sum of the raw response of all towers belonging to the cluster. If this quantity has to be used to estimate the quadrimomentum of the originating parton, it is mandatory to account for all the effects that cause systematic shifts to its measured value. First of all one must remember the two main problems of the CDF calorimeter system: its nonuniform nature, due to the use of two different active media (plastic scintillator in the Central and gas in the Plug and Forward detectors), and the imperfect coverage of the solid angle around the interaction vertex, primarily at the interface between the Central and Plug calorimeters, where as much as 50% of a jet's energy can manage to escape through the crack undetected. The accuracy of the energy measurement is also greatly affected by the insufficient response of the hadronic compartments to soft hadrons: this causes an "energy scale" effect, *id est* a downward shift of the measured energy. Secondly, one must keep in mind that a jet energy measurement based on the sum of energy deposited inside a cone of fixed radius is systematically in defect, since some low momentum particles may be emitted at large angle from the jet axis, and fail to contribute to the cluster: this is called "out-of-cone" loss; charged particles may also fail to reach the calorimeter because of their low momentum and the high solenoidal field in the central tracking volume.

Two effects may give a positive shift to the measured energy in a cluster: the underlying event and the additional hard or semihard interactions occurring during the same bunch crossing. The underlying event is due to the spectator partons inside the two colliding hadrons: their recombination into colorless hadrons usually yields a soft radiation flux in the direction transverse to the beam; this radiation may travel in the same direction of a jet, and add energy to it. Additional interactions in the same bunch crossing that originated the triggering event are just as much unavoidable: the higher the instantaneous luminosity, the more additional interactions one is likely to collect at each crossing. Most of the additional interactions give a negligible contribution to the energy flow; but their cumulative effect can become quite substantial at high instantaneous luminosity.

Finally, one must cite the problems connected to our understanding ot the fragmentation mechanisms of the partons, which determine their appearance as sprays of hadrons and their problematic measurement in the calorimeter. In fact, our chance of simulating the energy measurement—and thence of being able to calibrate it to yield a precise estimate of the energy possessed by the originating parton—depends on our understanding of the nonperturbative phase of the parton shower.

To account for all the factors that may affect the measurement of jet energy the CDF Collaboration has devised a correction routine, QDJSCO[25]. The routine works in two steps. First, a geometric correction is applied to take care of calorimeter nonuniformities and of energy lost in uninstrumented regions, by transforming the measured jet momentum in what would have most likely been measured if the jet had hit the central region of the detector—the best studied and better calibrated system. After that, all other effects are dealt with, to scale the measured energy to the best estimate of the true parton energy.

To construct a relative correction function CDF uses dijet events collected with the jet triggers, where one of the jets has detector pseudorapidity[13] in the interval $0.2 < |\eta_d| < 0.7$, corresponding to a leak-free region of the Central detector[14]. This jet is named *trigger jet*, since it is the one most probably responsible for the acquisition of the event; the other is called *probe jet* (if both jets have $0.2 < |\eta_d| < 0.7$ the naming is assigned randomly). Transverse momentum conservation implies that the two jets must balance in P_T: therefore the measured momentum of the probe jet can be checked by using the well-measured trigger jet. In order to do that, the missing transverse energy ($\not\!\!E_T$) is used[15], with which the ratio

$$MPF = \frac{2\ \vec{\not\!\!E}_T \cdot \hat{P}_T^{probe}}{P_T^{trigger} + P_T^{probe}} \tag{2.3}$$

is computed. Since, on average, $\vec{\not\!\!E}_T \cdot \hat{P}_T^{probe} = P_T^{trigger} - P_T^{probe}$, it follows that

$$MPF = \frac{P_T^{trigger} - P_T^{probe}}{P_T^{trigger} + P_T^{probe}} \tag{2.4}$$

holds. One can therefore construct a correction factor β as follows:

$$\beta = \frac{P_T^{trigger}}{P_T^{probe}} = \frac{2 + MPF}{2 - MPF}. \tag{2.5}$$

The second expression for MPF (Eq. 2.4), rather than the first, is used to correct jet momenta, since it is possible to show that β, being independent from the clustering algorithm used for the computation of measured jet momentum, yields a more precise correction[26].

The values of the relative correction are shown as a function of rapidity in Fig. 2.14 for five bins of jet E_T: one may clearly see the effect of cracks at the detector interfaces, as well as the claimed nonuniformity of the system response.

Downstream to the geometric correction an absolute correction must be applied, to account for the low calorimeter response to hadrons, the finite detector depth, and the particles not reaching the calorimeter. This correction has to be monitored during data taking and constantly adjourned, since even small variations in the calibration of the energy response may yield conspicuous systematic shifts in derived quantities such as the jet cross section[16].

---

[13]Pseudorapidity with respect to the detector center ($z = 0$) is labeled η_d, to distinguish it from the true physics pseudorapidity η of a reconstructed object, which is computed with respect to the real interaction vertex.

[14]Small uninstrumented regions are located at the interface between the two halves of the central detector at $\eta_d = 0$ and at its two boundaries $\eta_d = \pm 1.1$.

[15]The missing transverse energy, $\not\!\!E_T$, is described in Sec. 2.6.3.

[16]Due to the rapid falloff of the inclusive jet cross section with jet energy, a 1% error in the energy measurement causes a 5% or larger error in the measured cross section[27].

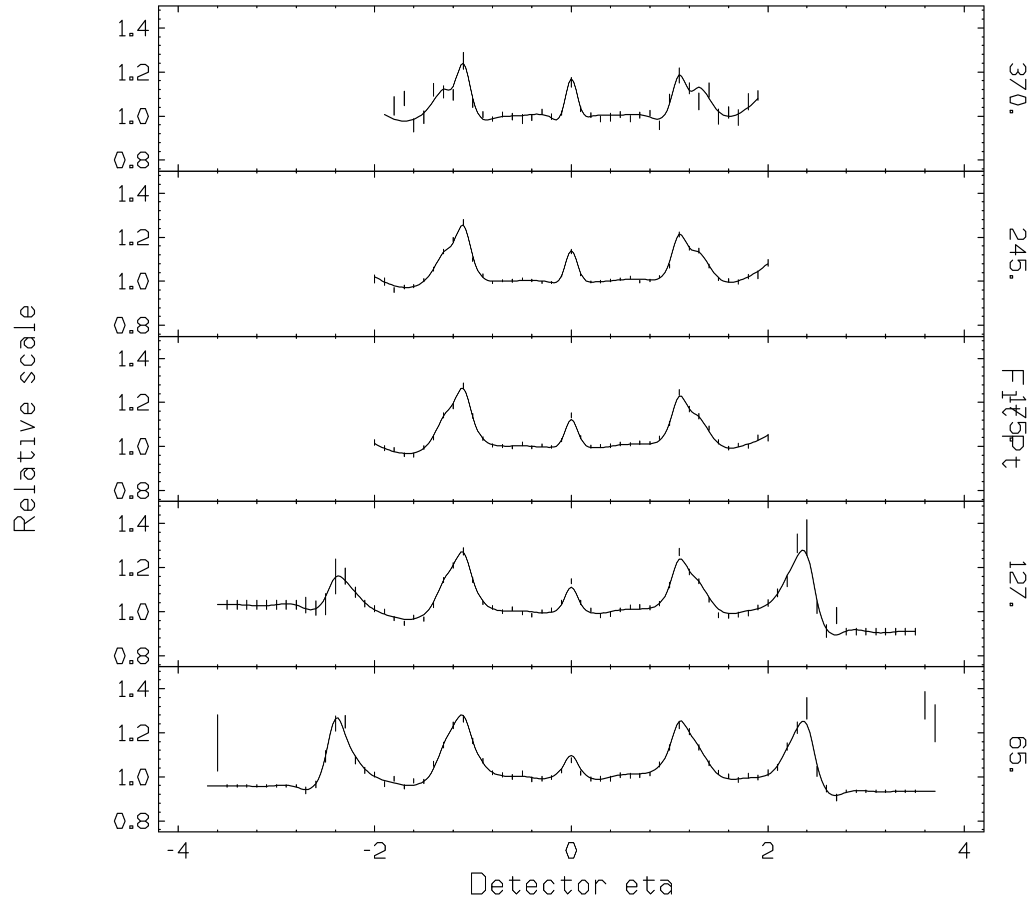

Figure 2.14: *Correction functions for the pseudorapidity dependence of the jet momentum measurement. Five different jet-P_T bins are displayed here (from top to bottom: 370, 245, 175, 127, and* 65 *GeV).*

To obtain an absolute correction of the jet energy scale CDF uses a simulation of the parton fragmentation, SETPRT, whose parameters are calibrated to reproduce the jet fragmentation function measured in the 1988-89 dataset[28]; a complete simulation of the calorimeter response is then applied using the QFL$'$ package[29]. The ratio between the true momentum of all particles included in the jet cone and the measured momentum is then computed:

$$\alpha(P_T^{jet}) = \langle \frac{P_T^{cone}}{P_T^{jet}} \rangle. \tag{2.6}$$

The values of α are used for a correction of the absolute energy response in the calorimeters for generic hadronic jets.

After the absolute correction, the user can choose to operate an additive correction to account for the energy lost out of the clustering circle. This correction is estimated from Monte Carlo and is also a function of the jet P_T.

Finally, it is possible to subtract from the jet energy the average expected contribution due to additional interactions occurring in the same bunch crossing. This is computed using the number of good primary vertices[17] reconstructed in the event: the higher this number, the larger the expected contribution inside the clustering circles. This correction, as well as the out-of-cone correction, is of course unable to increase the jet energy resolution, but at least prevents the jet measurement from depending on data taking conditions.

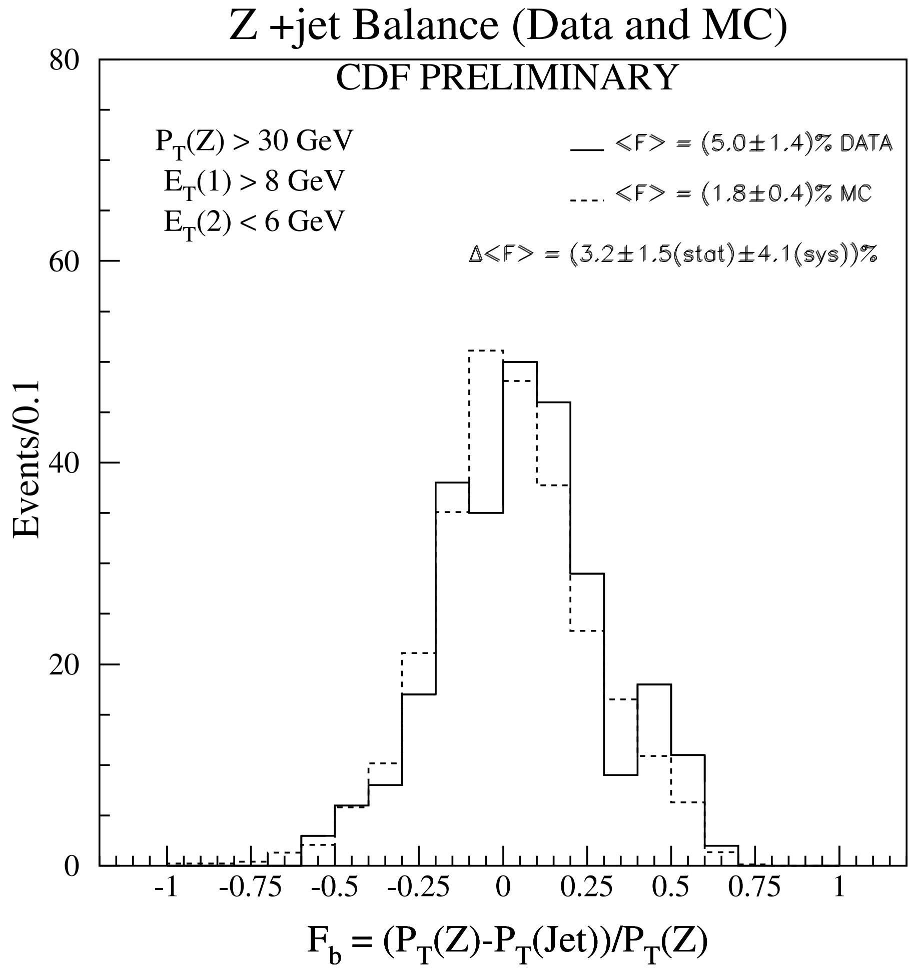

Figure 2.15: *Comparison of Z-jet balancing in the data and in the Monte Carlo simulation. The offset in the data may be used for an estimate of the systematic uncertainty in the jet energy measurement, while the difference in offset between the two samples yields the systematic uncertainty in jet energy-derived measurements relying on the comparison of distributions in data and Monte Carlo, such as that of the top quark mass[30].*

In order to estimate the uncertainty on the jet energy measurement due to the modeling of the absolute energy scale, one can use events containing a Z boson

[17] Primary vertices are reconstructed from VTX segments, and are classified according to the number of hits belonging to the tracks used in their reconstruction (see also Sec. 2.4.2 for a description of the VTX system and Sec. 2.6.4 for a description of the primary vertex finding algorithm).

decayed to a e^+e^- or $\mu^+\mu^-$pair, recoiling against a single jet: the P_T of the dilepton pair, well measured from tracking information, must then equate to the jet P_T. The relative difference between these two observables allow us to check if the jet energy scale is offset from its true value. As Fig. 2.15 shows, both data and Monte Carlo betray a small offset from zero. The 5% offset in the data may be taken as a systematic uncertainty due to the energy scale in the calorimeter.

2.6.2 Identification of Soft Leptons

Leptons are copiously produced in the decay of charmed and bottom hadrons: for a bottom meson, for instance, one expects an electron or a muon to be emitted 20% of the time in the decay. Downstream, if the b quark has produced a c quark[18], one expects a similar fraction of decays to electrons or muons, albeit with a softer momentum spectrum. Since, on the other hand, particles originating from the fragmentation of a light quark or a gluon seldom produce leptons (the main sources are kaon and pion decays to muons, but their long lifetime makes these decays not frequent inside the CDF volume), the identification of a clean lepton inside a jet cone provides a very useful tool to select events containing heavy quark decays.

Electrons are identified at CDF as energy deposits in the lead calorimeter consistent in transverse shape with electromagnetic showers. Their typical identification criteria vary sensibly as a function of their E_T and rapidity: in the central region they can be told apart from photons by the presence of a matching charged track with a value of E/P (ratio of energy in the calorimeter and measured track momentum) consistent with that of a single electron; outside it only high E_T electrons coming from W or Z decay can still be identified with reasonable purity, but photons—and occasionally jets that fluctuated to a high π^0 component—can no longer be ruled out as competitive processes, and the electron purity is process-dependent[19]. For the lower E_T electrons coming from the semileptonic decay of a heavy flavor, a matching track reconstructed in the CTC is essential to reduce the background, and therefore only the central rapidity region is considered. The other crucial parameter in their identification is the isolation of the electron signal. The isolation of the energy deposit in the e.m. calorimeter may be quantified by the total P_T of tracks in a $R = 0.4$ cone around the electron direction or by the total E_T deposited in surrounding calorimeter towers. The first definition is appropriate to banish jets faking high E_T electrons when one searches for leptonic W and Z boson decays, since some track activity around the electron candidate is expected in jet backgrounds; but it is impossible to use it if one wishes to select heavy flavors, which are expected to have a number of nearby tracks. On the other hand, an E_T deposit in the e.m. calorimeter cannot tag with certainty the presence of an electron even in the presence of a matching track, because the background due to other particles having released most

[18]The $b \to c$ decay is largely dominant, since $|V_{cb}|^2 \sim 200|V_{ub}|^2$[36].

[19]For instance, if the event contains another electron candidate and their combined mass is close to 91 GeV/c^2, the background is still extremely small; W candidates with a Plug electron (without a matched track) suffer on the contrary from larger backgrounds, due to events where the $\not{E}_T$ is due to a fluctuation of jet energy response and the electron candidate is actually a jet.

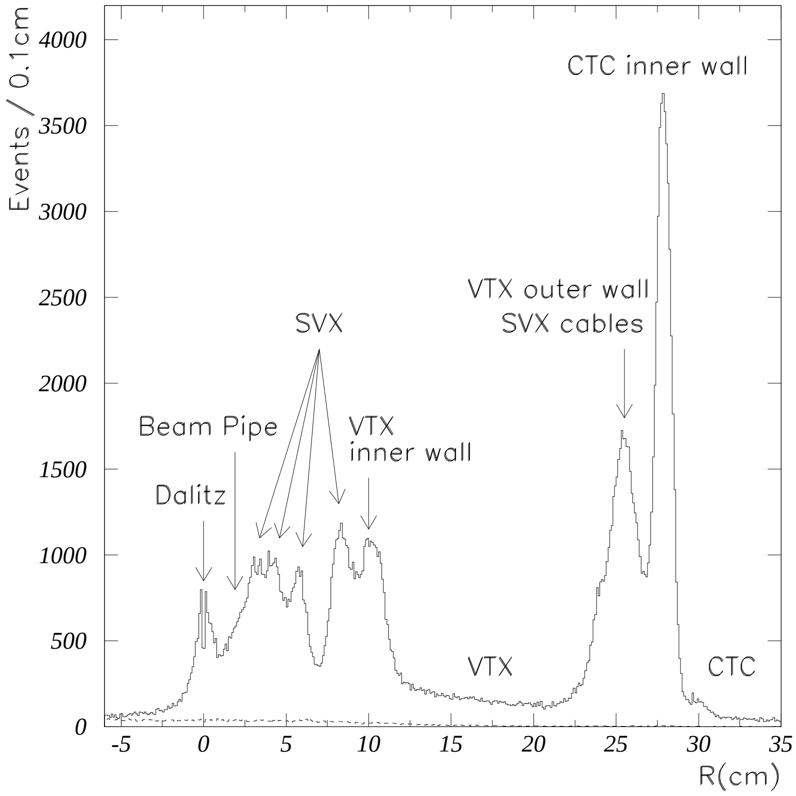

Figure 2.16: *This plot shows the r distribution of reconstructed conversion vertices. Spikes identify the distance from the beamline where a significant amount of material is present (beampipe wall, subdetector boundaries, etc.).*

of their energy in the first layers of the e.m. calorimeter is too high. If the deposit is isolated from other energy deposits, electrons are however still the most probable cause. A variable called $HAD\_EM_{3\times3}$ is used to select good electrons originated from semileptonic decays of heavy flavors: it is computed as the ratio of hadronic over e.m. energy in the square of nine towers having the electron seed tower as its center. Most analyses looking for a sample of semileptonic b-quark decays require this ratio to be less than 4%. This cut, however, affects the acceptance of b quarks in an E_T-dependent way, as can be seen in Fig. B.1 in App. B; this is the main reason of the low acceptance of the inclusive electron trigger for $Z \to b\bar{b}$ decays. The set of cuts usually applied to select soft electron candidates are listed in Table 2.5. Apart from isolation, there are cuts on the χ^2 test of resemblance between the Central Strip Chamber signal shape and the one expected from electrons, cuts on the spatial mismatch of electron track and signal in the CES, and a cut on L_{shr}, a variable quantifying the consistency of the lateral profile of the energy in adjacent calorimeter cells with what is expected by considering the CES signal and the rz angle of incidence of the electron track in the calorimeter.

At the end of this selection chain, a further cleanup of the electron candidates is necessary to reject conversion candidates. As we already pointed out in Sec. 2.4.2, a $p\bar{p}$ collision at the Tevatron produces on average 30 photons from the decay of neutral hadrons: these photons have a nonzero chance of converting to an electron-positron pair when they traverse the beampipe, the SVX layers, and the other material they find on their way to the e.m. calorimeter. A routine associates electron track candidates to any other reconstructed charged track of opposite charge, and the $r\phi$ separation with each is computed at the point where their ϕ opening angle and difference in $\cot\theta$ are at a minimum: if a zero-$r\phi$ match is found, the electron candidate is regarded as a conversion product and is discarded.

Variable	cut
E_T	$> 7.5\ GeV$
P_T	$> 6\ GeV/c$
$HAD\_EM_{3\times3}$	< 0.04
L_{shr}	< 0.2
Δx (track,CES)	$< 1.5\ cm$
Δz (track,CES)	$< 3\ cm$
Strip profile χ^2	< 10
Wire profile χ^2	< 10

Table 2.5: *Standard cuts used to select central electron candidates. See the text for details.*

Variable	Cut
P_T	$> 6\ GeV/c$
Δx (CMU)	$< 9\ cm$
Δz (CMU)	$< 12\ cm$
Δx (CMP)	$< 9\ cm$
d_μ	$< 3\ mm$
$E_{had} - \Sigma_{0.2}P$	$< 6\ GeV$

Table 2.6: *Standard cuts applied to CMUP muon candidates for the selection of good muons inside b jets. The P_T cut is a standard cleanup cut in the data collected by the muon triggers, whose thresholds are generally higher (see Sec. 3.2).*

The identification of soft muons inside jet cones is less problematic, since the central muon system is shielded by more than five pion interaction lengths. Occasionally, some secondary charged pions may still leak through the hadron calorimeter and give a signal in the CMU or CMX chambers, but in order to reach the CMP they need to pass through three additional interaction lengths. A coincidence between CMU and CMP signals suffers therefore from very little background, most of which is caused by decays in flight of kaons or pions to muons inside the calorimeter. To reduce these, a cut on the energy measured in the calorimeter tower traversed by the track is normally applied to muon candidates of high P_T. The standard cleanup cuts involve matching requirements to the stub in the muon chambers to a corresponding track in the CTC; further, the impact parameter of the muon track must be loosely consistent with the primary vertex or with a secondary vertex from heavy quark decay: that helps reducing cosmic ray background and decays in flight. Finally, a variable similar to the isolation described above for electrons, $E_{had} - \Sigma_{0.2}P$ (where E_{had} is the energy deposited in the hadronic calorimeter and $\Sigma_{0.2}P$ is the sum of momenta of tracks inside a cone of radius 0.2 in $\eta\phi$ space, centered around the muon track) is used to allow a maximum value of energy deposit in the calorimeter tower traversed by the muon track: the muon is a minimum ionizing particle, and is not expected to release a high fraction of its momentum in the hadron calorimeter; the amount of energy in the tower may be due to other hadrons traveling near the muon, and therefore the total momentum of any near track is subtracted to the tower energy before imposing a cut at 6 GeV. A list of the standard cuts used to select good CMUP muons is given in Table 2.6.

2.6.3 Missing Transverse Energy

The energy measured in the calorimeters is used to compute a quantity of great interest at hadron colliders: the missing transverse energy, $\not{E}_T$. It is computed as the negative of the vector sum of all calorimeter tower energies in the plane transverse to the beamline:

$$\vec{\not{E}_T} = -\Sigma_{i=1}^{N_{tower}} \vec{E}_T^{tower}. \tag{2.7}$$

A value of $\not{E}_T$ significantly different from zero may signal the presence of particles that escaped the calorimeter system undetected. This is the case of energetic neutrinos, emitted for instance in the leptonic decay of a W boson; however, before a significant[20] value of $\not{E}_T$ can be attributed to the presence of one such particle, a careful correction must be applied. In fact, the most common source of $\not{E}_T$ is a fluctuation in the energy measurement of hadronic jets; the other common source is the presence of energetic muons, that can be identified in the outer muon chambers: they normally deposit a constant amount of energy in the calorimeters, regardless of their momentum[21]. Therefore, once muon candidates are identified (see Sec. 2.6.2), and jet momenta are corrected with the routine described in Sec. 2.6.1, the $\not{E}_T$ can itself be corrected to yield a better measurement of P_T of a possible neutrino. Even so, its resolution on the neutrino P_T is typically[22] only of the order of 100%: an extremely careful modeling of its measurement is therefore mandatory in analyses that use its value for precision measurements, as is the case of the determination of the t-quark or the W boson mass[30, 33].

[20] The common measure of the missing E_T significance is given by its absolute value divided by the square root of the total transverse energy measured in the calorimeters: $S = \not{E}_T/\sqrt{\Sigma E_T}$.

[21] See Sec. 2.6.2. A plot that demonstrates the independence of the energy deposit for muons on their P_T is given in Fig. 5.1 (Ch. 5).

[22] The resolution actually depends strongly on the nature of the system recoiling against the $\not{E}_T$: when jets are not present, as is the case of leptonic W and Z decays, the $\not{E}_T$ is known with a considerably higher resolution than what is quoted above.

2.6.4 Secondary Vertex Tagging

As we have described in Sec. 2.6.2, lepton triggers can provide a good sample of heavy flavor decays. But by far the most powerful tool to identify jets originating from b quarks is the reconstruction of a decay vertex using the charged tracks contained in the jet cone. This is made possible at CDF by the very precise Silicon Vertex detector, which allows a measurement of the impact parameter of tracks in the plane transverse to the beam[23] with an error as small as 13 μm for tracks of high momentum.

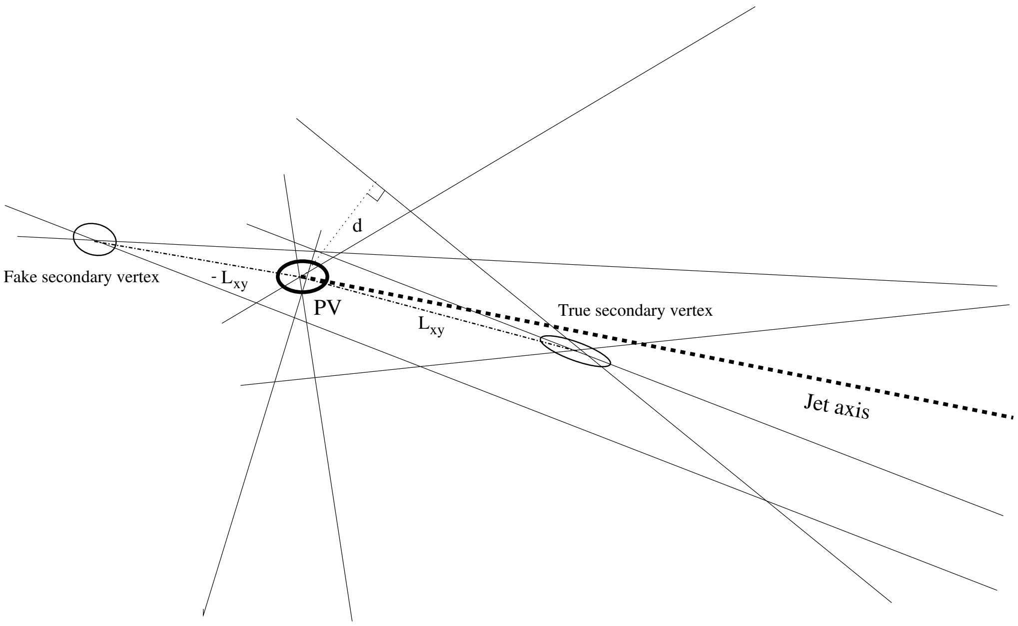

Figure 2.17: *Graphical explanation of SECVTX tagging. Tracks with significant impact parameter d are used to fit a secondary vertex. When the angle between the jet axis and the resulting L_{xy} is acute, L_{xy} gets a positive sign and the vertex is called a "positive SECVTX tag".*

Before starting a description of SECVTX, the algorithm used to find b-decay vertices[24], we need to describe the procedure used at CDF to identify the location of the primary interaction vertex in the transverse plane. As we have mentioned in Sec. 2.4.2 in the description of the VTX chambers, the z location of the interaction vertex is provided by a fit of track segments identified in this device. The z coordinate of the vertex is used to compute the actual pseudorapidity of each physics object reconstructed in the event, which in turn is used to extract the transverse energy of jets from the raw measurement of total energy in the calorimeters: however,

[23]The SVX measures track hits in the $r\phi$ plane, as described in Sec. 2.4.1.

[24]Along with SECVTX, other algorithms have been devised to tag b-quark jets. They are all based on the selection of jets containing tracks with high impact parameter; one of them, JPBTAG, has been thoroughly investigated as a possible alternative to tag b-quark jets in $t\bar{t}$ decays[34]. Other algorithms are tuned for the reconstruction of low P_T b-quark decays, and are mostly used for b-physics measurements.

the resolution of this quantity does not have a big impact on the event reconstruction. On the contrary, the xy location of the vertex—and its resolution as well—are critical parameters in the identification of secondary vertices. The proton-antiproton luminous region has a Gaussian spread of about 35 μm in the transverse plane, and its location is known with good precision, but a more precise determination can be obtained with the track parameters. Instead of relying on the beam spot location, use is made of a Fortran algorithm named VXPRIM[35]. VXPRIM uses charged tracks in the CTC matched to hits in the SVX detector (SVX tracks) to fit the primary vertex by an iterative procedure. The algorithm starts by selecting all SVX tracks pointing to the highest ΣP_T primary vertex among those reconstructed in the z coordinate by VTX information alone. A P_T-weighted fit is performed with the tracks in the xy plane; tracks with large impact parameter with respect to the fit vertex are discarded, and the fit is repeated until stability is reached. The resulting resolution in the x and y coordinates of the primary vertex ranges from 6 to 26 μm, depending on the topology of the event and on the number of tracks used in the fit. This determination of the primary vertex represents a significant improvement over the beam spot information alone, and it is more reliable on an event-by-event basis.

The algorithm used to identify secondary vertices is called SECVTX. It is the result of detailed optimization studies performed at CDF to obtain a b-quark tagging capable of high background rejection and high acceptance for real b-quark jets, such as those coming from $t\bar{t}$ decay[47]. To reconstruct a vertex in a jet, the algorithm starts with a list of all charged tracks reconstructed inside the jet cone (of radius $R = 0.4$ or $R = 0.7$, depending on the way jets are defined in a given analysis). A first search is made for all tracks having:

- $P_T > 0.5\ GeV/c$,
- impact parameter significance[25] $S_d > 2.5$, and
- hits in at least two layers of the SVX.

Tracks passing these criteria are ordered by P_T and S_d, and the first two tracks locate a *seed vertex* where their trajectories cross in the transverse plane. The impact parameter of the other tracks with respect to the seed vertex is then computed, and tracks with $S_d^{seed} < 3$ are associated to the seed vertex. If no such tracks are found, a different pair of tracks is used to define a new seed vertex, and the procedure is repeated. At the end of this routine, if a vertex with three or more tracks has not been found, a second phase begins. A new list of tracks with more restrictive requirements is compiled:

- $P_T > 1.5\ GeV/c$,
- $S_d > 4.0$,

[25]The impact parameter significance of a charged track is defined as $S_d = |d|/\sigma_d$, where d is the track's impact parameter and σ_d is the uncertainty in its determination, obtained from the uncertainty in the track helix parameters and the resolution on the location of the primary vertex in the plane transverse to the beam.

- track has hits in at least three layers of SVX.

A two-track vertex is then fit with the tracks passing these tight requirements.

At the end of the procedure, if a vertex has been found it is classified according to the reconstructed decay length L_{xy} in the transverse plane, computed as the distance between the primary and secondary vertex. The L_{xy} is required to be significantly

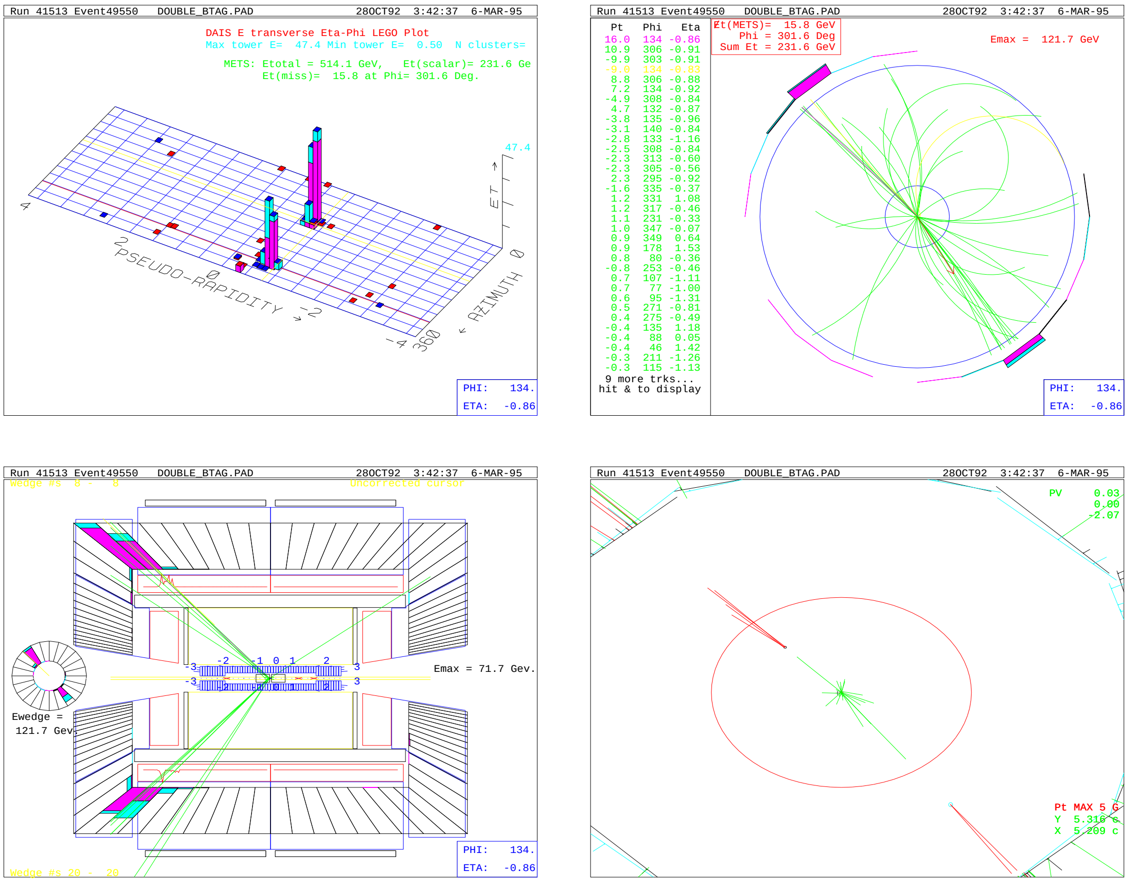

Figure 2.18: *Display of an event containing two high E_T jets, both tagged by SECVTX. The upper left plot shows the calorimeter depositions of the two jets, the upper right plot shows the tracks reconstructed in the CTC, the lower left plot shows a side view of the central detector, and the lower right plot shows the two reconstructed vertices from a blow-up of the interaction region in the transverse plane.*

greater than zero, $|L_{xy}|/\sigma_L > 3$. For the classification of the secondary vertex it is useful to give a sign to the value of L_{xy}: the decay length is given the sign of the scalar product between the vector pointing from primary to secondary vertex and the vector obtained as the sum of the momenta of all tracks fitted to the secondary vertex (see Fig. 2.17). If the vertex tags the location of a real heavy quark decay, the sign of the decay length is positive[26], while negative decay lengths are mostly

[26] Actually, 6% of the time a vertex with negative decay length is due to a heavy flavor decay:

due to fake vertices. Finally, two-track vertices consistent with coming from the decay of a K^0_S meson or a Λ baryon are removed using the decay distance[27] and the vertex mass and charge. A jet with positive and significant decay length is called a b-tagged jet, since its most likely origin is a b quark. Figure 2.18 shows four event displays for an event containing two SECVTX-tagged jets.

A remark on the typical terminology used in connection with SECVTX tagging is appropriate here, since we shall use it widely in what follows. A *taggable track* is defined as a track passing the criteria of one of the two selections listed above; a *taggable jet* is subsequently defined as a jet containing two or more taggable tracks inside the cone used in the clustering algorithm.

To give an example of the performance of the SECVTX algorithm on b and c quark jets we can use three Monte Carlo datasets of electroweak signals with heavy quarks in the final state: $Z \to b\bar{b}$, $Z \to c\bar{c}$, and $W^+ \to c\bar{s}$ decays[28]. The efficiency of requiring one or more jets with a b-tag in these events is shown in Table 2.7. One notes that, not surprisingly, the efficiency is very high for events having two b quarks in the final state, while it is almost four times lower for c quarks[29]. The algorithm can then be thought to select mainly b quarks.

Sample	$Z \to b\bar{b}$	$Z \to c\bar{c}$	$W^+ \to c\bar{s}$
$\epsilon_{\geq\ 1\ tag}$	60.7%	17.7%	10.4%

Table 2.7: *Efficiency of the requirement $N_{tag} \geq 1$ for electroweak processes producing heavy flavors in the final state (for the characteristics of the Monte Carlo datasets used for this estimate see Sec. 3.3.1). Tracking inefficiency effects and the SECVTX scale factor (see text) have not been applied here.*

The numbers quoted in Table 2.7 are to be considered only a first-order estimate of the tagging probability in heavy flavor jets. In fact, the detector simulation performed on the Monte Carlo data has been shown to be unable to reproduce with sufficient accuracy the rate of tagging b-quark jets found in the data. This is due to two competing sources:

1. Inefficiency effects in the Central Tracking Chamber;

2. Overefficiency effects in the selection of SVX tracks with significant impact parameter.

---

this is due to the bad reconstruction of the decay vertex, particularly when tracks from the b vertex and the c vertex are fitted together, or when the hadron momentum is small.

[27] $\tau(K^0{}_S) \sim 9 \times 10^{-11}s$, $\tau(\Lambda) \sim 2.6 \times 10^{-10}s$[36].

[28] These datasets are described in detail in Ch. 3.

[29] The most informative number would actually be the tagging efficiency *per jet*: it can be easily derived from the numbers in Table 2.7: since $\epsilon_{\geq 1} = \epsilon_1 + \epsilon_2 = 2P_j \times (1 - P_j) + P_j^2$, one finds $P_j = 1 - \sqrt{1 - \epsilon_{\geq 1}}$, so that for a b quark from Z decay one has $P_j = 37.3\%$ (compare to what is found in Ch. 3), and for a c quark from Z decay $P_j = 9.3\%$. For $W^+ \to c\bar{s}$ events the probability of tagging the jet coming from the s quark is very small, and the number quoted is in fact compatible with the c quark probability extracted from $Z \to c\bar{c}$ decays.

Occupancy effects arise when two or more charged tracks travel next to each other (with distance of the order of a few millimeters) inside a jet cone: it becomes then difficult for the algorithm performing the track reconstruction to accurately identify all of the tracks from the hits in the CTC, so that occasionally a track gets lost, *id est* is not reconstructed by the tracking algorithm. In the Monte Carlo simulation of the Central Tracking Chamber these effects are not simulated with enough accuracy: therefore, a b-quark jet from the simulation may contain more SVX tracks[30] than an equivalent jet in the real data. Since SVX tracks are the starting point for the search of secondary vertices in the SECVTX algorithm, the Monte Carlo ends up being overefficient in tagging b-quark jets.

The problem of high CTC occupancy has been studied in detail at CDF. The technique used to measure this effect has been based on the random embedding of Monte Carlo tracks inside the jet cones in experimental events, by a modification of the data banks containing the raw CTC hit information. Running the reconstruction algorithm on the new banks allows us to pinpoint the situations where the reconstruction has an efficiency smaller than 100%: these effects have been included in a modified version of the SECVTX algorithm, suitable for Monte Carlo events. There, a random removal of charged tracks is performed according to a parametrization of the inefficiency, simulating the effect and allowing a safe comparison of rates of tagging in data and simulations. The typical correction operated by the parametrization for b-quark jets in the characteristic energy range of $t\bar{t}$ decays has been measured to be $k = 0.87 \pm 0.07$[37, 38] or $k = 0.92 \pm 0.03$[39]: these measurements are consistent with the output of the modified SECVTX algorithm.

The second problem cited above has been the subject of a careful investigation at CDF recently. Since the agreement between tagging efficiency in data and simulations is a factor of crucial importance in the accurate determination of the cross section of processes that need SECVTX tagging to be isolated, such as $t\bar{t}$ production[40], there have been a number of detailed measurements of the tagging efficiency for b-quark jets in the data, to verify the reliability of the Monte Carlo simulations. A perfect Monte Carlo simulation—where the tracks had distributions of impact parameter, impact parameter significance, track momentum, and charged track multiplicity completely equivalent to what the data show—would no doubt give the exact value of tag probability; but a perfect Monte Carlo hardly exists at all.

To obtain a measurement of the SECVTX tagging probability for b quarks from the data themselves, data samples rich in b-quark jets have been obtained by collecting dijet events, where at least one of them was tagged by SECVTX and the other contained a soft lepton. Careful sample composition studies, that take into account all the different possible sources of SECVTX tags and soft leptons—the three leading order processes of direct production, gluon splitting and flavor excitation (see Fig. 4.1, page 56), both for $b\bar{b}$ and $c\bar{c}$ production, plus fake secondary vertices and fake leptons—allow to estimate the b purity of the lepton jet at around 40%, which in turn may yield the tagging probability of SECVTX by a comparison of the number of events with two SECVTX tags to the number of events where

[30]As was defined above, a SVX track is a CTC charged track matched to hits in the silicon layers of the SVX.

the lepton jet is only defined as taggable. Since many different datasets can be used for this purpose, the rate of tagging is now known with high accuracy. The results[41, 42] speak in favor of a ratio of tagging efficiency between data and Monte Carlo (named "*SECVTX scale factor*") equal to $f = 1.25 \pm 0.13$ for run 1b data, and $f = 1.21 \pm 0.13$ for run 1a data. Nevertheless, these numbers must be regarded as a preliminary estimate, since they are still being investigated for the purpose of a new determination of the $t\bar{t}$ cross section.

Two issues, in particular, have been recently under scrutiny. The claim that these factors are totally independent on jet E_T has now—after lengthy debates—won total agreement in the collaboration[31]: the scale factor f has been for a while thought to be decreasing from 1.25 at low E_T down to 1.0 for high E_T jets, due to the confusion caused by assuming that the heavy flavor fraction of the lepton jet was constant—something that has been disproven recently[32]. The second issue concerns the real cause of the discrepancy of f from 1.0: the hypothesis that the simulation of the SVX detector yielded a too high resolution on the track impact parameter, with the consequence that fewer tracks with large d/σ_d could be found in the Monte Carlo, and fewer secondary vertices were then reconstructed, has been found not sufficient to explain the value of f. The discrepancy is therefore still not completely understood; on the other hand, since f has been *measured* with the data themselves, one could jolly well forget about understanding the reason for its value.

[31]The stability of this number for different jet transverse energies is of crucial importance, given that most of the statistics used for the comparisons of tag probability lies at low values of jet E_T (10 to 20 GeV), while the typical particle signals one wishes to quantify—such as the $t\bar{t}$ or the $Z \to b\bar{b}$ signals—are characterized by b jets of high energy (typically over 30 GeV).

[32]In fact, the scale factor keeps constant, but the heavy flavor fraction decreases, making the tagging rate decrease as well.

Chapter 3

The Datasets

We describe in this chapter the datasets we use for our search of $Z \to b\bar{b}$ decays. In Sec. 3.1 we explain our choice of the starting dataset; the latter is described in Sec. 3.2. We use for our search some Monte Carlo simulations of the $Z \to b\bar{b}$ signal, to estimate the amount of that process at the various levels of selection and to study its characteristics, to try and find the best variables to discriminate it from background processes: these simulations are described in Sec. 3.3.1. There, we also describe other useful event samples that will be used in Sec. 4.2 for a better understanding of some of the characteristics of the $Z \to b\bar{b}$ signal.

3.1 The Choice of the Trigger

The first decision one has to take, getting started in a search for $Z \to b\bar{b}$ decays at CDF, is what dataset to look in. As we have seen in Sec. 2.5, the three-level trigger system used at CDF collects $p\bar{p}$ collisions in many different output streams, according to the characteristics of the events. The Level 3 trigger is managed by fast processors that divide the events to be collected into several different outputs; the user interested in a particular physics object may therefore look in the relevant dataset, where all of the events judged likely to contain that object are gathered.

For a search of Z decays to b-quark pairs the most obvious choice would appear to be a dataset collected by triggers requiring the presence of two jets. During run 1 CDF collected jet data principally with its four single-jet triggers, that accepted events having at least one calorimeter cluster above a predefined transverse energy threshold. These thresholds were set at 20, 50, 70, and 100 GeV during both run 1a and run 1b. For the hadronic decay of a Z boson one would then choose the lowest energy jet trigger, *i.e.* JET_20, since the higher ones are evidently inefficient for a process like Z decay, which gives rise to jets of transverse energy often smaller than 30 GeV[1]. Unfortunately, the cross section for background processes giving rise to a

[1] As a matter of fact, even the JET_20 trigger is not fully efficient for the Z signal, since the measurement of cluster energy at Level 2 is managed by a hardware cluster finder and thus is not precise, and the location of the primary vertex in the z axis is not yet known and assumed to be $z = 0$ (the z coordinate is used when converting total into transverse energy in the calorimeters): these approximations cause a slow turn-on in the efficiency curve as a function of the offline-

20 GeV jet is unmanageably high at the Tevatron: the production rate exceeds by almost three orders of magnitude the effective storage capability, and the physics one can do with those events is at any rate not very exciting. The CDF trigger is in fact designed to collect with the highest efficiency those much more interesting data containing a well identified lepton, or a photon, or *very high* E_T jets: objects that may be the signal of new physics processes, or that may grant precise measurements of quantities whose value is not yet well known. The low-energy jet triggers are then prescaled, so that a reasonably sized dataset is collected without the risk of saturating the bandwidth. In particular, the 20 GeV jet trigger has been subjected during run 1 to a typical prescaling factor of 200 to 1000. This makes that dataset completely useless for us: once one accounts for the prescale, the total effective integrated luminosity amounts to a bleak 0.3 pb^{-1}.

At the end of run 1c a field wire in the CTC broke, and a whole sector of the tracking volume needed to be turned off. Replacing the wire would have taken more time than what was still left for collider running, so the trigger system was modified, in order to collect data that did not need tracking information. In particular, about 2 pb^{-1} were collected by a very low transverse energy dijet trigger, DIJET_12, fired by events having two jets with E_T over 12 GeV. Many efforts have gone into trying to find a signal of W bosons in that sample[45, 46], with no success. The small S/N and luminosity and the insufficient knowledge of the turn-on shape of the trigger in the mass distribution, in fact, conspire to make the background shape uncertainty larger than the expected signal at masses around 80 GeV/c^2. A $Z \rightarrow b\bar{b}$ search in that dataset seems *a fortiori* not feasible, since the $Z \rightarrow b\bar{b}$ cross section is about 15 times smaller than the $W \rightarrow jj$ one at 1.8 TeV.

A jet trigger is therefore not useful for our purpose: we must investigate other options. To search for $Z \rightarrow b\bar{b}$ events one must rely on triggers that collected b-quark jets in large amounts, were not prescaled, and affected the mass distribution in a controllable way. The largest datasets of b quarks, widely used for many b-physics searches and measurements, are provided at CDF by the low energy electron and muon triggers. The two principal triggers had a P_T threshold of about 8 GeV/c for both electron and muon candidates.

As was described in some detail in Sec. 2.2, electrons are identified at CDF with good accuracy when they have high E_T and are isolated from other energy deposits in the calorimeters: they are then used with success for precise measurements of the W mass[33] and for the identification of top quark decays[40]. On the other hand, when they come from semileptonic heavy quark decays their energy may be low, and they can be faked by other particles, such as neutral pions with a charged track traveling nearby (providing for a matching charged track); some discriminating power is still possible with the signals in the strip chambers, but—in order to attain a reasonable S/N ratio—the electron candidates need to be required to be isolated from other activity in the calorimeter. The $HAD\_EM_{3\times3}$ isolation (see Sec. 2.6.2) is still an efficient tool when the jet containing them has low energy; on the other

reconstructed jet E_T. Moreover, the reader must keep in mind that these energies are not yet corrected for the non-uniform response of the calorimeters and for the other effects described in Sec. 2.6.1. The overall efficiency of the JET_20 trigger for the $Z \rightarrow b\bar{b}$ decay is about 60%.

hand, for a 40 GeV jet—such as one of those produced in a Z boson decay—the efficiency of the typical cut drops dramatically (see Fig.B.1 on page 180).

Muons suffer their own identification problems in the CDF detector, but their purity does not vary as much, and they end up being spotted much more efficiently inside the jets in the energy range characteristic of hadronic Z decay. The interested reader will find in App. B a detailed analysis of the comparison between the electron and muon datasets for the purpose of a Z search. At the end of the day, the combined result of acceptance and identification efficiency make the muon dataset twice as efficient for $Z \to b\bar{b}$ decays as the electron dataset. Therefore, from now on we will concentrate our analysis on the former. In App. A a similar search performed in the electron dataset is summarized.

3.2 The Inclusive Muon Dataset

For the reasons described in the previous section, the starting point of our search is a data sample collected with triggers requiring the presence of a muon *stub*[2] in the muon system, matched to a track reconstructed in the Central Tracking Chamber. The run 1a dataset consists in 2,724,593 events collected in the COMBINED_MUO1 stream, which contains all the events that satisfied at least one of the Level 3 muon triggers; the run 1b dataset consists in 2,690,162 events collected in the COMBINED_MULB stream, again the logical OR of many different Level 3 triggers[3]. The principal Level 2 and Level 3 triggers contributing to the 1a and 1b datasets are described in Table 3.1 and Table 3.2. All in all, the total run 1 data amount therefore to about 5,4 million events. These data exist in two different formats: the DST and the PAD. The DST format includes all the possible information obtained by the detector readout for each event, while the PAD format contains only the information that the typical user needs to access. Both DST and PAD data are stored in high density 8 mm tapes.

3.2.1 Estimation of the Integrated Luminosity

For a preliminary computation of the number of $Z \to b\bar{b}$ events collected in our sample we need to estimate the integrated luminosity the two datasets correspond to. We use for that purpose a utility module, LUM_CONTROL, that accesses a database containing all the information concerning the luminosity integrated during each data taking cycle, the activity of subdetector components, the relevant trigger tables, and the dynamic prescaling of specific triggers. This program accepts as input the list of files in the tapes containing the data whose integrated luminosity one wants to estimate; one must thereafter specify which Level 2 trigger he or she is interested in, because many have a static or a dynamic prescale that needs to be

[2]A stub is a track segment reconstructed from the collection of hits in the four logical layers of the muon chambers.

[3]The 1a dataset corresponds to an integrated luminosity about four times lower than the 1b dataset; the equivalence of the two datasets in number of events is due to the looser requirements applied to muon tracks during run 1a.

Trigger name	Description
CMUP_CFT_6*	A prescaled trigger requiring a 6 GeV/c CFT track and hits in the CMU and CMP chambers as a prerequisite.
CMUP_CFT_9_2*	The main CMUP muon trigger for run 1a, with a 9.2 GeV/c threshold on CFT P_T and a 5-degrees matching requirement with the muon chamber signals.
CMU_CMP_CFT_9_2*	This trigger, very similar to the one above, did not require the track-list board matching of CTC track and muon stubs. It was substituted by the former after the first part of data taking.

Table 3.1: *Description of the principal Level 2 triggers contributing to the run 1a COMBINED_MUO1 stream; the asterisks remind that these triggers existed in different versions. CMX triggers and other muon triggers are not included since we require our muons to be CMUP.*

taken into account for the estimation; besides, some triggers have been active for only a fraction of the data taking. Unfortunately, our run 1a and run 1b inclusive muon datasets are a composition of several different trigger streams, each of whose has a different acceptance for Z events. To simplify this complicated issue, we have to rely on an approximation.

We note that for each of the two data taking periods there is a Level 2 trigger whose contribution is largely dominant for the inclusive muon dataset: CMU_CMP_CFT_9_2 for run 1a and CMUP_CFT_12_5DEG for run 1b. The luminosities integrated with these, according to LUM_CONTROL, amount respectively to 13.97 pb^{-1} and 85.81 pb^{-1}. The total integrated luminosities corresponding to our file lists amount instead to 20.03 pb^{-1} (run 1a) and 86.33 pb^{-1} (run 1b)[4]. We may take the average of these numbers for each data taking to get our estimate[5], and give as a preliminary systematic error on the integrated luminosity half the difference. That way we get $\int_{1a} \mathcal{L}dt = 17.0 \pm 3.0 \pm 0.7\ pb^{-1}$ and $\int_{1b} \mathcal{L}dt = 86.1 \pm 0.2 \pm 6.5\ pb^{-1}$, which can be combined to a total of $\int \mathcal{L}dt = 103 \pm 7\ pb^{-1}$.

The final ingredient we need, in order to be able to use the Monte Carlo efficiencies for an estimate of the amount of signal in our dataset, is the cross section for Z production at the Tevatron. The cross section has been measured at CDF[51, 52]

[4]The luminosity estimates quoted are those relevant for data collected with the CMU, CMP and SVX subdetectors operating properly [48, 49].

[5]In principle, one should instead obtain the integrated luminosity for all the contributing triggers, evaluate the correlations between each of them, and use a Monte Carlo simulation to understand their individual contributions to the total dataset. However, due to the large uncertainties inherent to the estimations of the integrated luminosity[20], this complicated task would be completely useless. Averaging the total luminosity with the value integrated by the most important trigger is safe in our particular case, and does not cause a large increase in the total error.

Trigger name	Description
CMUP_CFT_12_5DEG*	The main trigger designed for the collection of central muon events in run 1b. It requred a Level 1 high P_T muon trigger, with hits in the muon chambers associated to a CFT track of $P_T > 12\ GeV/c$.
CMUP_CFT_7_5*	This trigger had very similar requirements to the former, but a lower CFT P_T threshold of 7.5 GeV/c. It was dynamically prescaled, with an average prescale factor of ~ 1.7.
CMUP_JET_15*	This trigger had the same requirements on the muon track as the 12 GeV/c one, but required also the presence of a 15 GeV calorimeter cluster.

Table 3.2: *Description of the principal Level 2 triggers contributing to the run 1b COMBINED_MUOB stream; the asterisks remind that these triggers existed in different versions. CMX triggers and other muon triggers are not included since we require our muons to be CMUP.*

and D0[53] both in the dielectron and in the dimuon channel. The most precise determination comes from $Z \to e^+e^-$ decays detected by the CDF detector during run 1[54]: the measured value of Drell-Yan production in the mass window $66 \div 116\ GeV/c^2$ is $\sigma_Z \times BR(e^+e^-) = 245.3 \pm 3.9 \pm 3.5 \pm 17.6\ pb$, where the first uncertainty is statistical, the second is the systematic error due to the selection efficiencies, and the last is the systematic uncertainty due to the integrated luminosity. By scaling this number by the ratio of branching fractions BR($b\bar{b}$)/BR(e^+e^-)[6] one finds that the cross section times branching ratio for the $Z \to b\bar{b}$ process amounts to $1.110 \pm 0.031(stat. \oplus syst.) \pm 0.090(lum.)$ nanobarns: therefore, in 103 pb^{-1} of $p\bar{p}$ collisions one expects to have produced about $114,000$ $Z \to b\bar{b}$ events.

3.2.2 The First Requirements

Even after the application of the identification criteria for muon candidates at Level 3, the dataset is still rich with fake muons: tighter requests have to be applied on the muon identification variables to obtain a reasonably pure sample[7]. Therefore,

[6]We can use for that purpose the fitted values found by the Particle Data Group in 1998[36]: $\Gamma_{Z\to b\bar{b}}/\Gamma_{Z\to hadrons} = 0.2169 \pm 0.0012$ and $\Gamma_{Z\to hadrons}/\Gamma_{Z\to e^+e^-} = 20.77 \pm 0.08$.

[7]In principle, given that we later select events by requiring the presence of two SECVTX tagged jets (see Sec. 4.1), we could avoid discarding fake muon events from our dataset, since what we want to select are $b\bar{b}$ events regardless of their decay. All in all this is not convenient, given two observations: first of all, we would not increase the double tagged dataset by a large amount (fake muon events are much less likely to contain tagged jets than the real muon ones, of course); moreover, we are actually bound to be confident of the nature of our candidate muons, since we correct jet momenta to take care of the energy taken away by the muon-neutrino pair, as explained in Ch. 5. Moreover, evaluating the efficiency to collect real Z events containing a fake muon in our

after a general cleanup[8], we filter the events by requiring the muon track to fulfill a set of criteria carefully tuned to discard tracks having high probability of belonging to fake muon candidates (these criteria are listed and explained in Table 2.6 on page 35). We also select at this stage the events that contain CMUP muons, *id est* muons that yielded signals in both the CMU and the CMP chambers. The choice to select only events with CMUP muons—and therefore discard events that could have good muons but were out of the common acceptance of the CMU and CMP chambers (muons traversing the CMX, or passing through uninstrumented regions of either the CMU or the CMP[9]—was dictated by many reasons. CMX muons do not add very much to the overall acceptance, due to the limited coverage of these chambers and to the combined effect of the rapidity distribution of muons from the $Z \to b\bar{b}$ decay and of the lower tagging efficiency of jets close to the boundary of the Central Tracking Chamber (see Fig. B.3, page 182); moreover, the CMX triggers were subjected to a static prescaling factor equal to 2.

In series with the muon filter, we require the events to have at least one jet tagged with SECVTX, with positive decay length ($L_{xy} > 0$, see Fig. 2.17, page 37). The options used in SECVTX are listed in Table 3.3; they are somewhat different from the "standard" ones, used for instance in all top search analyses by CDF[30, 40]. When using SECVTX one must first of all decide the cone radius of the clustering algorithm, since the search for secondary vertices is performed inside jet cones after these have been reconstructed. For the identification of a vector boson decaying to a pair of jets the most spontaneous choice would be that of a $R = 1.0$ cone: this would reduce the uncertainty due to the amount of radiation emitted by the final state partons at large angle with respect to their original direction; on the other hand, at high instantaneous luminosity one needs to worry about the amount of energy coming from satellite interactions from the same bunch crossing, collected accidentally inside the clustering cone[10]. We cannot choose this radius, though, since the SECVTX algorithm has been tuned to work with clusters of radius 0.4 or 0.7 only. Furthermore, one of the variables we use in our kinematic selection (we will discuss it in Sec. 4.2), the sum of clustered E_T outside of the two leading jets, becomes less discriminant when the jet cone is chosen to be 1.0, since the variable is sensitive to the amount of radiation in our events, and a cone too wide obliterates too big a portion of the phase space in the calorimeter. Therefore our choice for the reconstruction of the jets is a cone of radius 0.7, which dictates the radius for the search of secondary vertices inside them. As for the E_T threshold in SECVTX (normally fixed at 15 GeV of uncorrected energy for R=0.4 clusters), given that we wish to accept in our selection events with dijet invariant masses as low as possible, for a better understanding of the background shape in the mass distribution, we

Monte Carlo would be very difficult.

[8]Before any physics analysis can be undertaken, a general cleanup must be performed in order to discard events coming from bad accelerator runs (for whose there is no luminosity information) and to check that there are no multiple instances of the same event stored accidentally in the 8 mm tapes.

[9]See Fig. 2.11, page 23

[10]The correction due to this effect scales quadratically with jet radius.

set the jet E_T threshold at 10 GeV; besides, a 15 GeV threshold would be much too restrictive for jets that contain a $P_T \geq 8$ GeV/c muon and a $P_T \sim 5$ GeV/c neutrino.

Variable	Cut
Cluster E_T	> 10 GeV
Cluster radius	$= 0.7$
$S_d^1 = d/\sigma_d$ (pass 1)	> 2.5
$S_d^2 = d/\sigma_d$ (pass 2)	> 3.0
L_{xy}	> 0.0
$S_L = \lvert L_{xy}\rvert/\sigma(L_{xy})$	> 2.0

Table 3.3: *SECVTX parameters used for the selection of secondary vertex tags in the SECVTX sample.*

The requirement of at least one jet with a secondary vertex found by SECVTX (with positive decay length) with the options just discussed leaves us with 105,782 events. We will refer to this sample as the "SECVTX" dataset: it is the starting point of our search for Z bosons in the inclusive muon data, and at the same time an excellent sample for background studies, given the very small S/N ratio at this level.

In the following section we describe the other datasets we have used for our search. The most important one is the $Z \to b\bar{b}$ Monte Carlo simulation, used to compute the number of signal events at the various selection levels and to search for the most discriminating variables between the signal and the QCD background.

3.3 Other Datasets

3.3.1 The $Z \to b\bar{b}$ Monte Carlo Simulation

To study the kinematical properties of the $Z \to b\bar{b}$ decay we used PYTHIA V5_7[43] with the MRSD$-$ parametrization of the structure functions, and generated 1,673,000 signal events ($1.4M$ for run 1b and $273k$ for run 1a), choosing the Drell-Yan process and forcing the decay of Z bosons to b-quark pairs. The events were then passed through QQ V9_0[55], a package that simulates b decays according to the known branching ratios and composition; after that, a full detector simulation appropriate for run 1b (1a) was performed with the QFL′ (QFL) package[29].

We filter the simulated events by requiring the presence of a tight muon candidate (see Table 2.6 on page 35 for the cuts), to which we apply a trigger simulation. The 1a simulated events are tested against the Level 2 efficiency curve of the dominant CMUP trigger (CMUP_CFT_9_2), derived using a large dataset of $J/\psi \to \mu\mu$ decays collected by a low momentum dimuon trigger[56] during run 1a; on top of that, the Level 3 cut of 7.5 GeV/c is applied to the muon track P_T. The run 1b fraction is instead tested against both the CMUP_CFT_7_5 and the CMUP_CFT_12 curves, similarly derived from $J/\psi \rightarrow \mu\mu$ decays[57] collected during 1b; Table 3.4 describes their functional form and shows their parameters; the curves are drawn in Fig. 3.1.

Since the lower momentum trigger for run 1b is prescaled dynamically, we need to correctly model that effect in our Monte Carlo. That is done by looking in the real data at how many events passing the higher P_T trigger (not prescaled) fired the lower P_T one as well: we find this fraction to be 0.5678 ± 0.0012, independent on muon P_T as expected[11]; we therefore multiply the CMUP_CFT_7_5 curve by its average prescale factor before using it. We can finally take the logical OR of the two triggers for each of the 1b Monte Carlo events, and apply a $P_T > 8\ GeV/c$ cut to simulate the P_T requirement of the Level 3 triggers during run 1b. We believe the above procedure correctly accounts for the trigger bias to the muon momentum spectrum of $Z \rightarrow b\bar{b}$ events.

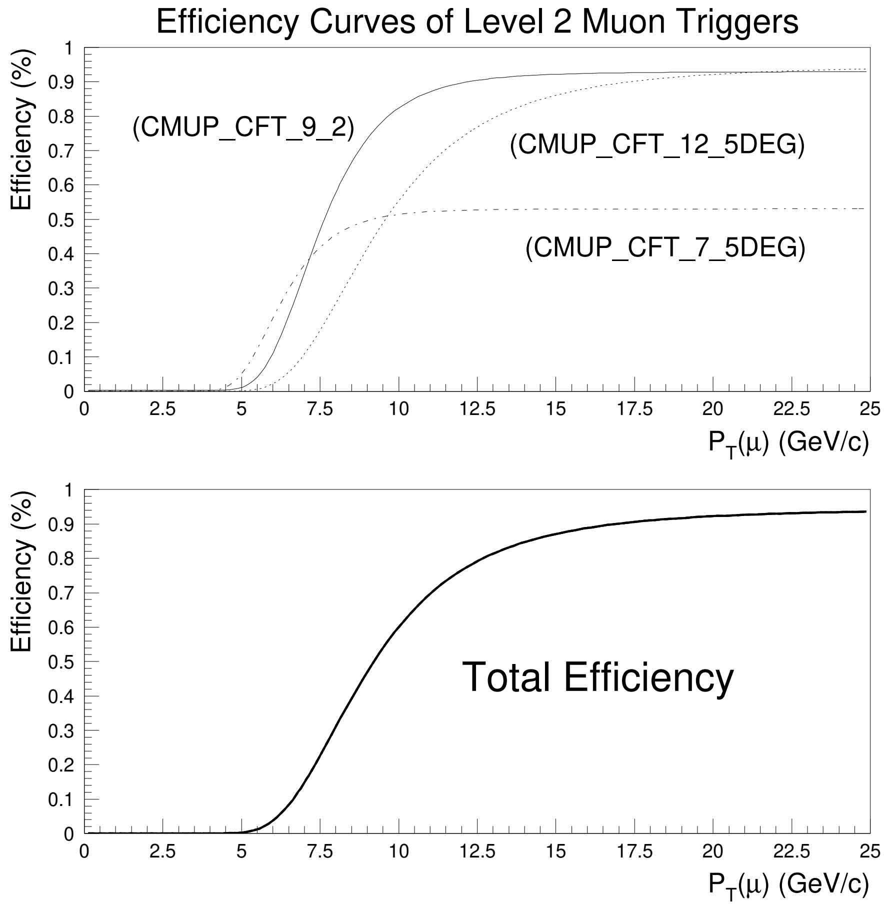

Figure 3.1: *Top: efficiency curves for the dominant Level 2 muon triggers. The 7.5 GeV/c trigger is scaled according to the dynamic prescale applied to it. Bottom: the combined acceptance of the three curves.*

[11]This number is in excellent agreement with other independent determinations of the average prescale of this trigger[58].

The dataset is filtered by requiring the presence of at least one SECVTX tag with positive decay length. Tracking inefficiencies in the data—due to high occupancy of the Central Tracking Chamber and ageing effects—are parametrized in SECVTX as described in Sec. 2.6.4; the "scale factor" effect, also described there, needs instead to be taken into account by rescaling the number of events with tags by the appropriate factor.

Trigger name	A_1	A_2	A_3	A_4	Dyn. Pr.
CMUP_CFT_9_2	0.929	7.49	0.0278	0.003	1.0
CMUP_CFT_12	0.951	9.435	0.03023	0.024	1.0
CMUP_CFT_7_5	0.934	6.304	0.03139	0.183	$(0.5678)^{-1}$

Table 3.4: *Parameters of the efficiency curves for the Level 2 muon triggers used in the simulation of run 1a and run 1b data. The efficiency is parametrized as* $\epsilon_{E_T} = A_1 \times Freq((1/A_2 - 1/P_T)/A_3) - A_4 \times (1 - A_1 \times Freq((1/A_2 - 1/P_T)/A_3)$, *where Freq is the well-known CERN library routine[50].*

After the whole machinery, 1400 events from the run 1a simulation and 7585 from the run 1b simulation are left in the dataset; of these, 261 and 1644 have two SECVTX tags, respectively. To take into account the effect of the SECVTX scale factor, we need to apply some algebra. First, we find the probability of tagging a b-quark jet from Z decay in run 1a and run 1b: if we call T the total number of events before the selection of tags in each dataset, N_1 (N_2) the number of events with exactly one (two) SECVTX tags, and $\mathcal{P}$ the probability of tagging a jet containing a b quark, we have

$$N_2 = \mathcal{P}^2 T \tag{3.1}$$

$$N_1 = 2T\mathcal{P} \times (1 - \mathcal{P}). \tag{3.2}$$

We can solve for T the first one, to obtain

$$T = N_2/\mathcal{P}^2 \tag{3.3}$$

$$\mathcal{P} \times (\mathcal{P}\frac{N_1}{N_2} + 2) - 2 = 0 \tag{3.4}$$

and thence $\mathcal{P} = \frac{2}{2+N_1/N_2}$; we thus find for run 1a $\mathcal{P}^{1a} = 31.4 \pm 1.5\%$ and for run 1b $\mathcal{P}^{1b} = 35.6 \pm 0.6\%$. If now we call f the scale factor for SECVTX tagging ($f^{1a} = 1.21 \pm 0.13$ and $f^{1b} = 1.25 \pm 0.13$ for run 1a and 1b, respectively), these numbers need to be scaled up as follows:

$$N_2' = f^2 N_2 \tag{3.5}$$

$$N_1' = 2Tf\mathcal{P} \times (1 - f\mathcal{P}) = \tag{3.6}$$

$$= 2fN_2 + fN_1 - 2f^2 N_2. \tag{3.7}$$

Inserting the appropriate values for f and $\mathcal{P}$ we finally obtain

$$(N_1')_{1a} = 1245.5 \pm 233.4 \tag{3.8}$$

$$
\begin{aligned}
(N'_2)_{1a} &= 382.1 \pm 85.4 & (3.9)\\
(N'_1)_{1b} &= 6398.7 \pm 1387.3 & (3.10)\\
(N'_2)_{1b} &= 2568.7 \pm 538.0 & (3.11)
\end{aligned}
$$

and the corrected number of single and double tags in the total dataset is therefore $(N_1)_{tot} = 7644 \pm 1407$ and $(N_2)_{tot} = 2950 \pm 545$, respectively; the total number of events with *at least* one SECVTX tag is instead $N_{SECVTX} = 10,595 \pm 1509$. The errors have been propagated by keeping 100% correlated the uncertainties on the run 1a and run 1b SECVTX scale factors.

We can finally use these results to extract the number of $Z \to b\bar{b}$ events expected in the SECVTX dataset. To find it, we just need one more ingredient: the effective integrated luminosity of the Monte Carlo dataset. This can be computed by dividing the number of generated events by the cross section times branching ratio of the process, that we have already computed (Sec. 3.2) from earlier CDF and LEP-SLC results to be $1.110 \pm 0.031(stat \oplus syst) \pm 0.090(lum) nb$. We find that the integrated luminosity[12] of the simulation amounts to $1507 \pm 39\ pb^{-1}$, so that

$$N^{Z\to b\bar{b}}_{SECVTX} = (10,595 \pm 1509) \times \frac{103 \pm 8\ pb^{-1}}{1507 \pm 39\ pb^{-1}} = 724 \pm 121. \tag{3.12}$$

We also display here the number of expected signal events with two SECVTX tags; we will need this number later on:

$$N^{Z\to b\bar{b}}_{2\,SECVTX} = (2950 \pm 545) \times \frac{103 \pm 8\ pb^{-1}}{1507 \pm 39\ pb^{-1}} = 202 \pm 40. \tag{3.13}$$

3.3.2 Other Useful Datasets

Along with the $Z \to b\bar{b}$ simulation, to better understand the behavior of the signal, we have used real Z decays to electron-positron pairs collected during run 1b. These events were selected from high E_T electron triggers by requiring the presence of a tight central electron plus another electron candidate passing looser cuts in the wider rapidity region covered by the Central and Plug calorimeters ($|\eta| < 2.4$). The dataset consists in 6722 events, and can be useful for kinematical studies, particularly when the variables are difficult to model with Monte Carlo.

To study the contamination of other boson signals in our dataset we also generated 500,000 W boson decays to $c\bar{s}$ pairs with PYTHIA V5_7. That process may contribute to the muon dataset, although SECVTX tagging should kill most of it. We further generated 500,000 Z decays to charm pairs, to check how much contamination it could give to the single and double tagged data. These datasets were subjected to the same treatment (tracking degradation, muon trigger simulation, etc.) described for the $Z \to b\bar{b}$ data. We will give an estimate of the contamination of our dataset from these processes in Sec. 6.6.

[12] We neglect the uncertainty due to the integrated luminosity of the dataset used to measure the Z cross section in the e^+e^- final state at CDF, since it is strongly correlated to the uncertainty of the integrated luminosity of our experimental dataset (which enters the computation of expected signal events on the same footing).

For the studies described in Ch. 5 we generated three high statistics datasets: $2,000,000$ $Z \to b\bar{b}$ decays with PYTHIA V5_7, $1,000,000$ $Z \to b\bar{b}$ decays with HERWIG V5_6[44], and $1,000,000$ direct H^0 events with a decay to b-quark pairs, again with HERWIG. These datasets were not searched for SECVTX tags, in order to retain statistics high enough to perform detailed studies.

60,000 $Z \to b\bar{b}$ events were generated at particle level (no detector simulation) for some studies of raw lepton distributions needed for the checks described in App. B.

Further, 1,000,000 $Z \to b\bar{b}$ decays were generated and those containing an electron in the final state were filtered. We used this dataset for checks in the Inclusive Electron sample. A description of the result of the same analysis on the inclusive electron data is given in App. A.

Finally, we used a sample of about 20,000 events collected with a minimum bias trigger[13] for the studies of multiple interactions described in App. C.

[13]Basically, the minimum bias trigger required just an east-west coincidence in the Beam-Beam Counters.

Chapter 4

Event Selection

In the present chapter we discuss our data selection, from the point where we left our experimental dataset (that we named "SECVTX" sample) to our final sample of 588 events. In Sec. 4.1 we use SECVTX tagging to further reduce the QCD background, while in Sec. 4.2 we study the kinematic characteristics of the $Z \to b\bar{b}$ signal, and discuss two variables that may be used to discriminate from QCD production those processes yielding a low color radiation flux besides the two leading jets. These variables are not easy to simulate with Monte Carlo programs, so we need to study them with the experimental data, as is described in Sec. 4.2.2. We summarize our data selection in Sec. 4.2.3.

4.1 From Single to Double Tags

As can be seen in Table 4.1, the signal-to-noise ratio is of the order of 1/150 in the SECVTX sample. The picture is actually not so bleak, since what matters is the signal fraction in a limited invariant mass window around the Z pole: we may expect the S/N ratio to be as much as three times larger there, since the background is spread over a much wider range of dijet masses than the resonance for which we are looking. Nevertheless, the signal is still too small: even if our knowledge of the shape of the mass distribution for the background were perfect, we could not search the experimental spectrum for a small deviation and call it a signal of the searched process, because the statistical fluctuations of the background spectrum would make it insignificant. Accordingly, we need both to understand as precisely as possible the background spectrum and to devise a set of selection criteria capable of increasing the S/N ratio to a value larger than the relative uncertainty of the background shape. We shall describe our method of increasing the S/N ratio here, while our method of modeling the background shape will be described in Ch. 6.

In order to find the best discriminating tools for the $Z \to b\bar{b}$ signal, we must understand what is the dataset composed of at the SECVTX level. Coming to our help is a detailed analysis performed by another group at CDF, whose aim was the understanding of the b purity of the SECVTX dataset, used in that instance for $B^0-\bar{B}^0$ mixing studies. For that analysis a large sample of Monte Carlo $b\bar{b}$ and $c\bar{c}$ events was generated, and two story-telling distributions were studied: the muon P_T^{rel} and the reconstructed $c\tau$ distribution of the secondary vertices. The P_T^{rel} is defined in their analysis[1] as the projection of the muon momentum vector orthogonally to the vector sum of all other charged tracks belonging to the same jet: it is a quantity sensitive to the mass of the decayed particle, and therefore a good discriminant between muons coming from the semileptonic decay of b quarks, muons coming from c quarks, fake muons, and muons coming from the decay of a light meson. The distribution of $c\tau$, reconstructed from the transverse distance between the primary and the secondary vertex[2], is of course another good discriminant between the cited contributions. A simultaneous fit of these distributions to the sum of the various contributing processes allowed the understanding of the sample composition for the subsample of SECVTX events having the muon in the b-tagged jet[59]: it was found that the largest contribution to the sample was due to QCD $b\bar{b}$ production, with fake and $c\bar{c}$ components at the level of $5 \div 10\%$.

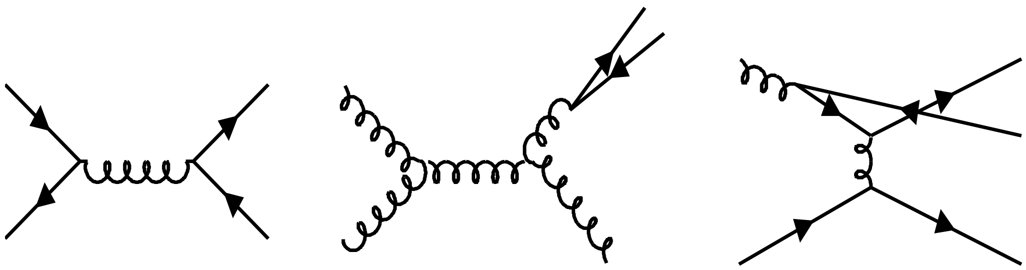

Figure 4.1: *Feynman diagrams of leading order QCD processes producing final state b-quark jets: direct production, gluon splitting, and flavor excitation.*

The $b\bar{b}$ component is generally thought of as a combination of the three leading-order processes of *direct production*, *gluon splitting*, and *flavor excitation* (Fig. 4.1). Of these, only direct production may give a sizable fraction of events where two well-separated, central jets both correspond to a b quark. The gluon splitting process will instead yield most frequently a pair of b-quark jets close in angle and recoiling against another jet, which most of the time comes from a high-P_T gluon; the flavor excitation process is characterized by a b-quark jet emitted at very high rapidity, and hence very difficult to tag by the SECVTX algorithm. The three processes are roughly equivalent in yield if we select—as we did—a sample with a SECVTX tag and a muon, without asking them to belong to one and the same jet; therefore, we have reason to believe that the experimental data will have a much smaller fraction of double SECVTX tags than the $Z \to b\bar{b}$ decays, which on the contrary give a pure $b\bar{b}$ final state with two back-to-back, central jets both containing a bottom meson or baryon. Moreover, the $c\bar{c}$ and fake components of the background will be swept away by the requirement of two SECVTX tags, given the much lower probability of tagging a c quark (Sec. 2.6.4) and *a fortiori* a light quark or gluon: only the

[1]Different definitions of this quantity are found in the literature, according to what is agreed to be defined the "jet axis".

[2]To obtain the true $c\tau$ the b-quark momentum has to be inferred from the total momentum of the tracks belonging to the secondary vertex.

direct $b\bar{b}$ production process will survive. In fact, we do see a large difference in the double tag rate in the experimental sample as opposed to the Monte Carlo dataset: only $5,479$ events out of $105,782$ have a second SECVTX tag in the data (5.2%), while $2,950$ $Z \rightarrow b\bar{b}$ Monte Carlo events out of $10,595$ (27.8%) carry a second SECVTX tag[3]. We therefore choose to restrict our search to events with two jets both containing a secondary vertex with a positive decay length: the signal-to-noise ratio in this sample should be close to 1/27.

Sample	Run 1a	Run 1b	Total	$Z \rightarrow b\bar{b}$ PYTHIA	Eff. (%)	Events in 103 pb^{-1}	N/S
Initial				1,673,000		$(116 \pm 9)10^3$	
Trigger	2,724,593	2,690,162	5,414,755				
SECVTX	19,925	85,857	105,782	10,595	0.63	724 ± 119	146
2 tags	816	4,663	5,479	2,950	27.8	202 ± 40	27

Table 4.1: *Number of events selected by each of the cuts described in the text for the experimental data and for the $Z \rightarrow b\bar{b}$ Monte Carlo, efficiency of each cut for the Z signal, expected Z events in the total dataset, and noise-to-signal ratios. The uncertainties quoted are the quadrature sum of statistical errors and systematic contributions due to the determination of integrated luminosity (Sec. 3.2.1), muon trigger efficiency (Sec. 3.3.1), and the SECVTX scale factor (Sec. 2.6.4). The signal efficiency in the SECVTX selection is an absolute number, while the efficiency for the double tag selection is relative to the former.*

4.2 Kinematical Tools for the Signal Selection

4.2.1 Identification of Discriminating Variables

After having required that the two leading jets be tagged by SECVTX, with positive-reconstructed decay length, we can think of the dataset as a virtually pure sample of $b\bar{b}$ events. One could demonstrate it with a simple computation, starting from the measured components of $b\bar{b}$, $c\bar{c}$, and light quark QCD backgrounds found in the sample with one SECVTX tag and a good muon candidate, by multiplying their size by the expected tagging rate of the second jet in the three categories; however, a simple comparison of the characteristics of the secondary vertex and the muon found in the double tagged data with those of the PYTHIA $b\bar{b}$ Monte Carlo is simpler and visually more convincing. This is shown in Fig. 4.2: all distributions agree closely with the Monte Carlo expectations. In particular, the pseudo-lifetime of the secondary vertices and the muon P_T^{rel} show no space for a significant contamination from $c\bar{c}$ events[4]. The muon P_T shows a good agreement as well, and the E_T measured in the tower hit by the muon track speaks in favor of a very low contamination from fake muons, which would give an excess at high values of E_T .

[3]These numbers are corrected for the SECVTX scale factor: see Sec. 3.3.1 for their computation.

[4]These distributions are sensitive to the lifetime and the mass of the decayed quark, respectively: a sizable c-quark contribution would show up as an excess for small values of each variable.

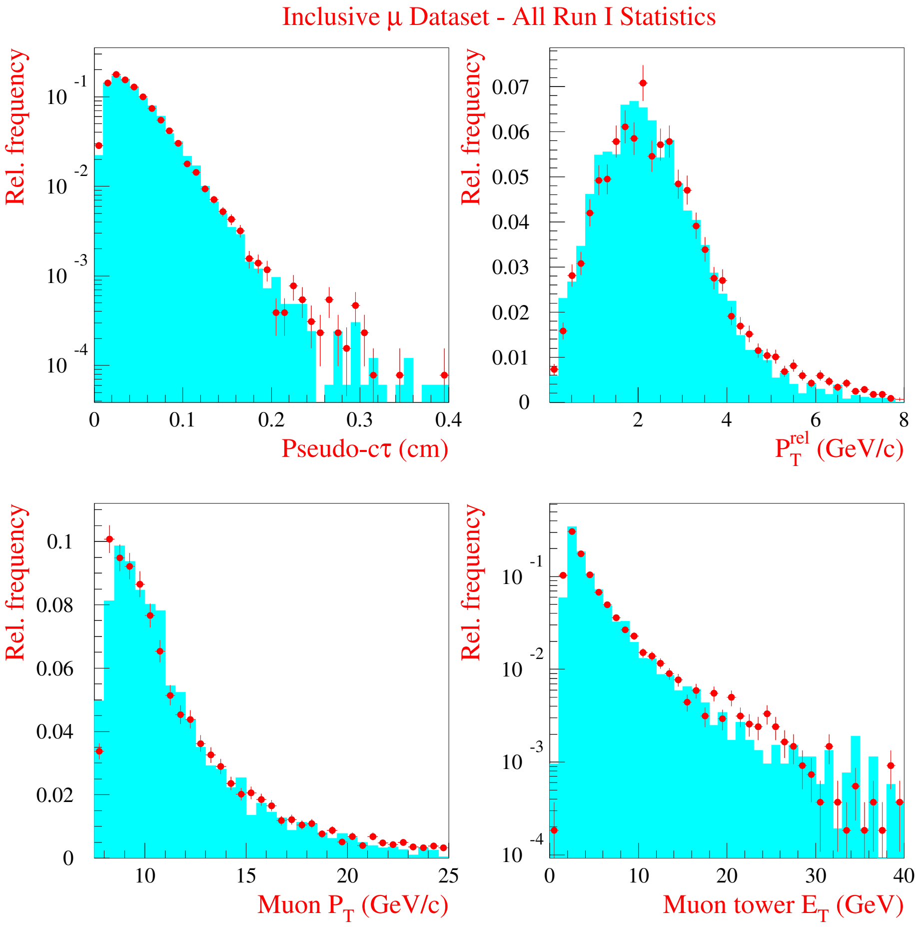

Figure 4.2: *These plots show comparisons between data and $b\bar{b}$ Monte Carlo events after the requirements that both leading jets contain a SECVTX tag. Top left: the pseudo-lifetime of the secondary vertex; top right: the muon P_T^{rel} spectrum; bottom left: the muon P_T spectrum; bottom right: the E_T measured in the tower hit by the muon track. All distributions are normalized to unit area.*

Figure 4.3 shows the dijet mass spectrum for events with two SECVTX tags[5]. Although there may be space in the invariant mass spectrum for the 202 ± 40 $Z \to b\bar{b}$ events expected at this level of selection[6], the overall agreement in shape with the spectrum of events containing one SECVTX tagged jet and one taggable jet

[5]Here, for the computation of the invariant mass, we use jet energies corrected with the standard routine (Sec. 2.6.1) plus the jet corrections described in Ch. 5. The latter, appropriate for b-quark jets in clean dijet events, increase the relative mass resolution by about 50%.

[6]The computation of the number of $Z \to b\bar{b}$ events expected after double SECVTX tagging has been given in Sec. 3.3.1.

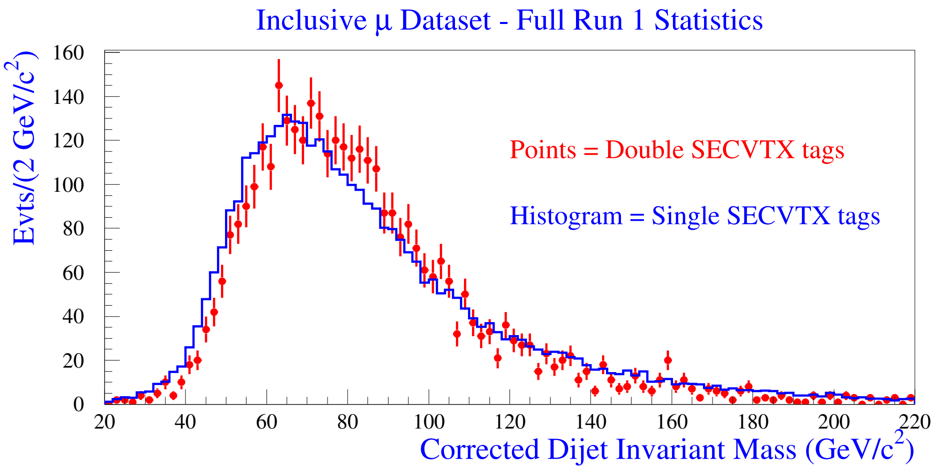

Figure 4.3: *Invariant mass spectrum of the events with two SECVTX tags (points). The agreement with the background-enriched sample of events with one tagged and one taggable jet (histogram) is good. Note the excess of single tags at low mass, due to the correlation between tagging probability and jet E_T (we will discuss it in detail in Ch. 6), and the small excess of double tags at* 90 GeV/c^2. *Jet momenta have been corrected as described in Ch. 5.*

(expected to contain a fraction of signal roughly five times smaller) is fair[7], and one can still claim no evidence for the presence in the experimental data of Z decays, that should cluster around 90 GeV/c^2, with a r.m.s of about 12 GeV/c^2. We must therefore find some other tool to increase the discrimination of the electroweak production of b-quark jets from the strong interaction.

From a theoretical standpoint, one expects that the two production processes be different in many ways. The two initial state partons of the Z production have to be a quark-antiquark pair, and the process is time-like. On the contrary, in the QCD creation of a $b\bar{b}$ pair both a quark-antiquark and a gluon-gluon pair can give rise to a time-like direct production process, and space-like diagrams may contribute as well (Fig. 4.4). Moreover, b-quark pairs produced by the splitting of a final state gluon may still be present in the double tagged sample, although in comparatively small amounts. Many of the QCD processes may thus result in a pair of outgoing partons with a flatter pseudorapidity distribution than those from the searched Z decay; however, the double SECVTX tagging and the requirement of a tight CMUP muon candidate in one of the two jets result in pseudorapidity distributions that are already very constrained and well peaked at zero, due to the acceptance of the SVX

[7]As we note in the caption of Fig. 4.3, the double tagged data also show a deficit at small invariant mass: that is due to the slight correlation between tagging probability and jet E_T . This bias will be discussed in detail in Ch. 6.

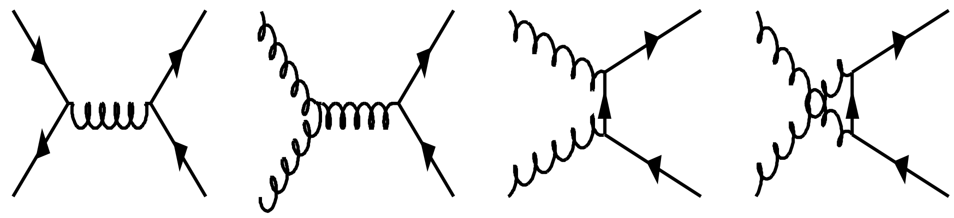

Figure 4.4: *Feynman diagrams of all tree-level QCD processes that may give rise to two back-to-back b-quark jets in the final state.*

and the CMU-CMP detectors.

If one examines the color structure of the involved diagrams, one however notices a marked difference between $Z \to b\bar{b}$ and $g \to b\bar{b}$. In the QCD processes there is a color connection between initial and final state, absent in the Z production. Furthermore, while both the initial and the final state in the $Z \to b\bar{b}$ process are in a color singlet configuration, the opposite is true for QCD tree level diagrams. Finally, we have to note that QCD processes have on average a higher color charge in the initial state, due to the contribution of gluon-gluon scattering. These considerations alone may lead us to believe that QCD gives rise to processes with a higher color radiation accompanying the two b-quark jets. In addition, the pattern of this radiation must in principle be different (see Fig. 4.5). In QCD processes, in fact,

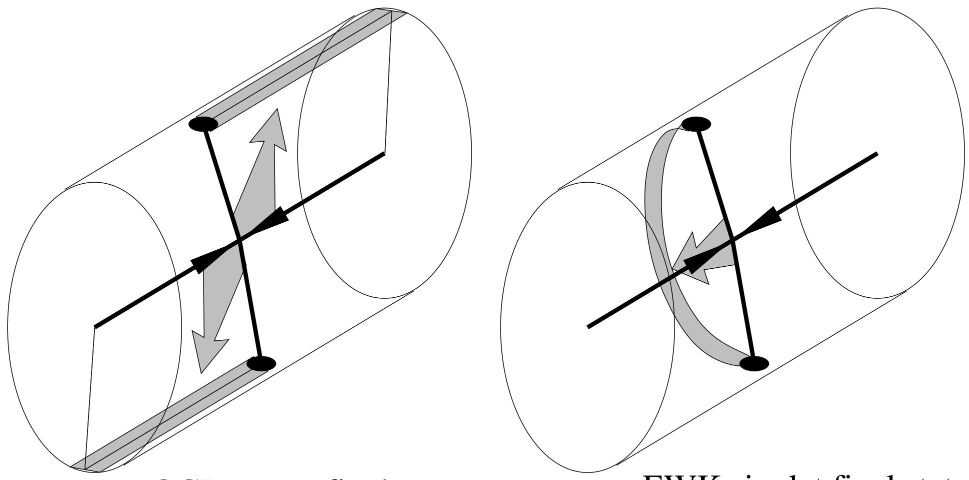

Figure 4.5: *The different patterns of radiation emission at leading order for the QCD and electroweak production of a pair of b quarks. On the left, the antenna pattern of the color strings in a QCD event create preferred radiation planes; on the right, the color singlet produced in a EWK interaction is shown to emit predominantly between the two final state partons.*

the color connection present at leading order between the two final state partons and the initial state should give rise to an enhanced radiation flow in the planes containing each of the two leading jets and the z axis: color coherence[61] prescribes that the radiation from the incoming and outgoing partons interferes constructively in the regions highlighted in Fig. 4.5, while the final state color singlet produced in the Z decay will be able to emit soft radiation mainly in the region between the two leading jets[60].

Inspired by these considerations, we have studied a wide variety of variables constructed with the calorimetric informations at our disposal, to try and separate the signal from the QCD background. We concluded that no event-by-event discrimination is possible by the use of variables meant to pinpoint the differences in the radiation pattern[8]: not surprisingly, the topological differences we have mentioned above amount to very subtle and non-leading effects; they can be detected only in very high energy events—where the effects of soft radiation are larger on an absolute scale—by statistical methods. A simpler and less arguable result for our purpose can be obtained with variables sensitive to the *total* amount of the color radiation. Again, a large number of variables has been investigated for this purpose. The two that seem to score better are the azimuthal angle between the two b jets, $\Delta\Phi_{12}$, and the sum of uncorrected transverse energies of all the JETCLU clusters in the event, subtracted by the E_T of the two leading b jets, $\Sigma_3 E_T$. Both variables may provide some discriminating power between a high radiation process and the colorless production of a dijet system (see Fig. 4.6). Unfortunately, they are very difficult to simulate with Monte Carlo programs, since they depend strongly on the detailed features of initial state radiation and underlying event.

4.2.2 Study of the Kinematic Variables

To verify that Monte Carlo simulations cannot be trusted in the modeling of the two cited variables, we cannot study $Z \to b\bar{b}$ decays, since we would then be unable to perform comparisons with the real data. Instead, we can use $Z \to e^+e^-$ decays, given the availability of a reasonably sized sample of these events in the CDF high-P_T electron dataset. $Z \to e^+e^-$ decays are of course identical to the $Z \to b\bar{b}$ ones as far as initial state radiation is concerned: by comparing these data to a simulation of the same process we can thus rate the reliability of the Monte Carlo in the modeling of initial state radiation. The dataset used for this purpose is a sample of 6722 $Z \to e^+e^-$ decays collected during run 1b (see Sec. 3.3.2). The Monte Carlo data comes from $10,000$ $Z \to e^+e^-$ decays generated with PYTHIA V5_7: the electron cuts and a simple trigger simulation select a sample of 4108 events. $Z \to e^+e^-$ data can be studied with ease with respect to the two kinematic variables described above: when reconstructed by the clustering algorithm, in fact, electrons appear as jets, so

[8]The different color fluxes pictured in Fig. 4.5 suggest to construct regional sums of the energy deposited in the calorimeter at the same azimuth of the two leading jets, and compare these to the amount of energy flowing in orthogonal planes. Other approaches involve the use of event-shape variables, such as Aplanarity or Sphericity; the rapidity of sub-leading energy clusters may also come to mind.

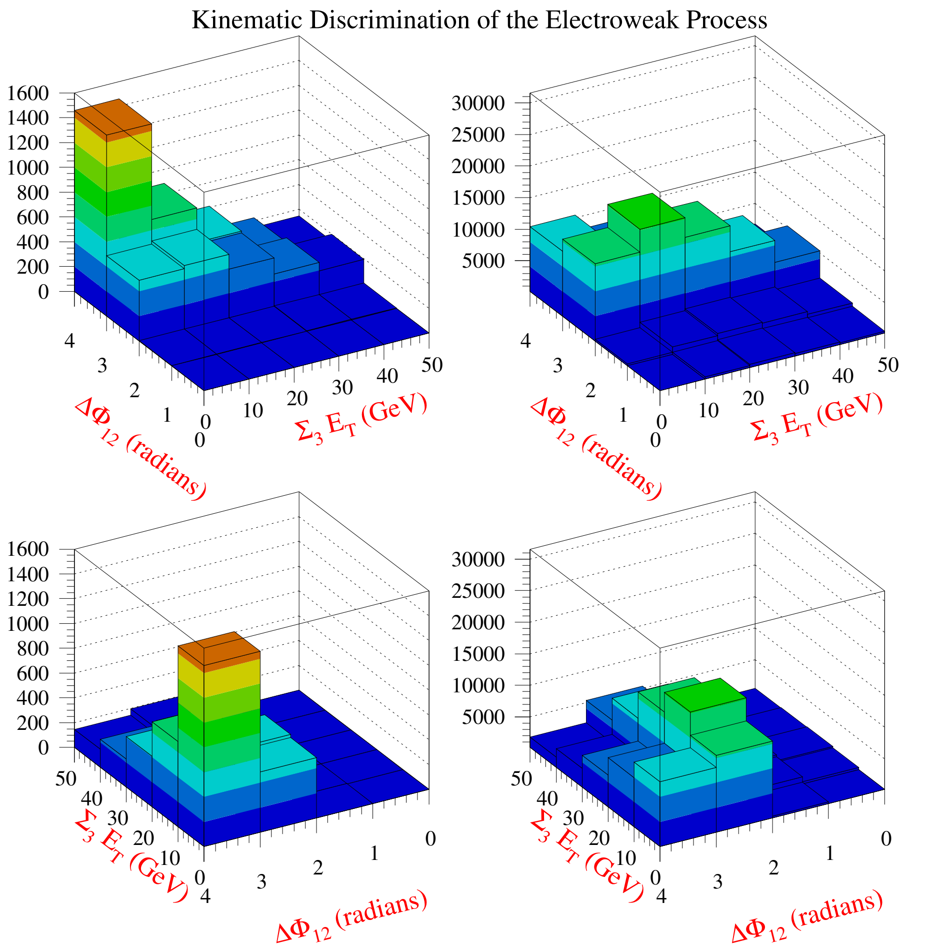

Figure 4.6: *These plots demonstrate the discriminating power of the two kinematic variables $\Sigma_3 E_T$ and $\Delta\Phi_{12}$ between a QCD and an electroweak process. On the left is shown the $Z \to e^+e^-$ data, on the right the SECVTX data. Each plot is shown from two different view angles.*

that the same definition may be used for $\Delta\Phi_{12}$ and $\Sigma_3 E_T$ in these datasets.

Before performing a comparison of the distribution of $\Delta\Phi_{12}$ and $\Sigma_3 E_T$, we must take into account the effect of multiple interactions. As is described in App. C, in order to correctly model the effect of additional minimum bias interactions in the same bunch crossing that produced the triggering $Z \to e^+e^-$ event, one must mix the calorimeter banks of the simulated event to those of real minimum bias events, taking good care to match the distribution of instantaneous luminosity of the latter to that of the data one wishes to compare the Monte Carlo events with.

Once the mixing is performed, any discrepancy between data and Monte Carlo $Z \to e^+e^-$ decays in the spectra of $\Delta\Phi_{12}$ and $\Sigma_3 E_T$ can be blamed with good reason on the modeling of initial state radiation and underlying event. As Fig. 4.7 shows, there is indeed a discrepancy in both variables: the Monte Carlo underestimates the amount of radiation emitted by the incoming partons, and cannot reproduce

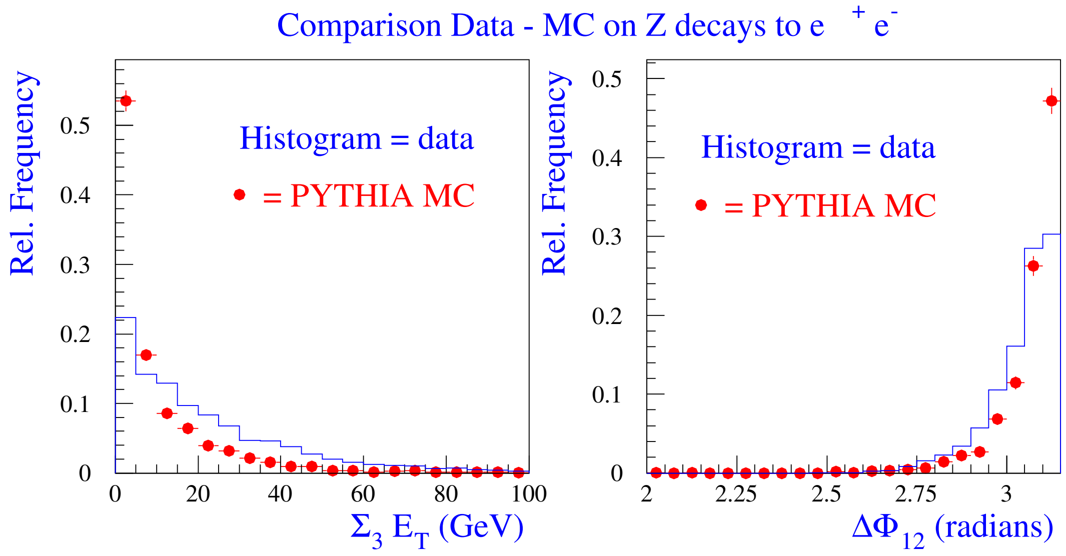

Figure 4.7: *Comparison of experimental $Z \to e^+e^-$ decays with the PYTHIA Monte Carlo in the kinematic variables $\Sigma_3 E_T$ and $\Delta\Phi_{12}$. The Monte Carlo underestimates the amount of soft radiation off the initial state partons, as well as the transverse boost of the produced bosons.*

the transverse momentum of the produced Z boson[9]. This is hardly a surprise, given that no fine-tuning of the Monte Carlo parameters governing these effects has been performed. Therefore, we cannot rely on a Monte Carlo simulation to define a selection strategy based on the two kinematic variables: we are bound to use the experimental data as far as we are able to.

Another useful check that can be performed to clarify the issue is a comparison between the $Z \to e^+e^-$ simulation and a $Z \to b\bar{b}$ simulation obtained with the same Monte Carlo and the same mixing procedure described above. A comparison of Z decays to dielectron pairs with decays to b-quark pairs allows us to focus on the effect of final state radiation off the color-singlet $b\bar{b}$ pair produced by the Z boson. We have reason to believe that final state radiation is well modeled by PYTHIA, which has been shown to correctly describe the color coherence effects displayed by the CDF data[62, 63]: therefore, a comparison of the leptonic and hadronic decays of the Z bosons enables us to understand how sensitive the two kinematic variables are to final state radiation in the data. This comparison is shown in Fig. 4.8: one notices that the difference is small, but not negligible. Both the $\Sigma_3 E_T$ and the $\Delta\Phi_{12}$ distributions are slightly broader in $Z \to b\bar{b}$ decays, due to the contribution from final state radiation. The acceptance of a cut $\Sigma_3 E_T < 10$ GeV is $63.8 \pm 3.2\%$ for $Z \to b\bar{b}$ decays and $69.8 \pm 2.2\%$ for $Z \to e^+e^-$ decays; after requiring $\Sigma_3 E_T < 10$ GeV, the acceptance of a cut $\Delta\Phi_{12} > 3$ is $76.1 \pm 4.5\%$ and $84.7 \pm 3.0\%$, respectively. The difference in acceptance of these requirements, attributable to the

[9]Transverse momentum is normally the variable used to study the production process of gauge bosons at a $p\bar{p}$ collider; the $\Delta\Phi_{12}$ between the two jets in the particle decay is linked to it by a fixed relation if the transverse mass of the dijet pair is given.

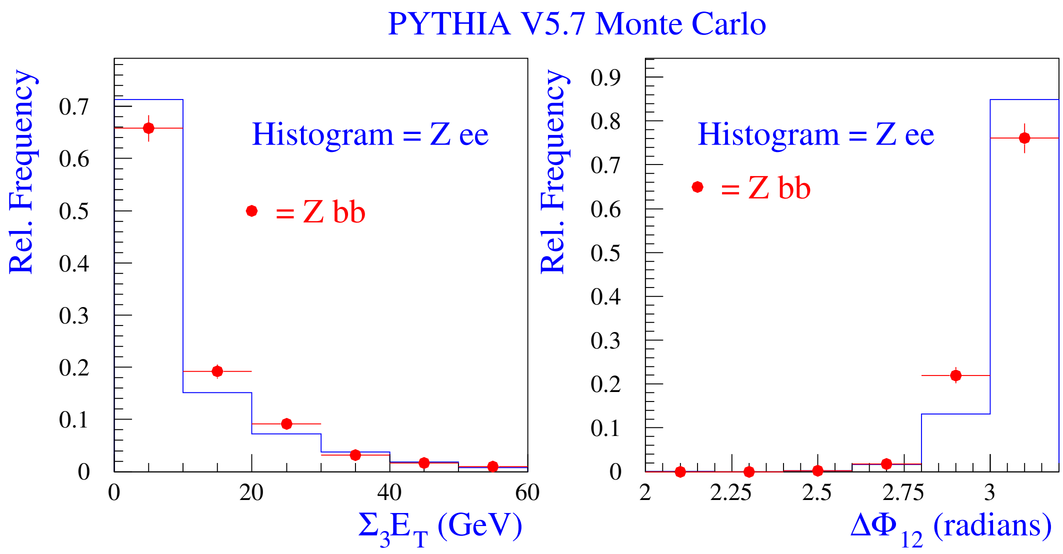

Figure 4.8: *Comparison between $Z \to e^+e^-$ and $Z \to b\bar{b}$ decays (PYTHIA V5_7 Monte Carlo) in the two kinematic variables $\Sigma_3 E_T$ and $\Delta\Phi_{12}$. Left: the $\Sigma_3 E_T$ distribution is slightly broader for $Z \to b\bar{b}$ decays. Right: $Z \to e^+e^-$ decays show a more back-to-back topology, underlining the effect of final state radiation in the azimuthal opening angle of the jets.*

difference between the two processes, is therefore of the order of 10%.

4.2.3 Selection of the Final Sample

The experimental dataset of $Z \to e^+e^-$ decays can be used for a first-order estimate of the efficiency of the selection cuts on the chosen kinematic variables: as we discussed above, we can assign a 10% systematic uncertainty on the efficiencies thus evaluated, to account for the absence of final state radiation in the $Z \to e^+e^-$ events used for the estimates. A comparison between the $\Sigma_3 E_T$ spectrum for the SECVTX data[10] and the $Z \to e^+e^-$ data is shown in Fig. 4.9. The value chosen for the cut, 10 GeV, will be justified in Sec. 6.3; its acceptance for SECVTX data (which must be understood as a pure background sample at this level of selection) is $18.9 \pm 0.3\%$, and its efficiency on $Z \to e^+e^-$ decays is $36.4 \pm 1.0 \pm 3.6\%$.

After the selection of events with $\Sigma_3 E_T < 10$ GeV, we requre the two leading jets to be separated in azimuth by more than three radians. The efficiency of the cut $\Delta\Phi_{12} > 3$ radians can be estimated to be $74.8 \pm 2.6 \pm 7.5\%$, again assigning a 10% systematic uncertainty due to the absence of final state radiation in the $Z \to e^+e^-$ events used for the computation of the efficiency; its acceptance for the SECVTX data passing the former requirement is instead $54.5 \pm 1.4\%$. The justification of the value chosen for the cut will be given in Sec. 6.3. The $\Delta\Phi_{12}$ distribution of the two samples is shown in Fig. 4.10.

[10]We select events having a dijet invariant mass close to the Z pole in order to picture the behavior of background events that are kinematically close to the signal: the energy flow outside the leading jets must depend—albeit not dramatically—on the Q^2 of the hard process.

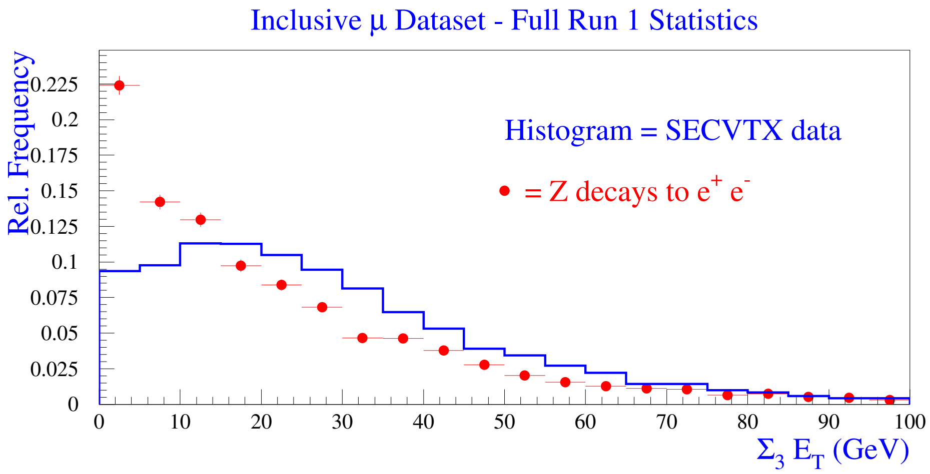

Figure 4.9: *$\Sigma_3 E_T$ spectrum for SECVTX data selected with the cut $|M_{jj} - 90| <$ 20 GeV/c^2 (histogram) and for $Z \to e^+e^-$ decays (points).*

The two cuts we have just discussed select 588 events out of the 5479 double tags in the inclusive muon sample. The number of $Z \to b\bar{b}$ events left in the sample after this selection can be computed with the help of the efficiency numbers quoted above: we expect to have collected 55 ± 14 signal events in the final dataset. The S/N ratio has improved to a value close to 1/5 in the Z mass region: in Ch. 6 we will try to demonstrate that this dataset really contains a significant fraction of $Z \to b\bar{b}$ signal. For the time being, before we end this chapter, we can take a look at the mass distribution of the selected events (Fig. 4.11). What was only a pale hint of a signal after the selection of double SECVTX tags appears now as a remarkable peak: the data seem to confirm our belief that the kinematic selection has increased the significance of the signal.

In the following chapter we will discuss some improved jet corrections that allow us to reconstruct the $Z \to b\bar{b}$ decay at the expected value, and to further improve the S/N ratio.

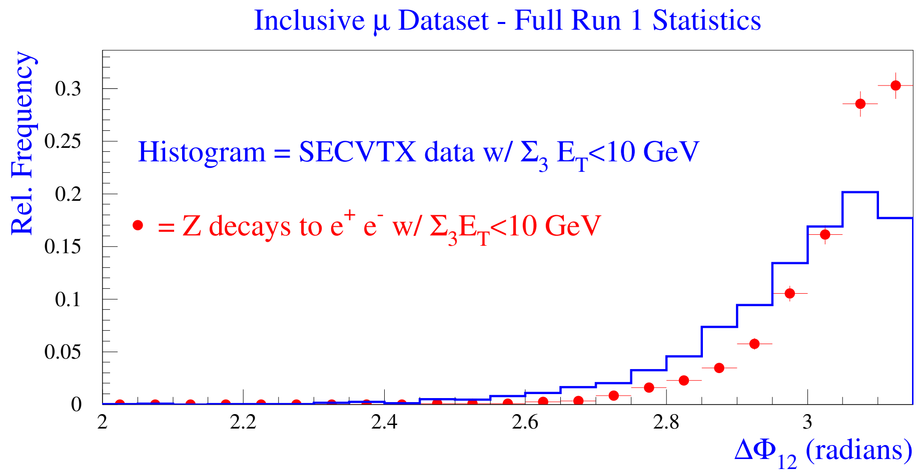

Figure 4.10: *$\Sigma_3 E_T$ spectrum for SECVTX data selected with the cut $|M_{jj} - 90| < 20$ GeV/c^2 (histogram) and for $Z \to e^+e^-$ decays (points), after the application of the cut $\Sigma_3 E_T < 10$ GeV.*

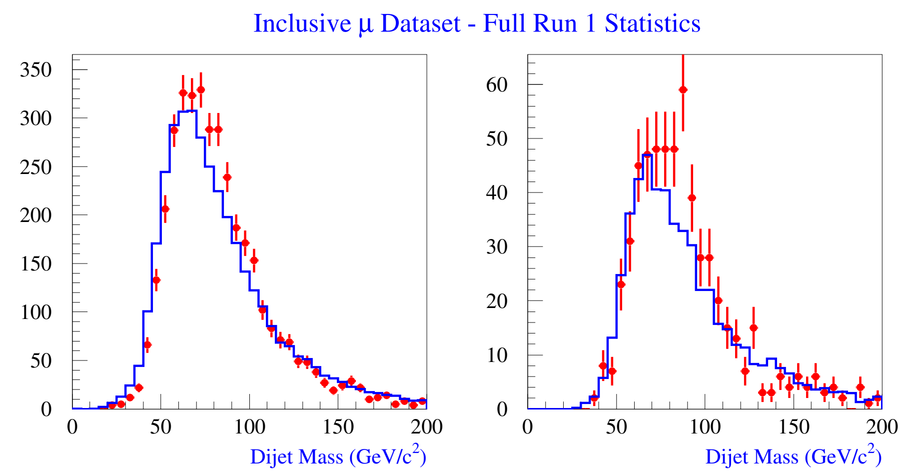

Figure 4.11: *Left: the dijet invariant mass distribution of the 5479 events having two positive SECVTX tags (points with error bars) is compared to events having one tagged jet and one taggable jet (histogram); right: after the kinematic selection the $Z \to b\bar{b}$ contribution can be clearly seen around 90 GeV/c^2. The normalization of the histograms is arbitrary; the improved jet corrections described in Ch. 5 have been applied.*

Chapter 5

Reconstruction of the Z Mass

In the present chapter we discuss the improvement of the mass resolution for those very special $Z \to b\bar{b}$ decays selected in our final dataset. After a brief introduction, we describe the Monte Carlo datasets used for our study in Sec. 5.2. In Sec. 5.3 we describe our correction method, by which we improve the jet momentum resolution: to do that, we use all those observables which show a correlation to the error in the jet energy measurement. Sec. 5.4 describes the results of the improved jet corrections, and Sec. 5.5 describes some tests of the performance of our corrections on other Monte Carlo datasets—and, of course, on the real data.

5.1 Introduction

We have described in Sec. 4.2 the event selection for the $Z \to b\bar{b}$ search in the muon dataset. To increase the signal significance we have selected a very peculiar sample of events by means of double SECVTX tagging and two kinematic cuts aimed at isolating a clean dijet final state: the dataset can thus be seen as a pure $b\bar{b}$ sample, where the two quarks have been produced in the absence of any hard gluon emission either in the initial or in the final state; moreover, at least one of the jets is required to contain a high-P_T muon track. We have thus exploited all but the most prominent among the discriminant characteristics of the signal: the fact that the Z decay products must have a total invariant mass equal to the Z pole mass, while the QCD $b\bar{b}$ continuum has no definite value for it, and will display a roughly exponential distribution, betraying the falloff with x of the parton distribution functions of the two colliding quarks or gluons in the initial state. Of course, we have saved the invariant mass till the end, because it is the most convincing and unequivocal variable to use in a counting experiment or in a multi-component fit: in the first case one can then use the absence of any signal outside the peak position to verify his or her understanding of the background normalization, by comparing the observed counts with the expectations for the background; in the second case one can profit from the very different shapes for the two processes, and use a statistical estimator to evaluate the relative composition of the sample.

The computation of the invariant mass of the dijet system is therefore a really important issue in our search. Both the counting experiment and the two-component fit are heavily dependent on the resolution we can attain in reconstructing the Z mass: a narrower peak can give a tangible contribution to the signal significance in a counting experiment by increasing the S/N ratio in the region of interest, and it may also reduce the uncertainty of the fit, due to the increased difference between the shape of the mass distribution of signal and background.

The mass resolution for a dijet system such as that coming from a Z decay—with a mostly back-to-back configuration—depends essentially on the resolution of the jet energy, the contribution due to angular resolution vanishing at first order. As we have seen in Sec. 2.2, CDF is endowed with good performance e.m. calorimeters, while the hadronic calorimeters have a rather poor resolution, their stochastic term ranging from $70\%/\sqrt{E}$ in the central scintillator towers to more than $100\%/\sqrt{E}$ in the argon-ethane Plug and Forward calorimeters. All in all, to a good approximation the typical knowledge of a parton's momentum we can attain by measuring the energy deposition in the calorimeters has an error of 10%, regardless of its energy. One arrives at this resolution after the use of CDF's standard jet correction algorithm[25], QDJSCO[1], a routine that corrects for the nonuniformity and leakage of the various components of the calorimeter system and scales the measured jet energies according to the known response function of the calorimeter to test beam pions and electrons. The standard algorithm, however, does not use all the information available from other detector components to improve the jet energy resolution. Moreover, it ignores any information relative to the presence of muons or the flavor of the originating parton. Muons are minimum ionizing particles and they lose on average only about $2 \div 3\ GeV$ of energy passing through the CDF calorimeters, regardless of their momentum. Further, the semileptonic decay that originates our CMUP muons gives rise to a neutrino with a broad momentum distribution, measured to have mean value around $5\ GeV/c$ in $Z \to b\bar{b}$ decays. Therefore, the muon jet energy in the calorimeter is rather heavily underestimated. If we wish to reconstruct a peaked signal from the Z in our sample, we must correct the muon jet to take into account these factors. But we can push that further, and also try to improve the correction of the other jet, using our acquired knowledge that it originated from a b quark, and exploiting the absence of any hard radiation—two facts peculiar to our final dataset, we emphasize again—to get the best possible resolution for both of them. The jet corrections that we wish to devise have not only to infer the true b-quark momentum from the detector observables at our disposal, but also increase the momentum resolution as much as possible: that will allow us to obtain a more significantly peaked structure in the mass spectrum.

In what follows we will describe the datasets used for the present study, our correction method, and the checks performed to verify the independence of the corrections on the model used in the Monte Carlo for the parton fragmentation and on the jet energy. We will defer to Ch. 6 the description of the results of our improved jet corrections on the experimental data.

[1] A description of the algorithm is given in Sec. 2.6.1.

5.2 The Data Samples

We use for this study a Monte Carlo dataset obtained from the simulation of two million $Z \to b\bar{b}$ events with the PYTHIA V5_7 Monte Carlo; the MRSD− parametrization of the structure functions is used to generate the process. The QFL′ (QFL) detector simulation is applied to $1,700,000$ $(300,000)$ simulated events for run 1b (1a), and we filter them by requiring the presence of a tight muon according to the requirements described in Sec. 2.6.2[2]. The selection leads to $44,938$ events. To these, a Level 2 trigger simulation is applied as described in Sec. 3.3.1: this reduces the sample to 30,555 events.

We then apply to the Monte Carlo events the two kinematic cuts that we use in the real data: we require the two leading jets to have an opening angle $\Delta\Phi > 3$ radians in the transverse plane, and the sum of additional clustered transverse energy to be $\Sigma_3 E_T < 10\ GeV$. These cuts lead to $18,720$ events. To be completely consistent with what we do in the real data, we should then require having two jets tagged by the SECVTX algorithm; we cannot do that, since we would end up with statistics too low to allow for precision studies of the jet momentum resolution. We therefore only require that the two leading jets in the Monte Carlo events be correctly associated (by a $\Delta R < 0.7$ requirement in $\eta\phi$ space) to the two final state b quarks emitted in the Z decay: that is, we discard the small fraction of spurious events where a gluon happened to form a leading jet. We end up with $18,524$ events: they will be the starting point of our jet correction study.

	PYTHIA $Z \to b\bar{b}$	HERWIG $Z \to b\bar{b}$	HERWIG $H \to b\bar{b}$
Generated	2,000,000	1,000,000	1,000,000
Trigger	30,555	14,322	15,244
Kin. Sel.	18,720	7,952	8,892
Two b jets	18,524	7,878	8,843

Table 5.1: *Statistics of the three Monte Carlo datasets.*

We also generated two datasets with the HERWIG V5_6 Monte Carlo (same PDF's, detector and decay simulation; 1b simulation only) for some useful checks: one million $Z \to b\bar{b}$ events and one million $H \to b\bar{b}$ events (we chose $M_H = 120\ GeV/c^2$ in that simulation). The same selection chain applied to these datasets leaves us with 7878 and 8843 events, respectively (see Table 5.1). The difference in acceptance (15%) between the HERWIG and PYTHIA Z datasets is due to the conspiracy of two separate effects: the different fragmentation models used by the two Monte Carlo programs, that causes PYTHIA to predict a higher acceptance for central muons, and the different simulation of initial state radiation, due in turn to the way the parameters governing it are tuned in the two generators; these factors, however, are not a relevant point for a jet resolution study, as the results of this chapter

[2]We do not discard the small fraction (2.2%) of reconstructed muons that do not match momentum and direction of a generated muon, consistently with what has been done with the real data. These extra muons are either fakes or they come from decays in flight of K and π mesons (the latter are usually called "punch-through" muons).

will soon show. On the other hand, the difference between the H and Z acceptance in HERWIG (11%) is mainly due to the higher momentum carried on average by the muons coming from the b quarks of a Higgs decay with $M_H = 120\ GeV/c^2$.

5.3 Correction Method

5.3.1 General Philosophy

As we have already pointed out, the jets of our dataset are really special. They come from b-quark hadronization, and have almost no radiation surrounding them. The events show a very clean dijet topology with one of the jets containing a $P_T > 6\ GeV/c$ CMUP muon, which in turn implies the presence of one energetic neutrino. The standard jet corrections (Sec. 2.6.1) do a poor job when dealing with these events, since they ignore their particular features; indeed, the reconstruction of the dijet mass after these corrections leads to a value of about 65 GeV/c^2 for the average Z mass (see for instance Fig. 5.10 on page 84).

The dijet mass resolution one expects for the Z decays in our sample with the standard jet corrections is about 14 GeV/c^2: this is due to the combined effect of the calorimeter resolution (which, as we have seen in Sec. 2.6.1, warrants no more than a 10% resolution on jet energies under normal conditions) and the presence of a semileptonic decay in one of the two jets, which smears massively the measured jet energy due to the flat response to the muon P_T and the total absence of a response to the neutrino P_T. A poor resolution limits our chance to see a well-defined and clear peak over the smooth background in the mass distributions; put in another way, the mass resolution affects the expected S/N ratio at the Z peak. We therefore set out to devise some improved jet corrections to be applied after the use of the standard corrections, in order to correctly take into account the characteristics of our events and improve mass resolution and peak S/N ratio. By doing that, we also intend, as a by-product, to shift the mass peak to the true Z mass value—a psychological and esthetic improvement.

We will correct jet momenta rather than jet energies, since the former are better modeled by the Monte Carlo, being less dependent on the fragmentation model[3]; the nominal b-quark mass will be used to switch to jet energies when these are required (*id est*, in the computation of the dijet mass). We have to note, though, that the difference for a 45 GeV jet is less than 1%.

When one looks at a typical jet momentum resolution plot, constructed by taking the difference between the b-quark momentum and the originated jet momentum (see for instance Fig. 5.8), one sees a distinctly non-Gaussian behavior in the tails of the distribution. We have no chance of correcting the jets in the tails enough to bring them back in the bulge of the distribution, thus "saving" them from getting lost in the smooth background mass spectrum: these events do not really count for us, and

[3]Jet masses, that are computed from the spread of the energy flux in the calorimeter around the reconstructed jet axis, are in fact poorly reproduced both by PYTHIA and HERWIG, the former being marginally better than the latter. The reason of that is the inadequate modeling of the dynamics of parton fragmentation transversely to the parton direction.

their contribution to any of the distributions we are going to study for the purpose of jet momentum corrections can be neglected: what we care for is to get a mass peak as high as possible, not the lowest possible r.m.s. for the total distribution.

We have studied many jet observables to find the most correlated ones to the jet momentum mismeasurement: the higher the correlation coefficient, the more information is extractable from them to correct the jet momenta. The variables that proved most useful and less model-dependent are the muon momentum (of course), the missing transverse energy—once a couple of tricks are performed—and the jet charged fraction. But, before we start with the description of the real momentum corrections, we must discuss a "zeroth-order" correction that we may apply to our events.

5.3.2 Correcting the Jet Direction

The muon coming from the semileptonic decay of a B hadron carries almost as much information on the parent momentum vector as the hadronic jet does, due to the selection bias in our dataset[4]. We have decided to use the momentum vector of the muon to correct the jet direction in order to obtain the best estimate of its parent b-quark direction, before turning to our main goal of improving the jet momentum resolution. The correction we obtain will prove of little impact on the dijet mass resolution for our events—which have been selected to be mostly back-to-back—since

$$dM^2/d\chi = 2E_1E_2\sin\chi \tag{5.1}$$

(E_1 and E_2 being the jet energies, and χ their angular separation), which vanishes for $\chi \to \pi$; but it certainly does no harm. We can obtain for each muon jet the muon vector $\vec{P}_\mu$, the vector of the energy deposited in the tower hit by the muon $\vec{E}_{tower}$, and the momentum vectors of the b quark and the associated jet. As far as the direction of the b quark is concerned, the first order estimate of the b-quark vector is given by the measured jet:

$$\vec{P}_b/|\vec{P}_b| = \vec{P}_{jet}/|\vec{P}_{jet}|, \tag{5.2}$$

while our improved estimate of the b-quark direction can be written as follows:

$$\vec{P}_b/|\vec{P}_b| = \frac{\vec{P}_{jet} + \alpha\vec{P}_\mu + \beta\vec{E}_{tower}}{|\vec{P}_{jet} + \alpha\vec{P}_\mu + \beta\vec{E}_{tower}|}. \tag{5.3}$$

In these expressions we have called $\vec{P}_b$ the reconstructed b-quark momentum, and $\vec{P}_{jet}$ the vector measured in the calorimeter. We find the two parameters α and β such that the average $\eta\phi$ distance between the true b-quark vector and the one written above is minimized, by scanning each parameter in search of a minimum separately; this is safe since the two observables can be shown to be uncorrelated (see Fig. 5.1). Not unexpectedly, we find that the energy deposited in the tower hit

[4]The muon momentum is in fact comparable, in our events, to that of the hadronic system the b quark has decayed to.

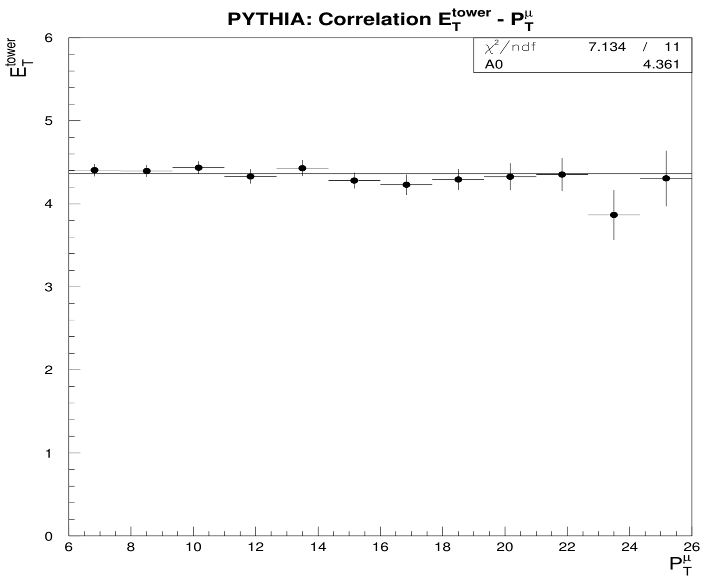

Figure 5.1: *The average energy in the tower hit by the muon track does not depend on the muon momentum, as expected.*

by the muon track is not sensitive to the b-quark direction[5]: the average distance is in fact minimized for a value $\beta = 0$. The best value of α is instead 1.0, showing that the muon carries indeed useful information for the b-quark direction[6] (see Fig. 5.2). We therefore correct the muon jets η and ϕ coordinates according to Eq. 5.3, *id est* by computing the momentum-weighted mean of the jet and muon vectors in each of the x, y, z coordinates, and using the direction of this as our best estimate of the b-quark initial direction.

5.3.3 The Jet Momentum Corrections

The method we use to correct the jet momenta is based on assuming a linear relationship between a given jet observable, X, and the mismeasurement $\Delta P = P_b - P_{jet}$ of that jet. To extract this dependence, we plot ΔP as a function of the value of the observable X, and fit a straight line. Once the intercept and slope of the line are known, jets falling outside a one-r.m.s. band on each side of the best fit line are discarded, and the remaining jets are fitted again. We thus obtain a dependence of

[5]This might actually look trivial: since the muon deposits in the calorimeter an amount of energy totally independent on its momentum (see Fig. 5.1), how can the energy in the traversed tower add information about the b direction ? As a matter of fact, it should. The distribution of the energy flow in the jet produced by a b quark that decayed semileptonically is certainly sensitive to the direction of the muon, and therefore the tower hit by the muon must retain some of that information (not by the amount of energy deposited, but by its location); however, it does not appear possible to extract it.

[6]The neutrino in the semileptonic decays we have selected is softer than the muon, and does not appear to degrade significantly the information that the muon direction provides. We trust our Monte Carlo for this result, since the b decay dynamics is very well modeled by QQ V9_0.

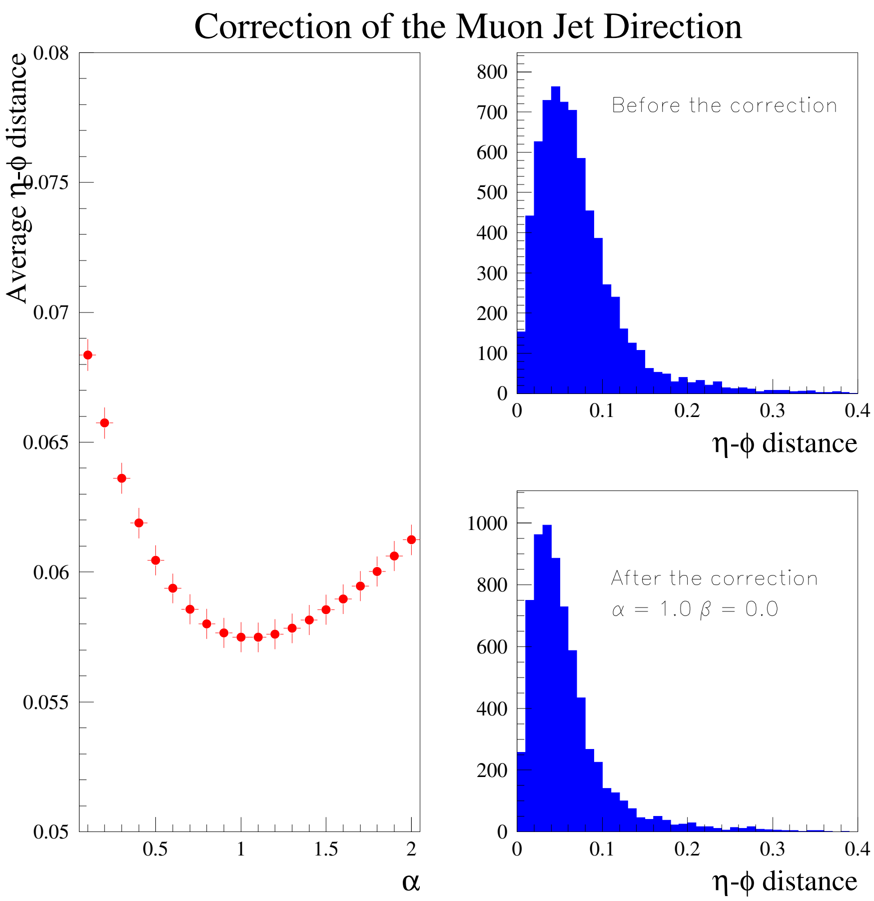

Figure 5.2: *These plots illustrate the minimization of the difference between reconstructed and true b-quark direction. The alpha parameter has to be chosen* $\alpha = 1.0$ *to get the best estimate of the b direction.*

jet momentum errors on the variable X,

$$\Delta P = mX + q, \tag{5.4}$$

that can be used to correct the measured momentum. The correlation coefficient of the line fit is an estimator of how dependent on X the mismeasurements actually are. A poor correlation coefficient means that no fruitful correction can be carried out with the use of X, while a correlation coefficient greater than 0.4, say, indicates that there is a true dependence that needs to be taken into account.

In what follows we will be dividing our data into disjoint classes on the basis of jet-jet correlations prior to extracting dependencies on the considered observables. We believe in fact that a jet momentum measurement is best used in conjunction with the measure of the opposite jet. In principle, if we were studying an object produced with no transverse momentum in the laboratory frame, and had reason to believe that the momentum of one of the two jets produced in its decay were measured with poor resolution (for instance if we observed it to be pointing at a calorimeter crack, or to have a charged fraction exceeding unity, or if we knew that it contained a neutrino by tagging a soft lepton inside the jet), we could just use its observed direction and rely on the momentum measurement of the other jet to reconstruct the true mass. The transverse momentum of Z production, a feature peculiar to hadron colliders, prevents us from being tempted by such unorthodox techniques[7]; anyway, we are still allowed to discriminate between well-measured jets and very badly measured ones, and to use that information somehow.

5.3.4 Correcting for the Muon Momentum

The first observable to deal with is the muon momentum. We found that the best correction to the muon jet momentum was obtained by first dividing the data in two disjoint classes:

1. events where the muon jet was second in rank of QDJSCO-corrected P_T;

2. events where the muon jet was the highest in QDJSCO-corrected P_T.

The partition allows us to profit from the anti-correlation of neutrino momentum and muon momentum, as will be made clear in a moment.

The total momentum difference between the parent b quark and the originated jet was plotted as a function of the muon momentum P_μ to get the two scatterplots shown in Fig. 5.3. A line fit was performed as explained on page 72.

Class	m	q	r	% events
1	0.793 ± 0.008	6.77 ± 0.11	0.68	86.0%
2	0.948 ± 0.028	-1.84 ± 0.31	0.58	14.0%

Table 5.2: *Parameters of the fits to the ΔP (b, jet) $\div P_\mu$ scatterplots; m is the slope of the fit lines, q the intercept ("offset"), and r is the correlation coefficient.*

The correlation coefficients given in Table 5.2 (0.68 and 0.58 for class 1 and 2, respectively) show that the muon momentum carries indeed useful information for the correction of the muon jets. The fits in the two samples give significantly different values for the slope of the lines ($m = 0.79, 0.95$) and for the offsets ($q = 6.8, -1.8$), underlining once again the correctness of our decision to treat the two classes separately. In fact, the muon momentum is anti-correlated with the neutrino momentum, and the negative correlation is much more evident in class 1 (see Fig. 5.4), where neutrinos are harder on average, reflecting the very way in which the two classes are defined: therefore in the first category the slope in the $\Delta P \div P_\mu$ scatterplots is significantly less than unity. The different offsets are also a result of the definition of the two classes: again, they reflect the different average neutrino momentum.

[7]It is the transverse boost of the boson that in fact prevents the use of such a method in the mass reconstruction of leptonic W boson decays at the Tevatron: the transverse lepton-neutrino mass is independent, to first order, to the W boson P_T, while the lepton P_T is not.

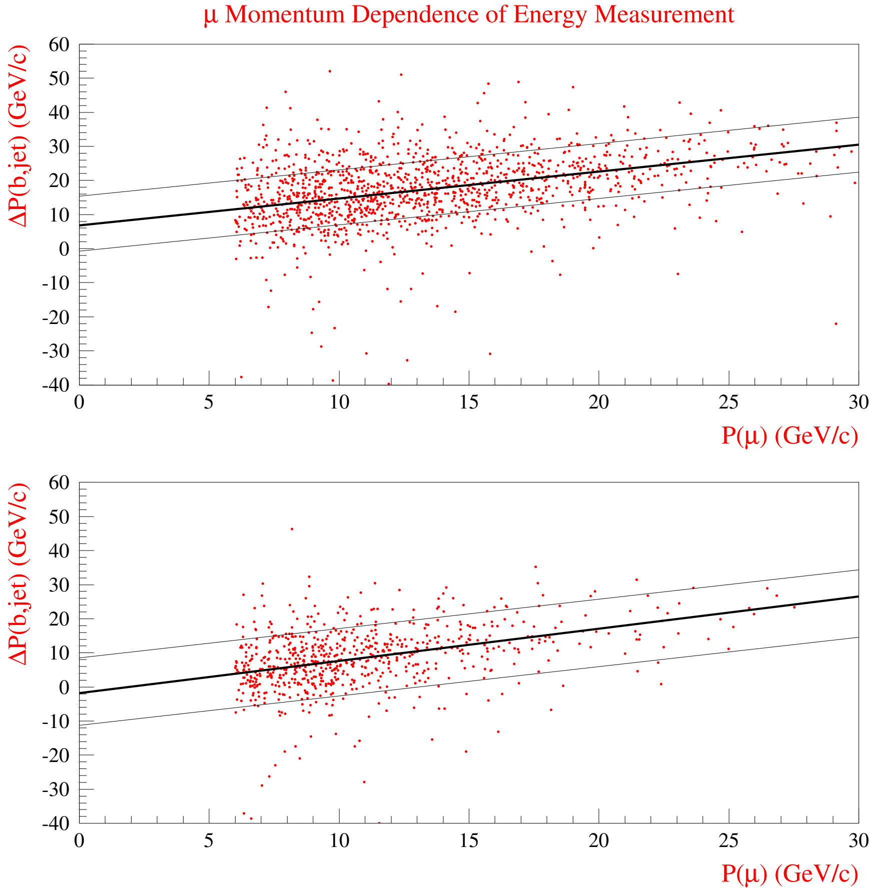

Figure 5.3: *Correction for the muon momentum in the two classes: events where the muon jet has P_T lower than the away jet (top) and events where the muon jet has higher P_T (bottom). Jets falling inside the one-r.m.s. band are used to get the fit shown by the thick line.*

The parameters extracted from the fits are used to correct the muon jet momenta as follows:

$$P_{new} = P_{old} + mP_{\mu}. \tag{5.5}$$

That is, we correct jet momenta for the muon momenta by factoring in only the slopes m. The offsets q (connected, as we have just said, to the neutrino momentum) are not used right away, since in what follows we want to treat muon and away jets uniformly: this will prove fruitful when we deal with the next observable, the missing transverse energy. Correcting the muon jets for the offsets at this stage would later result in a less clear separation of different contributions to the missing E_T.

The correction of jet momentum in the two classes does improve the mass resolution, as shown in Fig. 5.5. The invariant mass after the correction improves its relative resolution by 23.5%, the estimator σ_M/M being 0.166 (0.205) after (before) the correction.

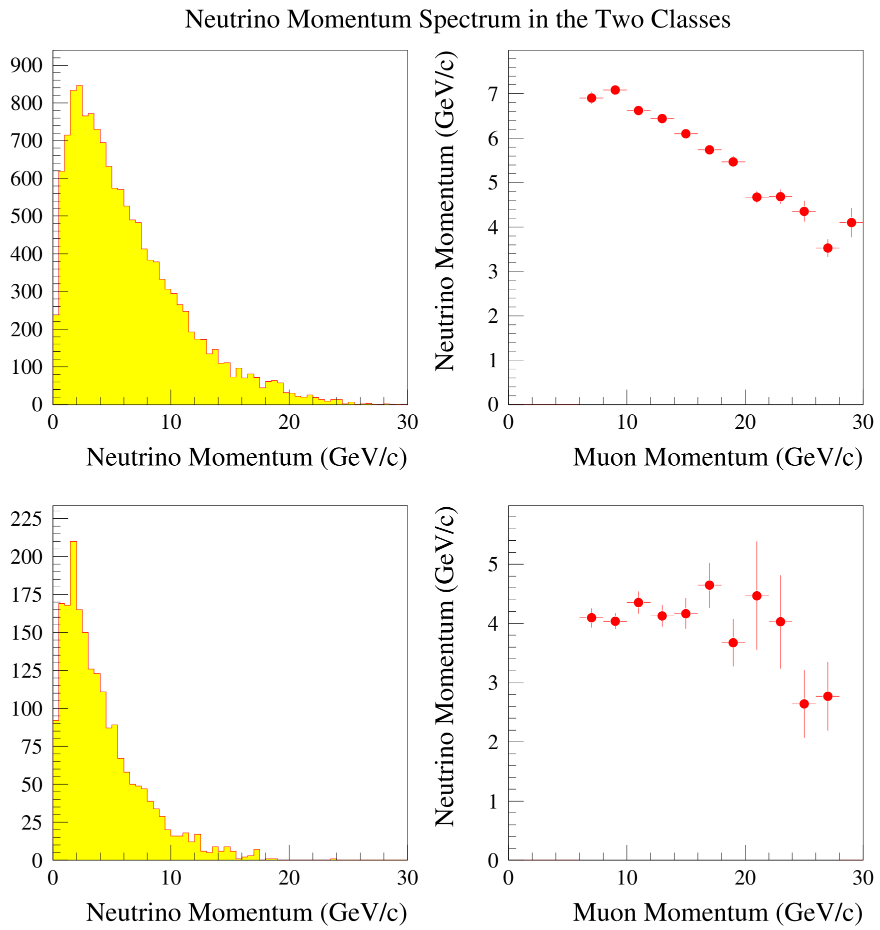

Figure 5.4: *These plots show that class 1 events (top) have the neutrino momentum anticorrelated with the muon momentum, while class 2 events (bottom) have the neutrino relatively insensitive on the muon momentum, justifying the division in classes. On the left are shown the neutrino momentum distributions, on the right the correlations with muon momentum.*

5.3.5 Correcting for the Missing Transverse Energy

Using the missing transverse energy for a jet momentum correction may appear at first a bit extravagant, if not nonsensical, to anyone familiar with hadron collider physics: the $\not\!\!E_T$ has lots of different possible sources, and its resolution is really poor for an event containing hadronic jets. Nevertheless, we will prove here that not only the $\not\!\!E_T$ can be dealt with, but also that it may actually turn out to be useful; moreover, its study reveals quite a bit of details on the characteristics of our data.

At CDF there are various different definitions of the missing transverse energy, reflecting its use in different contexts and in different datasets. A standard offline routine computes a missing E_T corrected for jet mismeasurements[8] and for muons

[8]The raw $\not\!\!E_T$, obtained from the vector sum of all calorimeter towers, is modified by using the

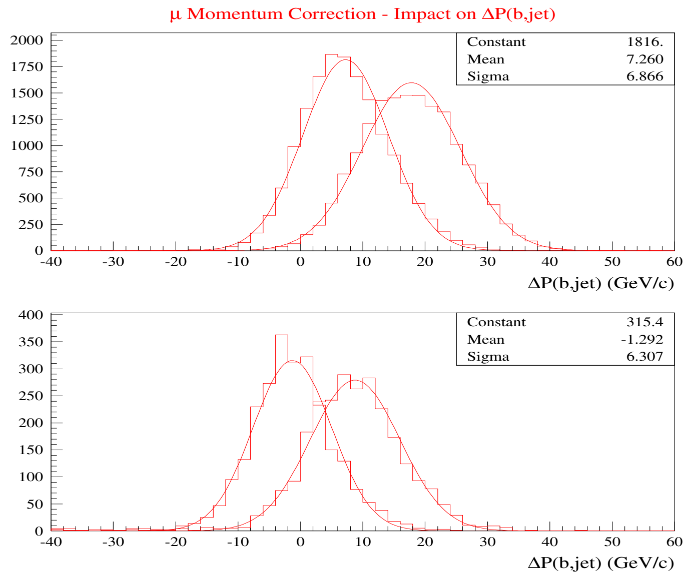

Figure 5.5: *The momentum resolution for the two classes of muon jets improves after the muon correction. Top: class 1 muon jets; bottom: class 2 muon jets. In both cases the Gaussians narrow thanks to the correction, and their means get closer to zero. The mean value of the Gaussian fits to the corrected distributions (*7.26 *and* -1.29 *GeV/c) shows a small mismatch from the value of the intercepts of the fit lines in Fig. 5.3: the difference is due to the different fit used here for display purposes.*

with transverse momentum higher than 10 GeV/c. We further correct it to take into account the presence of muons with P_T lower than that value in many of our events, by vectorially subtracting the muon transverse momentum decreased by 2 GeV/c (their typical E_T deposit in the calorimeters).

The missing E_T is an event variable, and we have to convert it to a jet variable in order to use it. We compute its projection along each jet direction, and define the jet to be corrected as the one yielding the larger projection. At this point, the value of the $\not\!\!E_T$ projection may be the result of many different sources, the dominant ones being:

- the momentum of a neutrino produced by a secondary decay in the jet;
- a mismeasured jet, *id est* a jet whose energy deposition in the calorimeter poorly matches the parton momentum, due to detector effects such as nonlinearity or energy lost in uninstrumented regions—this still happens even after the jet has been corrected (see Sec. 2.6.1);
- a non-MIP muon, that is, a muon that left an abnormally large deposit in the calorimeter (either because it was a fake muon, a punch-through muon, or a real muon in the tail of the Landau distribution);

vector differences between corrected and uncorrected jet energies.

- a badly measured hadronic recoil of the Z system: when the initial state radiation in the Z production is soft and spread over a wide region of the calorimeter, it can fail to be clustered and its contribution to the $\not{E}_T$ computation gets underestimated.

We want to disentangle as well as possible the first two listed sources. The best way to accomplish that is to divide the dataset into four classes:

1. events where the muon jet has the larger $\not{E}_T$ projection but its P_T (after the correction for the muon momentum) is the higher one;
2. events where the muon jet has the larger $\not{E}_T$ projection and its P_T (after the correction for the muon momentum) is the lower one;
3. events where the away jet has the larger $\not{E}_T$ projection and its P_T is the lower one;
4. events where the away jet has the larger $\not{E}_T$ projection but its P_T is the higher one.

Classes 1 and 4 contain events that have typically a small $\not{E}_T$ projection, whose source is not well defined and univocal; in class 2 the $\not{E}_T$ projection is predominantly due to the neutrino accompanying the muon, while in class 3 it is most likely caused by an undermeasurement of the away jet. We can demonstrate the truth of these statements by studying the correlation of missing E_T projection and neutrino P_T in classes 2 and 3: the correlation is positive in the former, while there is none in the latter (see Fig. 5.6 and Fig. 5.7). On the other hand, the mismeasurement of the jets is correlated to the missing transverse energy in classes 2 and 3, while it shows no correlation for classes 1 and 4. These observations allow us to confidently correct the muon jets in class 2 events, and the away jets in class 3 events; we will instead leave alone, for the purpose of a $\not{E}_T$ correction, the jets in events from classes 1 and 4. The parameters of the fits are shown in Table 5.3.

Class	Jet to correct	m	q	r	% events
2	μ jet	0.370 ± 0.009	6.63 ± 0.09	0.44	46.5%
3	away jet	1.034 ± 0.021	4.53 ± 0.20	0.57	33.6%

Table 5.3: *Parameters of the fit to the ΔP (b, jet) $\div$ $\not{E}_T{}^{proj}$ scatterplots for the two classes where the missing E_T is used; m is the slope of the line fits, q is the offset, and r is the correlation coefficient.*

The correlation coefficients of the fits in classes 2 and 3 are close to 0.5, showing that indeed there is good information to extract from the $\not{E}_T$. In classes 1 and 4, on the other hand, there is no correlation, the r factors being less than 0.1. The slopes are used to further correct the relevant jets in class 2 and class 3 events; again, we do not correct for the offsets: we will re-evaluate and use them at the last correction step. Figure 5.8 shows the fits and the improvement in the jet momentum resolution after these corrections.

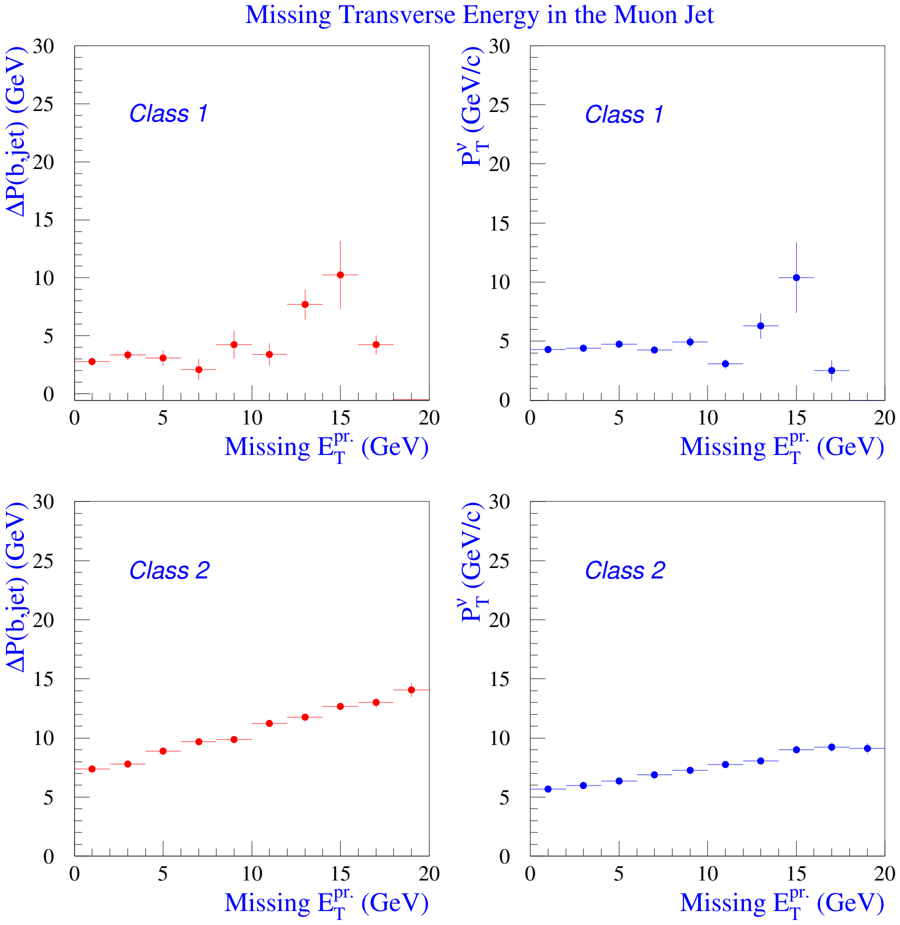

Figure 5.6: *These plots show the correlations of the projected missing transverse energy with muon jet momentum mismeasurement (left) and neutrino P_T (right). Class 1 muon jets are shown on the top plots; class 2 muon jets are on the bottom plots. One clearly sees that class 2 muon jets are to be corrected for the $\not\!\!E_T$ since the projected missing E_T has some sensitivity to the neutrino, while in class 1 there is no evident correlation between $\not\!\!E_T$ and P_T^ν.*

It is worth noting that the slope is exactly 1.0 for away jets of class 3: the missing E_T, although with very poor resolution, is a real measure of the away jet mismeasurement in that class. On the other hand, in class 2 muon jets have already been corrected for the muon P_T dependence; thus, due to the anti-correlation of P_T^μ and P_T^ν, the neutrino has been already partly corrected for (and in fact the slope we used in the P_T^μ correction was sensibly different from unity—0.793 ± 0.008, see Table 5.2—in class 2). What the $\not\!\!E_T$ can still do is contribute to the correction, with a slope lower than unity this time: this is not surprising, once we remember that the $\not\!\!E_T$ we are using is corrected for the muon P_T but does not take into account our previous muon correction to the jet P_T, having been computed with the help of jet momenta corrected only by QDJSCO.

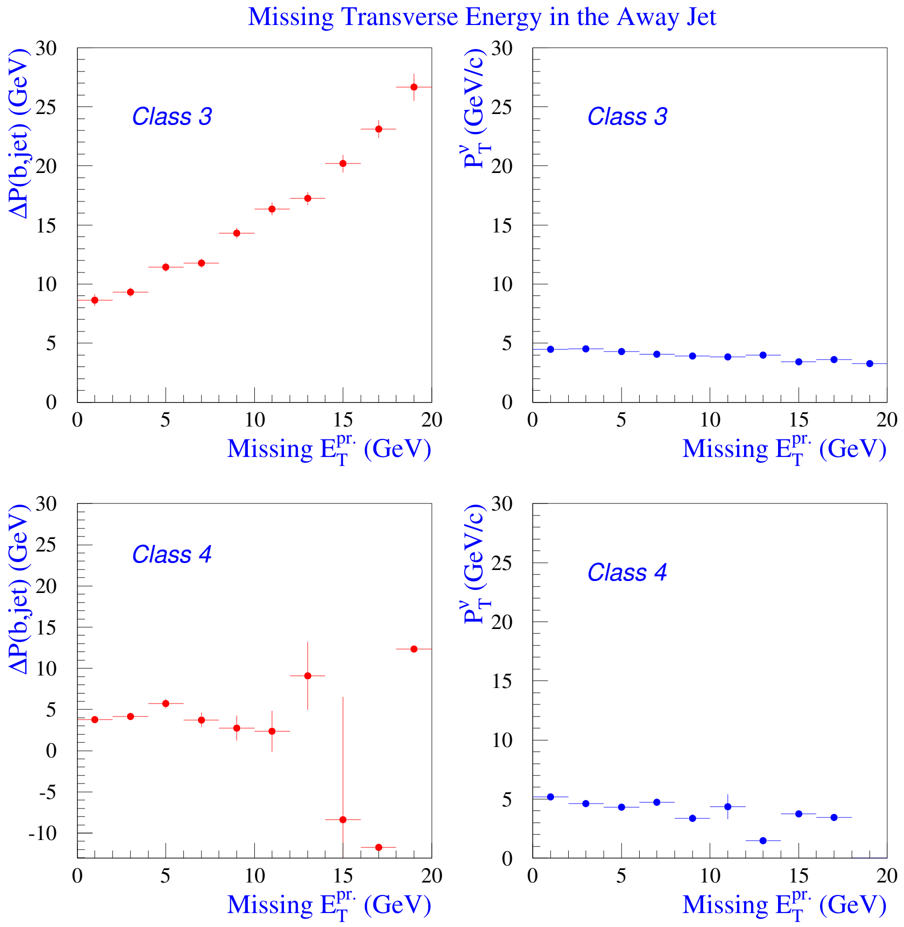

Figure 5.7: *These plots show the correlations of the projected missing transverse energy with away jet momentum mismeasurement (left) and neutrino P_T (right). Class 3 away jets are shown on the top plots; class 4 away jets are on the bottom plots. One can see how the P_T measurement of class 3 away jets is probed by the measurement of the projected missing E_T, while for class 4 away jets there is no effect to correct for. In both cases, there is no dependence on the neutrino P_T, as expected.*

The global effect of the corrections for the missing transverse energy alone is small: the absolute resolution in the invariant mass does not improve, while the estimator σ_M/M gets to 0.158, improving by 6%. Anyways, we consider this as a positive result: in fact, the correction of class 2 muon jets and class 3 away jets for the missing E_T shows a part of its usefulness only *a posteriori*, since it allows a more performant correction for the charged fraction for those jets.

Having corrected for both the muon and the neutrino momenta, we can now turn our attention to the last observable we found useful, the charged fraction.

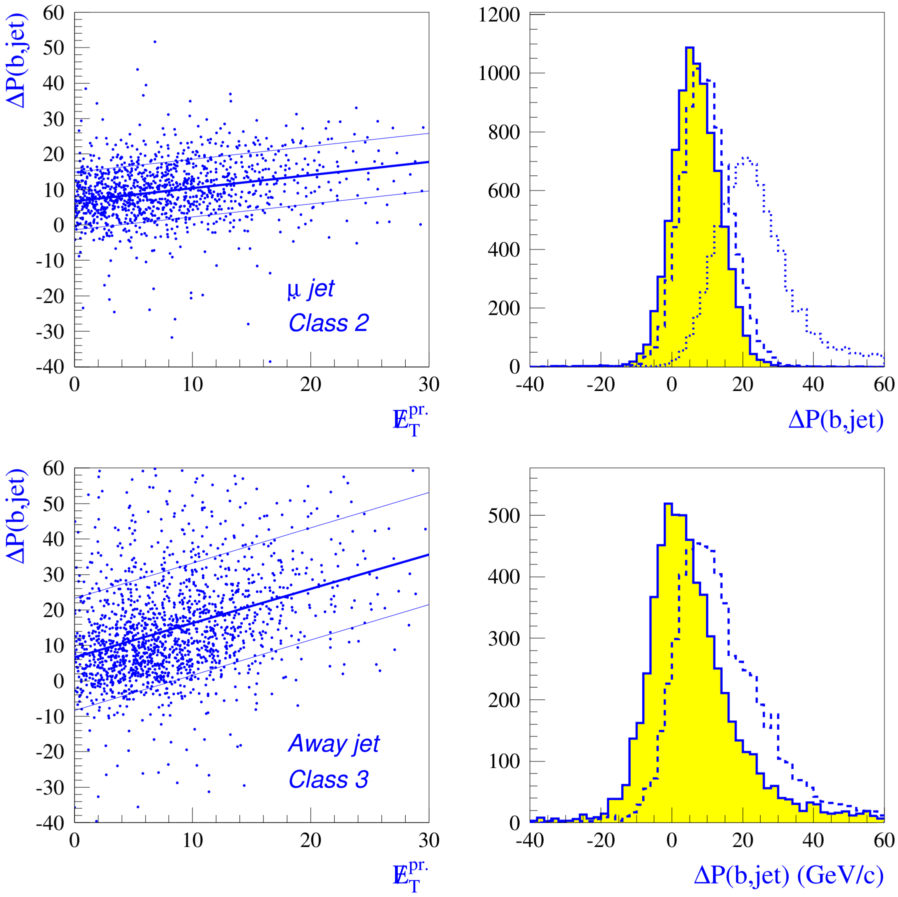

Figure 5.8: *These plots show the fits to class 2 muon jets (top) and class 3 away jets (bottom). On the right are shown the improvements in the momentum resolution: the dashed histograms are the distributions before the corrections; the top plot also shows the distribution before the muon momentum correction of muon jets belonging to class 2 (dotted histogram).*

5.3.6 The Charged Fraction Correction

The resolution in P_T for the charged tracks in the CTC acceptance is much higher than the resolution of the energy deposit they leave in the calorimeter; however, the fluctuations in the momentum fraction carried by neutral particles in the jets and the relatively small pseudorapidity acceptance of the central tracking prevent a direct use of the tracks P_T to infer the jet momentum. Nevertheless, some good information on the parton momentum is extractable from the sum of P_T of tracks included in the jet cone, when the jet falls inside the pseudorapidity region where the CTC is fully efficient. We therefore studied the jet charged fraction, modifying its definition from that of fashion at CDF. We define a corrected charged fraction as

$$CHFR_{corr} = CHFR_{JETS} \times P_T^{raw} / P_T^{corr}, \tag{5.6}$$

where CHFR$_{JETS}$ is the definition of the jet charged fraction used at CDF in most instances; it is computed as the ratio between the sum of P_T of all tracks included in the jet cone and raw (before the application of QDJSCO corrections) jet P_T, while P_T^{raw} and P_T^{corr} are respectively the raw jet transverse momentum and the P_T resulting after the application of all the corrections described up to this point. Thus defined, the charged fraction is decoupled from the effects dealt with precedently, and shows a stronger correlation to the ΔP of the jets.

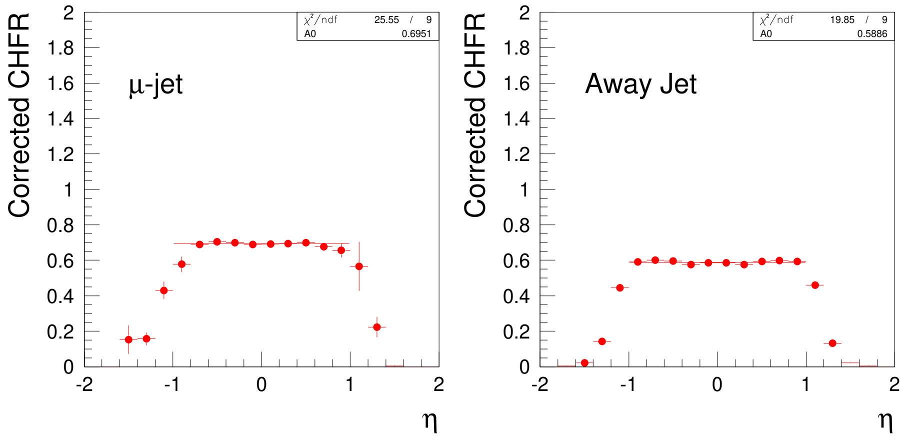

Figure 5.9: *Average charged fraction as a function of jet pseudorapidity, for the muon jets (left) and the away jets (right).*

We observe that the average charged fraction is constant with jet pseudorapidity for $|\eta_{jet}| < 1.0$ (see Fig. 5.9), falling steeply outside that interval, reflecting the limited acceptance of the central tracking. Accordingly, we correct for the charged fraction only jets in the central region ($|\eta_{jet}| < 1.0$): for jets outside this interval the charged fraction is close to zero and is very poorly measured.

We must keep separated the four event classes defined in the former section, because we wish to avoid treating the same way jets already corrected for the missing transverse energy and jets not corrected: the missing E_T measurement is in fact linked to the observed value of jet charged fraction, the two being both affected by the jet mismeasurement in the calorimeter. In each event class, muon jets and away jets also have to be treated separately, because their charged fraction is intrinsically different, since muon jets always have a high P_T track (compare again the two plots in Fig. 5.9). We end up with 8 different classes of jets, in each of whose we perform the same fits already described in the previous sections, using our redefined CHFR as the observable to correct for. All the fits show good correlation coefficients (see Table 5.4), which demonstrates the usefulness of the correction; moreover, each fit has sensibly different parameters, which also confirms—were it necessary—that treating the different classes separately is the correct procedure to undertake.

It is to be noted that a correction for the jet charged fraction is, in principle, energy-dependent: the slope of the fit line must depend on the typical momentum of the considered jets. This fact, together with the very peculiar nature of the events used to derive the corrections, limits in principle the usefulness of the corrections to our searched process, the $Z \to b\bar{b}$ decay. To be more general, we should have chosen to correct the *relative* error in the jet momentum with the jets charged fraction, or alternatively to use the bare sum of tracks P_T in the jet cones as the observable to correct for; anyways, one can notice from the comparisons shown in Sec. 5.5.1 that the impact of the charged fraction correction is quite similar even on the higher P_T jets from a 120 GeV/c^2 Higgs boson decay.

Since the CHFR is the last observable we are using, we correct the jet momenta also for the offsets we find in these fits: $P' = P + mX + q$. Those jets falling outside of the CTC acceptance are all "away jets" (our selected muons have $|\eta| < 0.6$, and therefore the jets containing them are already restricted in rapidity), and they have to be corrected for some offset as well, so we divide them into the four separate classes dictated by the missing E_T projection, and fit their ΔP distributions with Gaussian curves, extracting mean offsets $\overline{\Delta P}$ that we apply to the jet momentum in each class: $P' = P + \overline{\Delta P}$. The parameters are shown in Table 5.5.

Jet to correct	Class	m	q	r	% events
μ jet	(1)	15.50 ± 0.85	-7.87 ± 0.62	0.53	5.4%
(99.6%)	(2)	8.37 ± 0.23	0.93 ± 0.16	0.47	46.4%
	(3)	13.77 ± 0.16	-7.75 ± 0.11	0.89	33.6%
	(4)	12.70 ± 0.31	-3.63 ± 0.25	0.69	14.2%
away jet	(1)	17.03 ± 0.74	-3.39 ± 0.59	0.63	4.8%
(69.8%)	(2)	13.61 ± 0.21	-9.04 ± 0.12	0.74	36.0%
	(3)	13.62 ± 0.41	-5.47 ± 0.25	0.54	21.4%
	(4)	11.74 ± 0.59	-4.15 ± 0.38	0.51	7.6%

Table 5.4: *Parameters of the line fits to the ΔP (b, jet) $\div CHFR^{corr}$ scatterplots; m is the slope of the fit lines, q is the offset and r is the correlation coefficient.*

Jet to correct	Class	ΔP (b, jet)	% events
away jet	(1)	2.32 ± 0.77	0.9%
(30.2%)	(2)	2.94 ± 0.28	10.5%
	(3)	3.07 ± 0.37	12.2%
	(4)	2.89 ± 0.63	6.6%

Table 5.5: *Offsets evaluated as means of the Gaussian fits to the ΔP (b, jet) distributions for jets falling outside the CTC fiducial region, $|\eta_{jet}| > 1$.*

This last correction takes the ratio σ_M/M to 0.136, with a relative improvement of 16% with respect to the former corrections.

5.4 Results on the Mass Distribution

The result of the jet corrections we have discussed above is best presented as a plot of the mass of PYTHIA Z events after each step of correction. The mass distributions are shown in Fig. 5.10; the resolution improves from 13.81 to 12.27 GeV/c^2, but the measure of the effectiveness of our corrections is given by the relative resolution σ_M/M, that has improved by 50%, decreasing from 0.205 to 0.136. Table 5.6 also gives the numerical values of interest for the mass resolution improvements.

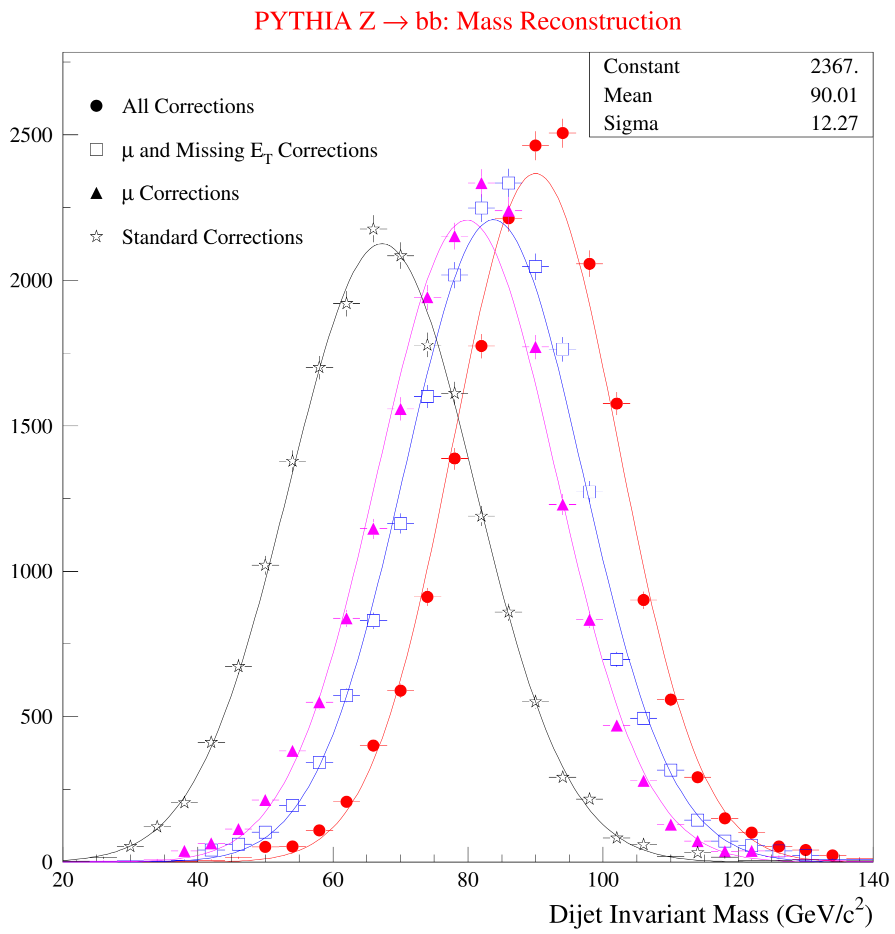

Figure 5.10: *Mass distributions of PYTHIA Z events at the various stages of correction.*

5.5 Checks and Tests

5.5.1 Checks with HERWIG

The results we have obtained could in principle depend on the model used by the Monte Carlo generator for the parton fragmentation. For instance, if the fragmen-

Correction	$M(GeV/c^2)$	$\sigma_{jj}\ (GeV/c^2)$	σ_{jj}/M_{jj}	$\Delta(\sigma_{jj}/M_{jj})\%$
Standard	67.25 ± 0.10	13.81 ± 0.07	0.205	
P_μ	79.81 ± 0.10	13.22 ± 0.08	0.166	-23.5%
$P_\mu + \not{E}_T$	83.75 ± 0.10	13.22 ± 0.08	0.158	-29.7%
Final	90.01 ± 0.09	12.27 ± 0.08	0.136	-50.7%

Table 5.6: *Improvement of the parameters of Gaussian fits to the invariant mass distribution in the $Z \to b\bar{b}$ dataset generated with PYTHIA, from the standard QDJSCO corrections to our final result.*

tation model gave always higher P_T tracks than in the real b-quark fragmentation, we could end up with a wrong determination of the slope in the fits to the charged fraction. To verify that the corrections are relatively independent on the model, we use the $Z \to b\bar{b}$ dataset generated with HERWIG V5_6. We apply to these events our jet corrections—using the fit parameters obtained with the PYTHIA Z events by means of the procedure described above—exactly as we would do on the real data.

Correction	$M(GeV/c^2)$	$\sigma_{jj}\ (GeV/c^2)$	σ_{jj}/M_{jj}	$\Delta(\sigma_{jj}/M_{jj})\%$
Standard	69.07 ± 0.15	12.84 ± 0.11	0.186	
P_μ	80.77 ± 0.14	12.42 ± 0.11	0.154	-20.8%
$P_\mu + \not{E}_T$	84.49 ± 0.14	12.58 ± 0.11	0.149	-24.8%
Final	90.74 ± 0.13	11.68 ± 0.10	0.129	-44.2%

Table 5.7: *Improvement of the parameters of Gaussian fits to the invariant mass distribution in the Z dataset generated with HERWIG, from the standard QDJSCO corrections to our final result.*

We find that there is no dependence of the results on the generator, as can be seen by a comparison of Fig. 5.10 with Fig. 5.11. The resolutions and the average mass values are compatible in the two Monte Carlo datasets, as can be checked in Table 5.7.

Maybe of greater interest is the fact that, applied to the direct Higgs boson production events we have generated with HERWIG ($M_H = 120\ GeV/c^2$), the corrections work in an astonishingly similar fashion, as can be seen in Table 5.8 and in Fig. 5.12. The Higgs boson is a scalar particle, so the angular distributions in its decay are in principle different from those of Z decay; moreover, the jet momenta are on average 30% higher, yet the corrections devised for the $Z \to b\bar{b}$ decay do not lose any of their effectiveness. They prove to be a bit more general than we would think.

5.5.2 The Test on the Real Data

We can finally correct the jet momenta in our data events using the parameters extracted from PYTHIA. The only difference in treatment is due to the fact that the

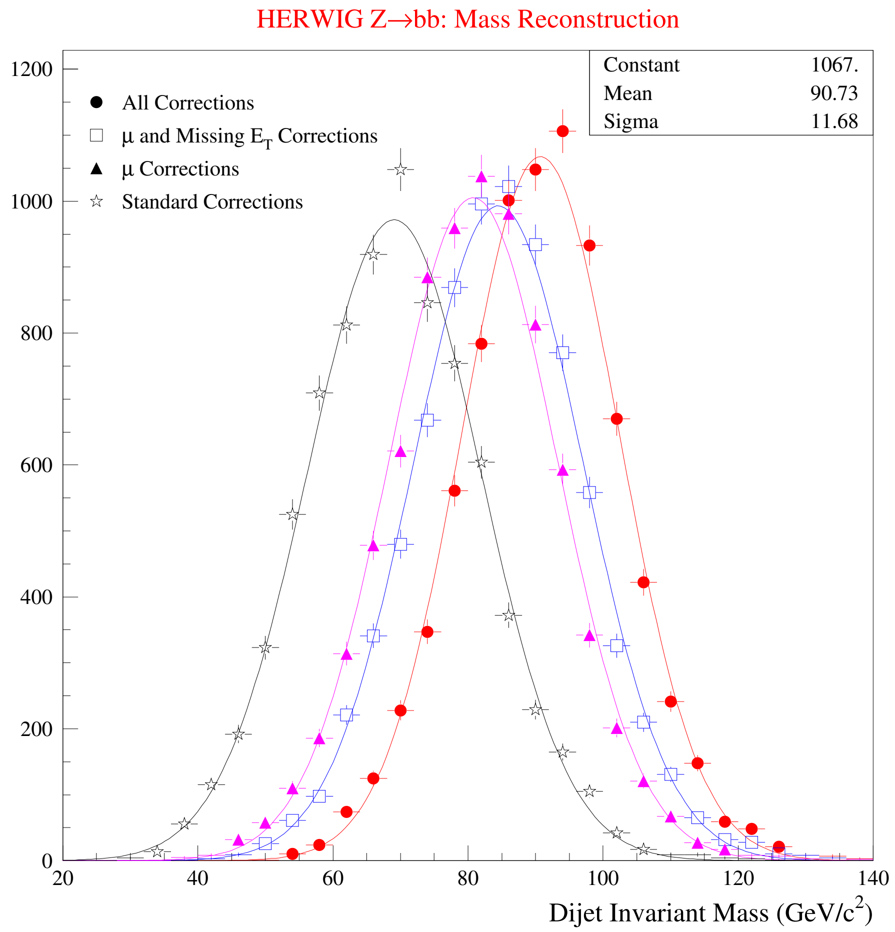

Figure 5.11: *Mass distributions at the various stages of the correction for $Z \to b\bar{b}$ decays generated with HERWIG.*

experimental data are selected by requiring that both jets contain a secondary vertex identified by the SECVTX algorithm. This may actually have an impact on our corrections, since these jets have necessarily at least two tracks with $P_T > 1.5\ GeV/c$. We believe this is a relatively small price to pay in order to be able to play with Monte Carlo datasets with one order of magnitude more events. Anyways, we can actually *see* from our data whether our corrections work as expected or not.

To obtain a peak and study its mass value and shape we perform a counting experiment, as will be explained in detail in Ch. 6; we anticipate here some of the results, in order to show the results of the corrections on the real data. To summarize the technique briefly, we obtain a double tagging rate in a signal-depleted sample (events discarded by the kinematic cuts) as a function of the invariant mass of the dijet system, and use it to compute an absolute prediction for the number of double tags in each bin of the mass spectrum; by choosing mass bins small enough we are allowed to see the shape of the excess of events. The results will be shown for the mass as computed at each step of the jet corrections described above. We expect to

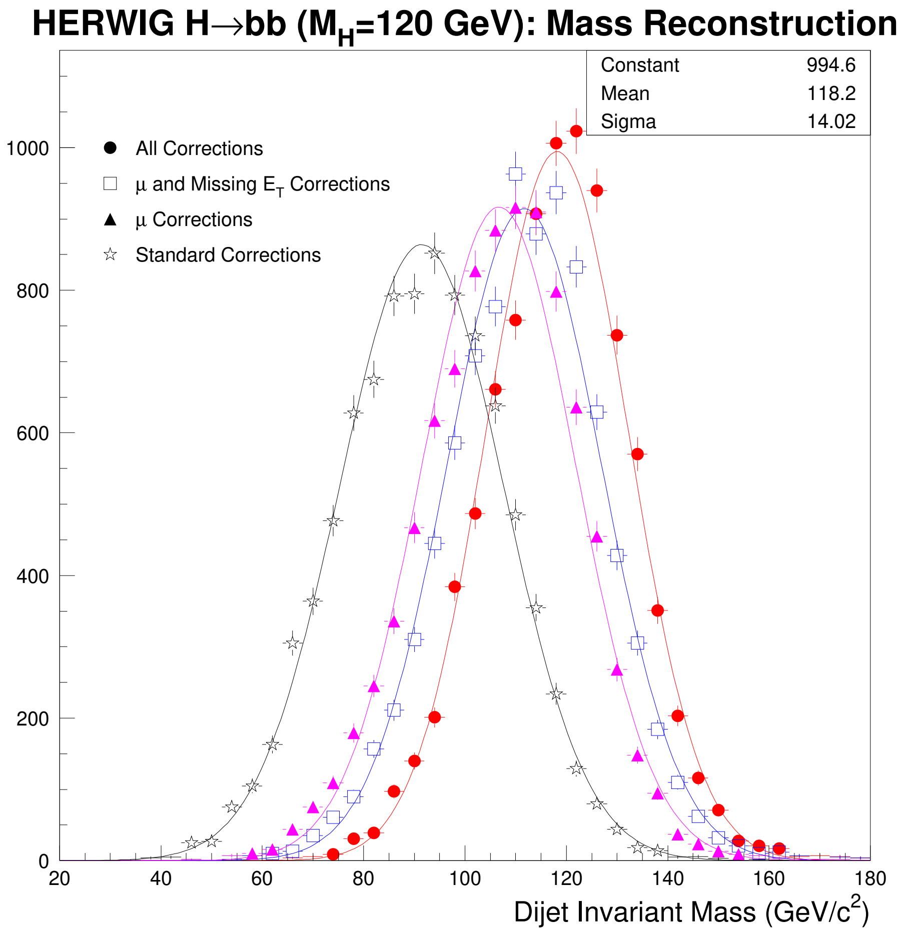

Figure 5.12: *Mass distributions at the four stages of correction for the $H \to b\bar{b}$ dataset generated with HERWIG ($M_H = 120\ GeV/c^2$).*

see a peak shifting gradually from 67 to 90 GeV/c^2, with a width improving slowly from 14 to 12 GeV/c^2.

The results of the counting experiment can be seen in Fig. 5.13 and Fig. 5.14. The results indeed show that the peak shifts to 88.5 GeV/c^2, more or less where the Monte Carlo predicts the Z to sit, and that the resolution appears to improve. We cannot tell with certainty whether the improvement in the mass resolution is real, since the error bars in the plots of the excess are too large. Nevertheless, we can try and fit them with Gaussian shapes (plus an offset, to take into account a possible bias in the prediction for the background—the bias turns out to be negligibly small[9]) and see whether there is compatibility or not with an improvement in the resolution. We find that it does appear to improve (see also Fig. 5.16) from $17.6 \pm 4.4\ GeV/c^2$ to $13.8 \pm 4.8\ GeV/c^2$: these values are larger than expected, but overall they are satisfactory.

[9]Checks with doubly taggable events show that the counting experiment has a systematic uncertainty of the order of 3% in the prediction. See Sec. 6.5 for the details.

Correction	$M(GeV/c^2)$	$\sigma_{jj}\ (GeV/c^2)$	σ_{jj}/M_{jj}	$\Delta(\sigma_{jj}/M_{jj})\%$
Standard	94.03 ± 0.18	16.43 ± 0.13	0.175	
P_μ	107.60 ± 0.17	15.10 ± 0.13	0.140	-24.5%
$P_\mu + \not{E}_T$	112.59 ± 0.17	15.14 ± 0.13	0.135	-29.9%
Final	118.74 ± 0.16	14.18 ± 0.12	0.119	-46.3%

Table 5.8: *Improvement of the parameters of Gaussian fits to the invariant mass distribution in the $H \to b\bar{b}$ dataset generated with HERWIG, from the standard QDJSCO corrections to our final result.*

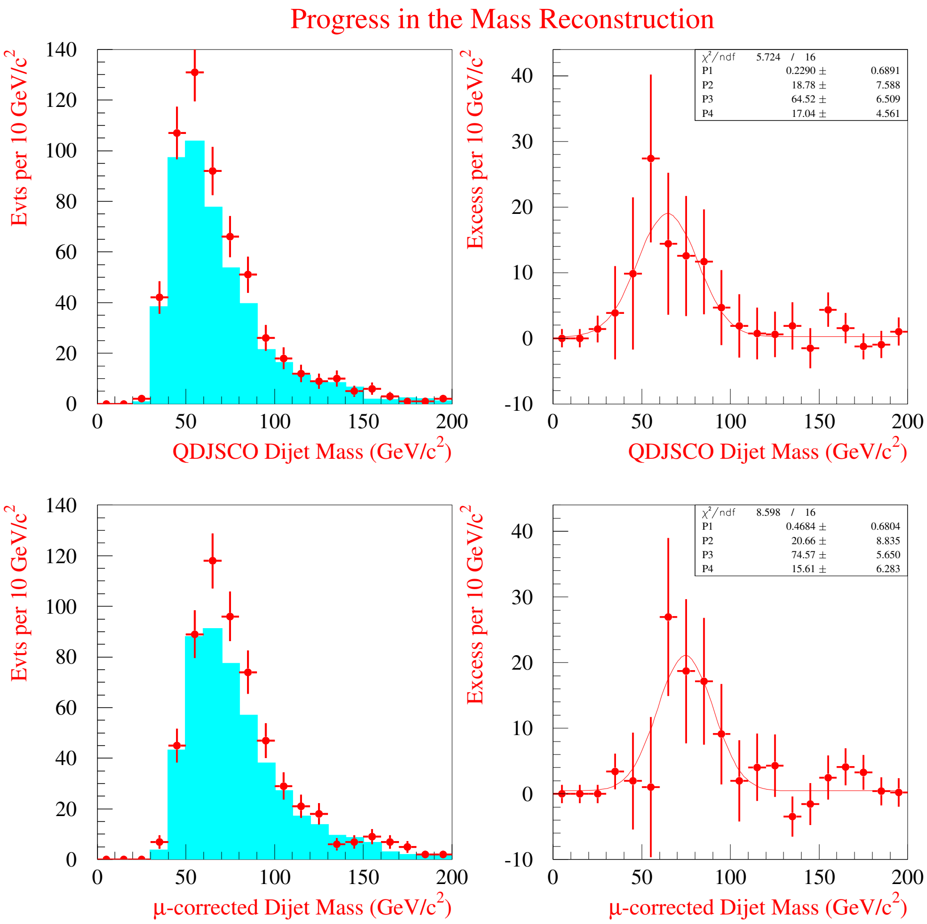

Figure 5.13: *Left: mass distribution for the double tagged data (red dots) and for the background prediction (full histogram); right: the excess of double tags is fit to a Gaussian plus an offset. The top plots are obtained with the standard jet corrections, the bottom ones are obtained with the jet energies corrected for the muon momentum.*

There are three final points to make before embarking in the topic of next chapter. The first concerns the background shape: as can be seen in the plots in Fig. 5.13 and Fig. 5.14, the predicted background peaks at 55 GeV/c^2 without corrections,

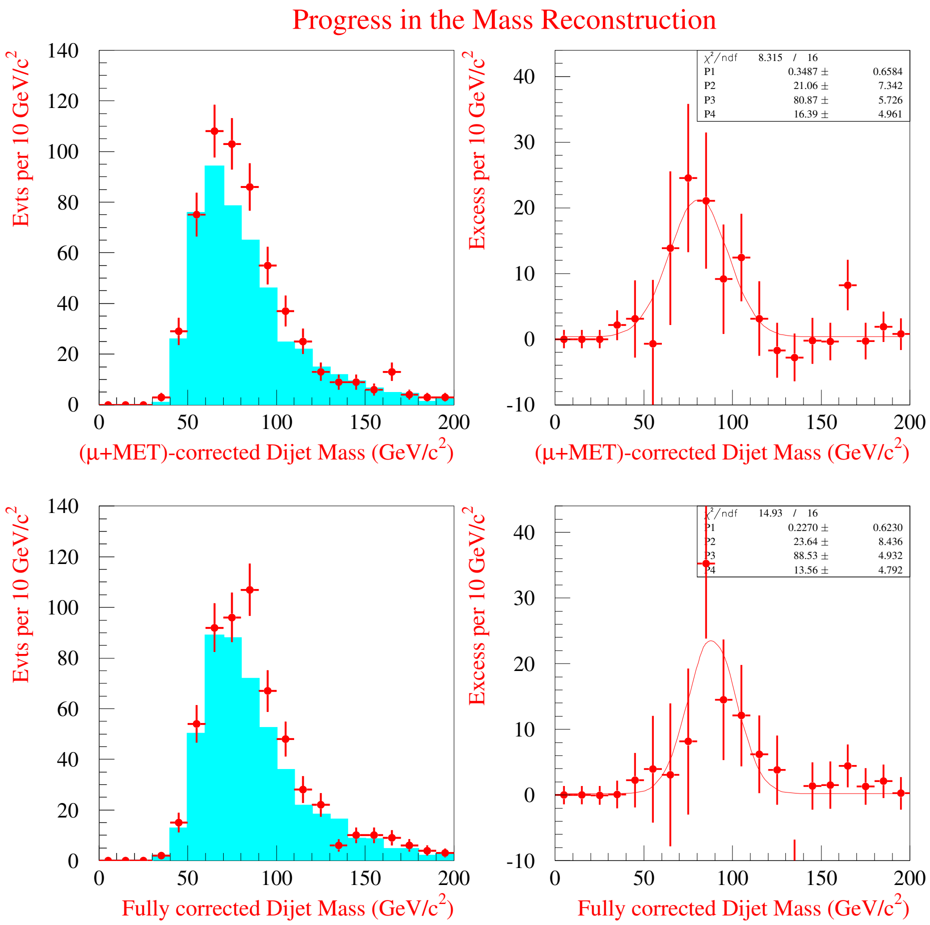

Figure 5.14: *Left: mass distribution for the double tagged data (red dots) and for the background prediction (full histogram); right: the excess of double tags is fit to a Gaussian plus an offset. The top plots are obtained with jet energies corrected for muon momentum and missing E_T, the bottom ones are obtained with the full jet corrections.*

and shifts to $65 \div 70\ GeV/c^2$ after the corrections. The shift is much less dramatic than the one of the excess (64.5 to 88.5 $GeV/c^2 = 24\ GeV/c^2$), underlining the fact that the double tagged data are really understandable as the sum of two different processes featuring a different behavior with respect to the variables used to correct the jet momenta. To understand what that means, we can plot the average shift in the mass due to each correction as a function of the mass before that correction, for data and Monte Carlo. As Fig. 5.15 shows, the behavior of background and signal is indeed strikingly different: the former gets a larger correction for larger masses, the latter is corrected more at smaller mass—as it should, since its true mass is always the same[10].

[10] In fact, a perfect correction would manifest itself in this plot as a $-45°$ line: all Z events would then be brought to 90 GeV/c^2.

The behavior of QCD events is due to their steeply falling cross section as a function of the mass (or, equivalently, the Q^2 of the scattering), that makes events firing the trigger more likely to come from a low mass region, even if there the probability of creating a $P_T > 8$ GeV/c muon in the B decay is smaller: for these events, the muon, the neutrino, and the calorimeter response conspire to give a small bias, and the "raw" mass is then closer to the true mass.

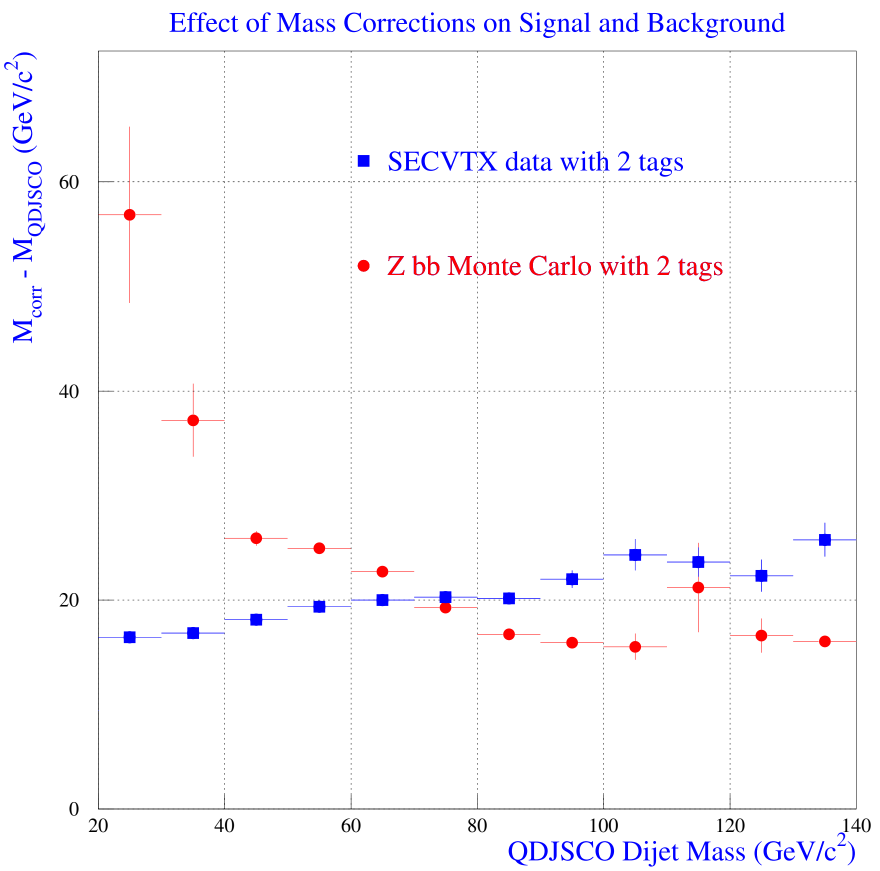

Figure 5.15: *This plot demonstrates that the QCD background and $Z \to b\bar{b}$ signal behave differently when their jet's energy is corrected according to the recipe described in the text. The increase in the value of mass is shown as a function of the mass value after the standard QDJSCO corrections.*

The second observation is that the observed S/N ratio does improve with the application of the jet corrections: Fig. 5.13 and Fig. 5.14 show that we have a Gaussian peak of 18.8 events on top of a background of about 90 events (in 10 GeV/c^2 bins) in the non-corrected spectrum (S/N $\sim$ 1/5), to be compared to a Gaussian peak of 23.6 events on top of about 70 expected background events (S/N $\sim$ 1/3) in the corrected spectrum. This is due in part to the narrower width of the Z peak after the corrections, and in part to the effect mentioned above, *id est* the different behavior of signal and background under the effect of the jet corrections.

The third observation concerns the total fitted excess: the tag probability is a function of the invariant mass, and therefore the jet corrections might in principle modify the predicted background and the size of the excess (which are computed starting from the tag probability). Nevertheless, we observe the latter to be invariant: the integrals of the Gaussians all amount to about 80 events. Table 5.9 gives the value of the fit parameters for the four masses.

Correction	$M(GeV/c^2)$	σ_{jj} (GeV/c^2)	σ_{jj}/M_{jj}	$\Delta(\sigma_{jj}/M_{jj})$	N_{fit}
Standard	64.5 ± 6.5	17.0 ± 4.6	0.264		80.0
P_μ	74.6 ± 5.6	15.6 ± 6.3	0.209	-26 %	80.6
$P_\mu + \not{E}_T$	80.7 ± 5.7	16.4 ± 5.0	0.203	-30 %	86.3
Final	88.5 ± 4.9	13.6 ± 4.8	0.154	-71 %	80.2

Table 5.9: *Improvement of the parameters of Gaussian fits to the invariant mass distribution in the experimental data from the standard QDJSCO corrections to our final result. The last column shows the integrals of the fit Gaussians.*

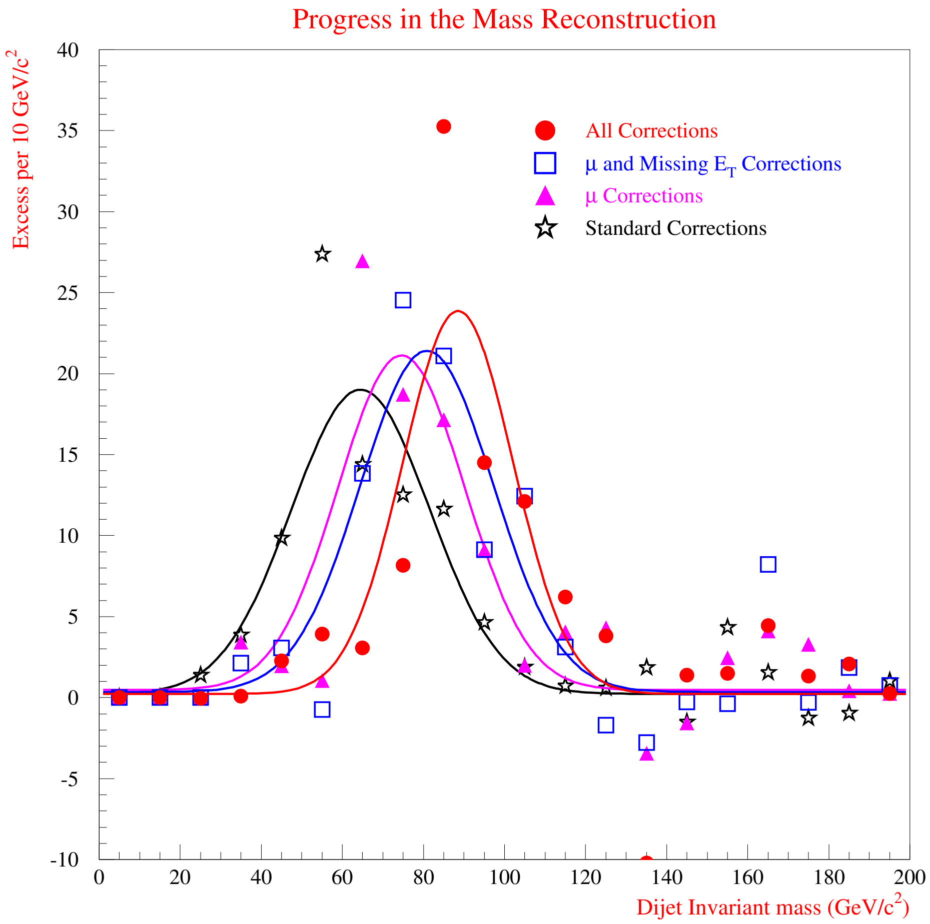

Figure 5.16: *Mass distributions of the excess at the four stages of the jet corrections. The four Gaussian fits already shown in Fig. 5.13 and Fig. 5.14 are compared to show the progress.*

In the next chapter we will describe our counting experiment, which allows to extract a significant signal from our selected dataset, as we have just given the reader a chance to see.

Chapter 6

The Counting Experiment

In the present chapter we describe a method to compute the number of background events selected by our kinematic cuts. After a brief introduction, the method is explained in Sec. 6.2. We describe the choice of bin size in the counting experiment, as well as the choice of the cuts on $\Sigma_3 E_T$ and $\Delta\Phi_{12}$, in Sec. 6.3. In Sec. 6.4 are given the results of the counting experiment, and in Sec. 6.5 are discussed some methods to estimate the systematic uncertainty in the method. In Sec. 6.6 we estimate the other possible contributions to the observed excess of events isolated with the counting experiment.

6.1 Introduction

After the kinematic selection and the double SECVTX tagging, we expect the dataset to be still dominated by QCD background events—almost exclusively direct $b\bar{b}$ production at this point—with a S/N ratio as high as 1/3 for events with $M_{jj} \sim 90\ GeV/c^2$. Under such circumstances, the mass distribution can be used to demonstrate the presence of a signal only if the background shape is very well under control.

The most obvious way to model the background shape is to use a sample of experimental events that are rich in background and are subjected to the same selection biases of the signal sample. To achieve that, we can remove the most discriminant selection cut, the double tag requirement, and look for events passing all the kinematic cuts imposed to the signal sample, having one b tag plus another jet that was not tagged but—here is a very important point—it had E_T higher than the minimum required by the tagging algorithm and it contained at least two tracks passing the P_T threshold set in SECVTX: such a jet is called *taggable*, to describe the fact that it possesses all the characteristics needed to contain an identificable secondary vertex in terms of E_T , track P_T , and pseudorapidity (jets falling out of the SVX acceptance are of course out of play since they do not feature SVX tracks). The dataset selected by these requirements therefore features exactly the same biases as the signal sample, but the fraction of signal it contains is expected to be about five times smaller.

By using events with one SECVTX tag and a second taggable jet passing all the kinematic requirements to model the background mass shape, it is indeed possible to fit a sizable Z peak in the signal sample: this will be the subject of Ch. 7. But we cannot assume without proof that events containing one SECVTX tag plus a taggable jet and events with two SECVTX tags behave absolutely the same way in the mass distribution: and alas, in fact they do not. The probability for a jet to be tagged by the SECVTX algorithm depends on many of the characteristics of that jet and of the event as a whole, and some of these are correlated, albeit only slightly, with the invariant mass of the dijet system. We can, in fact, expect the tag probability to be lower than average for low E_T jets—and they are of course most easily found in events with a low dijet invariant mass—since there are fewer charged tracks in these jets and they pass less frequently the track P_T thresholds set in the SECVTX algorithm. On the other hand, we may expect that for very high mass events the tag probability would be lower, due to the conspiracy of many different effects. First, there exists a positive correlation between the invariant mass of the dijet system and the absolute value of the pseudorapidity of the jets[1]: the tag probability falls with jet pseudorapidity as a result of the jet being less and less likely to contain many reconstructed charged tracks (due to the limited acceptance of the Central Tracking Chamber) giving a lower tagging rate for high mass events; second, very high energy jets are likely to feature very collimated tracks, so that the probability that a track is lost by the reconstruction algorithm increases, thereby lowering the tagging probability; third, highly boosted b quarks may travel past the first layer of the SVX detector before decaying[2], lowering the probability that the decay tracks are reconstructed; fourth, the fake and $c\bar{c}$ components of the SECVTX datasets have been shown to have higher E_T jets[3], and therefore to give a larger contribution to the high mass tail: these events have of course a much smaller probability of containing a second SECVTX tag[4].

Nevertheless, the choice of "(+0)" events as a background-enriched sample appears the best one. These events are as similar as possible to the double tagged events, with a sustainably small contamination from the $Z \to b\bar{b}$ signal. If we wished to have a background sample with an even smaller signal contamination, we could use events with *two* taggable jets (no SECVTX tags), and apply to them all the selection criteria used to obtain our signal sample: the fraction of $b\bar{b}$ events in such a dataset would be small by construction, and the fraction of signal consequently negligible. Doubly taggable events would manifest the aforementioned tagging bias elevated to the second power; but—more alarmingly—even if we could keep that

[1]This happens because the higher the pseudorapidity for a jet, the larger its energy for a given E_T (the latter is correlated to the tag rate, due to the P_T thresholds in SECVTX), and consequently the higher the reconstructed mass, if—as most often happens—the two jets have pseudorapidity of opposite sign.

[2]For a 50 GeV b quark this happens about 1% of the time; for a 250 GeV one, the probability increases to about 40%.

[3]For the charm component this is due to the increased probability of giving a well-displaced secondary vertex if the c quark is highly boosted. The fake component is instead likely to be composed of jets with high E_T because they have lots of high P_T tracks, and these have a higher probability to fake a muon (because they sometimes manage to make it through the calorimeters).

[4]A plot of the tag probability is shown in Fig. 6.4.

bias under control, the perspective of using a sample of events with light quark or gluon jets to describe a dataset of a pure $b\bar{b}$ final state would be very unattractive: demonstrating that a QCD event with light quarks or gluons in the final state behaves like a QCD $b\bar{b}$ event with respect to variables such as the dijet invariant mass or our selection variables would be very hard, if possible at all, and anyway outside the scope of the present work[5].

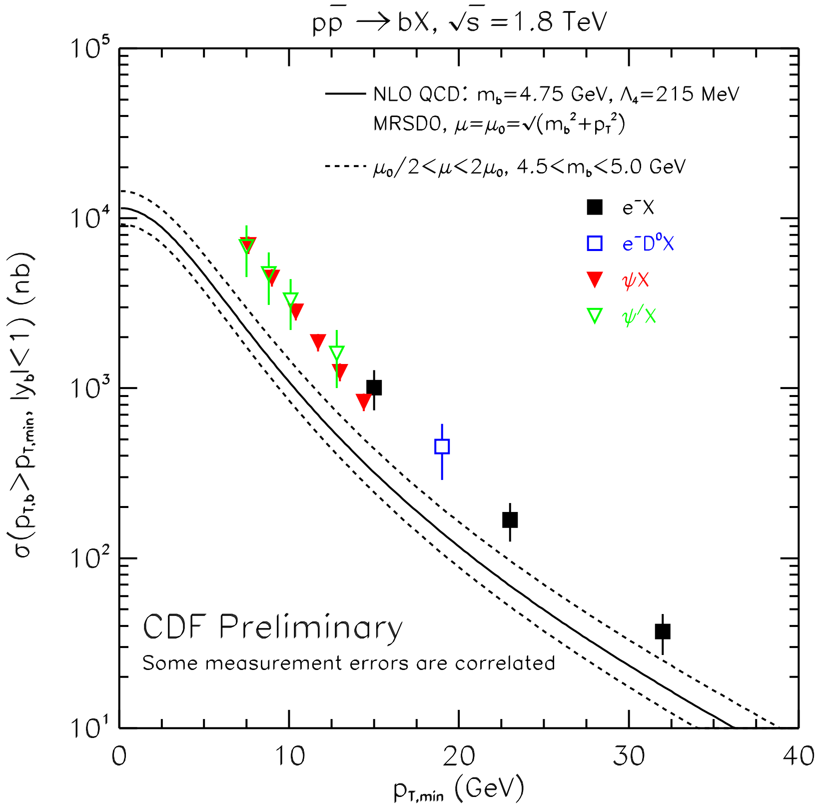

Figure 6.1: *Compilation of different CDF measurements of the integral b-quark production cross section as a function of the P_T threshold. The cross section at* 10 *GeV/c amounts to about* 4 μb.

We do not wish to convince the reader that (+0) events are the best possible dataset to represent the background in the (++) sample. In fact, a Monte Carlo sample of QCD $b\bar{b}$ events passing all our selection criteria, including the double SECVTX tagging, could represent a better choice, although we would then be concerned with the accuracy of the description of the simulated background processes. The unfeasibility of this plan is due anyway to the time required for such a simulation: to obtain a sample of (++) events of size similar to that of the (+0) experimental data—a reasonable guideline in order to have enough statistics[6]—we would have to generate

[5]The "(00)" sample will rather prove its usefulness for studies of systematic uncertainties connected with the tagging bias, as will be described in Sec. 6.5.

[6]We rely, in fact, on the large statistics of the (+0) sample to have a small error on our

about ten times the integrated luminosity of the experimental data, *id est* a really huge amount of $b\bar{b}$ events. A back-of-the-envelope computation of the needed CPU time requires the input of one fundamental parameter: the minimum P_T of the final state b quarks generated by the Monte Carlo. This number has to be set to a value small enough that no bias is generated in the final dataset: that is, small enough that a pair of final state b quarks having a P_T any smaller than that value have a vanishing probability of yielding a triggering muon and two tagged jets. Studies made for a different analysis at CDF have shown that the P_T^{min} must not be larger than 10 GeV/c. Setting a minimum P_T of 10 GeV/c corresponds to generating a process with a cross section of about four microbarns, as can be seen in Fig. 6.1: to obtain an integrated luminosity $\sim$ 10 times larger than our experimental data one should then produce several billion events, a task that would require more than a year of CPU time on the fastest computers at our disposal.

Having discussed why our choice of (+0) events as a background sample is forced, we can now turn to our project of using them to obtain a prediction for the number of background events having a second tag: a knowledge of this number will allow us to check whether the observed number of (++) events is compatible with the expected background alone, or there is an excess due—hopefully—to the searched signal; in the latter case we will be able to quantify the signal significance. What we need is a method to estimate the probability that, given an event with a SECVTX tag, the second jet is tagged as well. As we discussed in Sec. 4.1, a large part of the events in the SECVTX sample contain heavy flavor decays, but the fraction of double tags is nonetheless much smaller than that expected from the $Z \rightarrow b\bar{b}$ signal, because the fraction of events with two angularly separated central jets both coming from b-quark decay—essentially, the direct $b\bar{b}$ production component—is small. The data themselves can thus be used to estimate the tag probability, given the smallness of the S/N ratio at the SECVTX level.

At CDF the problem of estimating the amount of backgrounds yielding b-quark jets is not new at all: for instance, detailed studies of the rate of tagging in inclusive jet datasets have allowed accurate predictions of the amount of background events in $t\bar{t}$ -enriched samples, by means of the construction of multidimensional "tagging matrices" parametrizing the probability of a jet to contain a SECVTX tag as a function of jet and event observables such as the number of charged tracks in the jet, the jet E_T, the jet rapidity, and so on[7]. We cannot rely on a matrix method and simply reproduce that approach for our analysis, since what we need is not a single tag probability, but the probability that, given an event already containing a SECVTX tag, another jet is tagged. Now, in our dataset the conditional probability

background prediction: given an expected S/N ratio of 1/3 at the Z peak, the relative uncertainty in the background needs to be kept at the 5% level, or smaller, systematics incuded. As we have already said, the smaller the signal is, the better the background has to be known; and our signal is indeed small.

[7]These methods work by virtue of the much higher tagging probability of the signal one wishes to discriminate than the tagging probability of background processes. In principle, this is also our case (but see the text): as we have discussed in Sec. 4.1, the much lower tagging rate of the experimental data w.r.t that expected for the $Z \rightarrow b\bar{b}$ signal is a confirmation of our understanding of the sample composition at the SECVTX selection level.

of finding a tag in a jet, given the observation of a tag in another one, depends on too many variables: not only those relative to the jet under study, but also the corresponding ones for the jet that is already tagged; moreover, the presence of the triggering muon in one or the other makes a sensible difference, since the stiff muon track increases the tagging probability, and of course a jet that contains a muon is more likely to come from b-quark decay than a jet that does not. We do not have in our dataset sufficient statistics to fill the multidimensional space of all these variables and obtain a good estimate of the dependence of the tag probability on the value of them. We will instead rely on the observation that our tag probability, which can be easily defined *ad hoc* as a ratio between events with two SECVTX tags and $(+0)$ events, is independent on the value of the two kinematic variables we have selected our data with, as can be seen in Fig. 6.8 on page 106: for jet pairs not too far from being back-to-back, the tag probability is indeed flat as a function of their opening angle in the transverse plane, and the same is true regardless of how much additional clustered energy is present in the event. We note *a posteriori*, since we can only do it now, that our choice of the variables used in the kinematic selection was influenced by their "good" behavior with respect to the tagging probability.

6.2 The Method

We select four subsamples from our data: events that fall in the kinematic region selected by our cuts ($\Delta\Phi_{12} > 3$ radians, $\Sigma_3 E_T < 10\ GeV$, in the following referred to as the "*Signal Zone*"), and events that fail those requirements but fulfill the looser cuts $\Delta\Phi_{12} > 2.5$ radians, $\Sigma_3 E_T < 50\ GeV$ (hereafter "*Normalization Zone*"), from both the $(++)$ and the $(+0)$ datasets[8].

We can then define a tagging probability as the ratio between $(++)$ and $(+0)$ events that fall in the Normalization Zone, and extrapolate it into the Signal Zone (see Fig. 6.2), obtaining an absolute background prediction for the double tags in the Signal Zone:

$$N^{++}_{exp,in} = N^{+0}_{obs,in} \times (N^{++}_{obs,out}/N^{+0}_{obs,out}). \tag{6.1}$$

This procedure is more effectively carried out after a binning of each dataset in the reconstructed invariant mass of the dijet system: that way we obtain a tag probability as a function of mass,

$$F(M) = N^{++}_{obs,out}(M)/N^{+0}_{obs,out}(M), \tag{6.2}$$

and a prediction

$$N^{++}_{exp,in}(M) = N^{+0}_{obs,in}(M) \times F(M) \tag{6.3}$$

[8]The choice of discarding events with a too small azimuthal opening angle between the leading jets, or too high additional transverse energy, is not strictly necessary; besides, the cuts defining the Normalization Zone are very loose and very few events are excluded by them. We do it because we do not want to evaluate a tag probability using events that are "too distant" from the region where we want to extrapolate it, but the impact of these cuts is really small. An evaluation of the systematic error due to the somewhat arbitrary choice of these boundaries is described in Sec. 6.5.

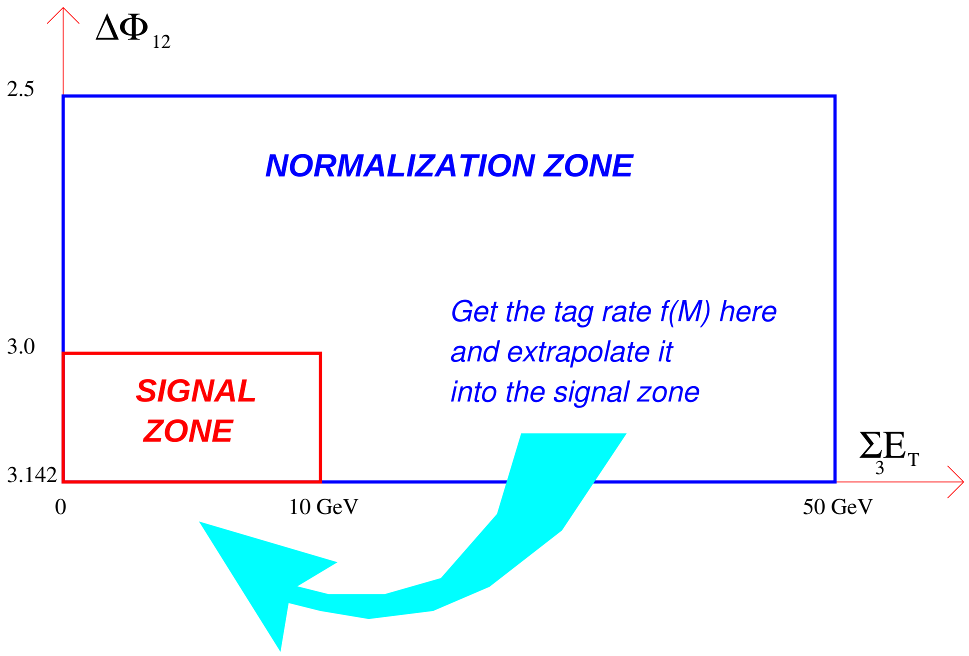

Figure 6.2: *Graphic explanation of the counting method. See the text for more details.*

that can be compared to the observed counts bin by bin in the mass histogram. An excess around the Z mass will be bias-free evidence for the signal, if away from the Z pole the excesses are compatible with zero.

It must be pointed out that the background prediction has to be considered an overestimate if—as we trust—there are Z events in our muon dataset: the $(+0)$ sample in the Signal Zone then contains some of these, and when the number of events in that sample is multiplied by the tag probability obtained from the Normalization Zone, we predict more background events than we should. Moreover, there have to be $Z \to b\bar{b}$ events also in the two samples used for the determination of the tag probability, and—due to the higher chance of identifying b-quark jets from Z decays than QCD events—the tag probability in the Z mass region is overestimated, giving a further negative bias to the observed excess. In this section we will use the quantitative results of the counting method only to show the inconsistency of the background prediction with the observed number of events, and therefore these two effects are not a concern. For the correct evaluation of the number of events in the dataset, and the subsequent measurement of a cross section for the searched process, we will use the likelihood fitting technique described in Ch. 7.

6.3 Maximization of the Expected Signal Significance

In order to extract the highest significance from a possible excess of events in our dataset we have studied the expected signal significance as a function of the width of the mass bin where we look for the signal. The Monte Carlo simulation for the Z signal predicts that after our jet momentum corrections we should obtain a Gaussian mass distribution, peaked at 90 GeV/c^2 with a width of 12 GeV/c^2: with this information, and by using our knowledge of the signal cross section, integrated luminosity of the dataset, and efficiency of the selection variables, we can easily obtain an estimate of the number of signal events included in the bin around 90 GeV/c^2:

$$N_Z(90-\Delta, 90+\Delta) = \frac{\sigma_Z \times BR(Z \to b\bar{b}) \times \int \mathcal{L}dt}{\epsilon_{kin}\epsilon_{\mu}\epsilon_{2SVX}} \times \frac{\int_{90-\Delta}^{90+\Delta} exp(-0.5*(\frac{m-90}{12.3})^2)dm}{\int_{-\inf}^{+\inf} exp(-0.5*(\frac{m-90}{12.3})^2)dm}.$$

To obtain the number of background events included in the bin as a function of its width, we use the method already outlined for the counting experiment: given the boundaries of the bin where we look for the signal, we compute the tag probability in that bin as the ratio $(++)/(+0)$ of events in the Normalization Zone falling in that mass interval, and multiply it by the number of $(+0)$ events observed in the Signal Zone:

$$N_{bgr}(90-\Delta, 90+\Delta) = \frac{N_{N.Z.}^{(++)}(90-\Delta, 90+\Delta)}{N_{N.Z.}^{(+0)}(90-\Delta, 90+\Delta)} \times N_{S.Z.}^{(+0)}(90-\Delta, 90+\Delta).$$

Our definition of the signal significance for the present purpose is an approximation[9]: we define it as $N_Z/(\sigma_{bgr}^{stat} \oplus \sigma_{bgr}^{syst} \oplus \sqrt{N_Z + N_{bgr}})$. The statistical error in the background prediction is evaluated with the actual number of events given by the counting method, while the systematic error is defined as a fixed percentage of the expected background; these are quadrature-summed to the Poisson fluctuation of the expected number of observed events—computed as the sum of the expected background plus the expected signal.

By scanning the bin size and computing the significance thus defined, we are able to decide what is the best possible bin size around 90 GeV/c^2. As shown in Fig. 6.3, we find that the best choice, as far as the maximization of the significance is concerned, is a bin size of 40 GeV/c^2. This is what we will use to quote our excess and significance in the counting experiment.

The same game can be played to justify the cuts on the kinematic variables, as we promised in Sec. 4.2, by fixing the bin size to 40 GeV/c^2 and varying instead the value of $\Sigma_3 E_T$ dividing the Signal Zone from the Normalization Zone; we use the

[9]In fact, the only accurate method involves the calculation of the probability for a Poisson fluctuation of the expected background, given its total error, to the number of events observed—in this case the number of background plus expected signal. We will use this method when quoting the final numbers for the significance.

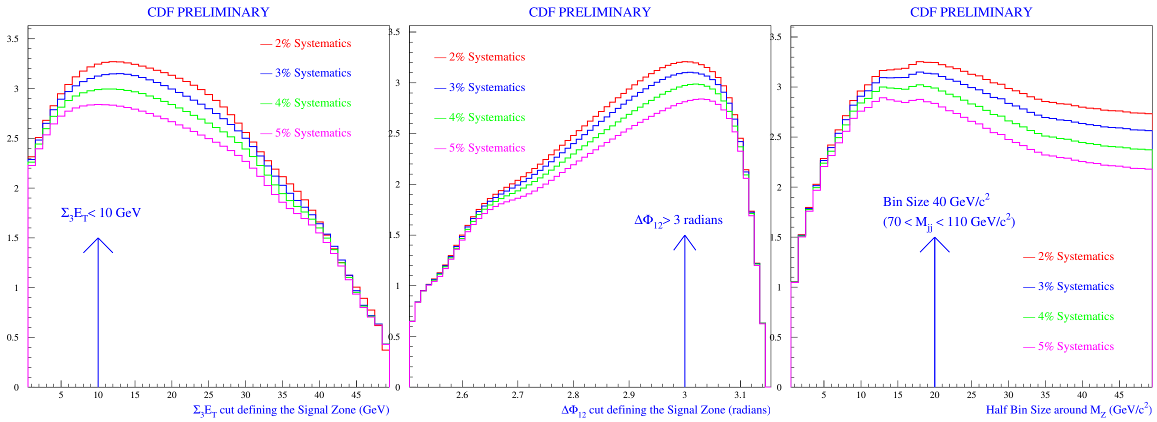

Figure 6.3: *Maximization of the expected signal significance as a function of the cut on the kinematic variables. Left: choice of cut on* $\Sigma_3 E_T$*; center: choice of the cut on* $\Delta\Phi$*; right: choice of the bin size. Four different possibilities for a systematic uncertainty in the background evaluation are considered.*

$\Sigma_3 E_T$ distribution obtained from $Z \rightarrow e^+e^-$ decays[10] to get the expected amount of $Z \rightarrow b\bar{b}$ decays in the Signal Zone as a function of the cut. The results are shown in Fig. 6.3, again for four different choices of a systematic uncertainty in the background prediction: as one can see, the chosen cut at 10 GeV does maximize the expected signal significance for the considered values of a systematic uncertainty. Finally, we can do the same for the $\Delta\Phi_{12}$ cut: again, the choice $\Delta\Phi_{12} > 3$ is close to the optimal value (Fig. 6.3).

6.4 Results of the Counting Experiment

Having justified the values of the cuts and decided the most favorable binning in the mass distribution, we can turn the crank and study the results. These are detailed in Table 6.1 and in Fig. 6.4.

The excess in the third bin is quite significant, if we do not consider any systematic uncertainty associated with the extrapolation of the tag probability: the Poisson probability of 248.49 ± 8.95 events fluctuating to 318 is barely 0.000095, corresponding to 3.6 standard deviations for a single-sided Gaussian distribution. In the following sections we will discuss the possible systematic errors associated with the counting experiment: a fair estimate for the total systematics may be placed at 3%, so that our background prediction becomes $N_{exp} = 248.49 \pm 11.65$. In that case we obtain a Poisson probability of 0.00032, equivalent to 3.2 standard deviations.

The numbers quoted above are already rather convincing evidence for the presence of $Z \rightarrow b\bar{b}$ decays in our dataset. But we can use a smaller bin size than the one we found optimal for significance purposes, to study the *shape* of the excess of

[10]This distribution is correctedfor the bias due to the higher instantaneous luminosity of the $Z \rightarrow e^+e^-$ data with respect to the SECVTX data. More details of the unbiasing procedure are given in Sec.8.2.

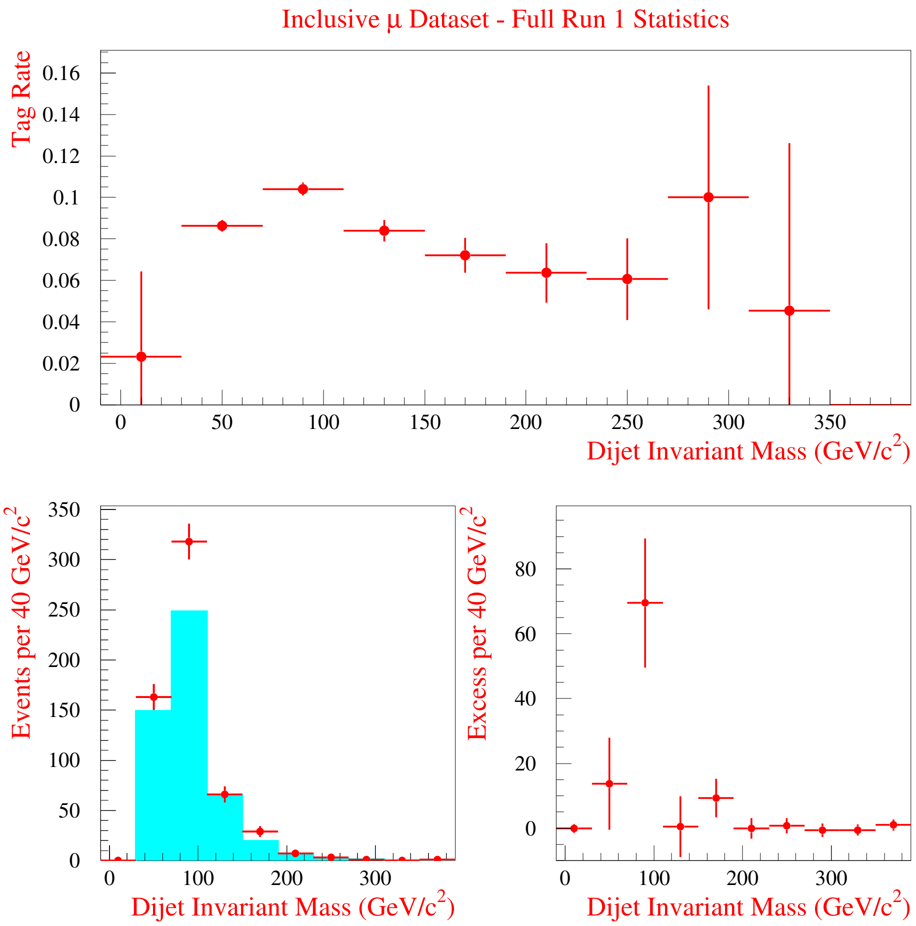

Figure 6.4: *Results of the counting experiment with a 40 GeV binning. Top: the tag probability obtained in the Normalization Zone. Bottom: on the left the double tagged data (points with error bars) are compared with the background prediction (histogram); on the right is shown the excess of events above predictions.*

events in the mass distribution: we then expect to see a Gaussian excess[11] peaked at 90 GeV/c^2, with a r.m.s. of 12.3 GeV/c^2 according to PYTHIA V5_7[12]. The results of the counting experiment with this smaller bin size confirm the expectations: the excess is very well fit by a Gaussian function, whose mean and width are in excel-

[11] As noted earlier, the tag probability is a function of the dijet invariant mass: therefore, strictly speaking one would expect that our signal in the (++) dataset is not anymore perfectly Gaussian. However, the tag probability is rather flat in the region of mass around 90 GeV/c^2, since the turn-on effect due to the track-P_T threshold set in SECVTX for the secondary vertex reconstruction affects events with masses much lower, and the competitive effect of a decrease at high mass is still negligible. We can therefore still speak of a Gaussian shape with very good approximation.

[12] See Sec. 5.4.

Mass Interval	Observed	Expected	Excess	Exp. Z
$0-30\ GeV/c^2$	0	0.05 ± 0.09	-0.05 ± 1.40	
$30-70\ GeV/c^2$	163	149.30 ± 6.13	13.70 ± 14.16	2.0 ± 0.5
$70-110\ GeV/c^2$	318	248.49 ± 8.95	69.51 ± 19.95	50.2 ± 12.7
$110-150\ GeV/c^2$	66	65.51 ± 4.68	0.49 ± 9.38	2.8 ± 0.7
$150-190\ GeV/c^2$	29	19.70 ± 2.59	9.30 ± 5.97	
$190-230\ GeV/c^2$	7	7.06 ± 1.73	-0.06 ± 3.16	
$230-270\ GeV/c^2$	3	2.24 ± 0.82	0.78 ± 2.38	
$270-310 GeV/c^2$	1	1.60 ± 0.95	-0.60 ± 2.00	
$310-350\ GeV/c^2$	0	0.55 ± 0.98	-0.55 ± 1.71	
$350-390\ GeV/c^2$	1	0.00 ± 0.00	1.00 ± 1.76	

Table 6.1: *Results of the counting experiment with a* $40\ GeV/c^2$ *binning.*

lent agreement with the values expected from Monte Carlo (Fig. 6.5, lower plots). We can also check if the application of our corrections to the jet momenta biases our prediction: the tag probability is a function of the invariant mass, and it will change for a different definition of the latter. The counting experiment, performed with jets corrected with the standard QDJSCO corrections alone, gives again a nice Gaussian excess, this time peaking, as expected, around $65\ GeV/c^2$, with a larger width (Fig. 6.5, upper plots). These excellent results make us confident that we have isolated a Z signal that behaves as we expect with regards to the variables we use to correct the jet energies[13].

We have further ways to convince ourselves that what we are observing is a Z signal, if we are not yet satisfied with the nice shape it shows as a function of the invariant mass of the dijet system. In fact, the counting method may allow us to study how the excess behaves in the same kinematic variables we have selected it with, which as we know have a discriminating shape for the electroweak process from QCD. To do that, we select $(++)$ and $(+0)$ events in the Signal Zone that fall in the interval $70 < M_{jj} < 110\ GeV/c^2$: as Table 6.1 shows, we have an excess of 69.5 ± 20.0 events there. We can then construct the $\Sigma_3 E_T$ and $\Delta\Phi$ distributions for $(++)$ events, and compare them to the $(+0)$ distributions scaled down by the tag probability $F(70 \div 110\ GeV/c^2)$ obtained in the Normalization Zone. The excess can then be studied as a function of each kinematic variable by subtracting from the $(++)$ spectrum the expected background distribution. If our excess were an artifact of the counting extrapolation, then it would hardly be distributed according to the expectations for $Z \to b\bar{b}$ events. But the agreement is very good, as can be seen in Fig. 6.6 and Fig. 6.7: this observation represents a further indication that we are indeed observing the signal sought.

[13]The same exercise can be played at each stage of the jet momentum corrections described in Ch. 5. The relevant plots are Fig. 5.13 and Fig. 5.14; see also Fig. 5.16

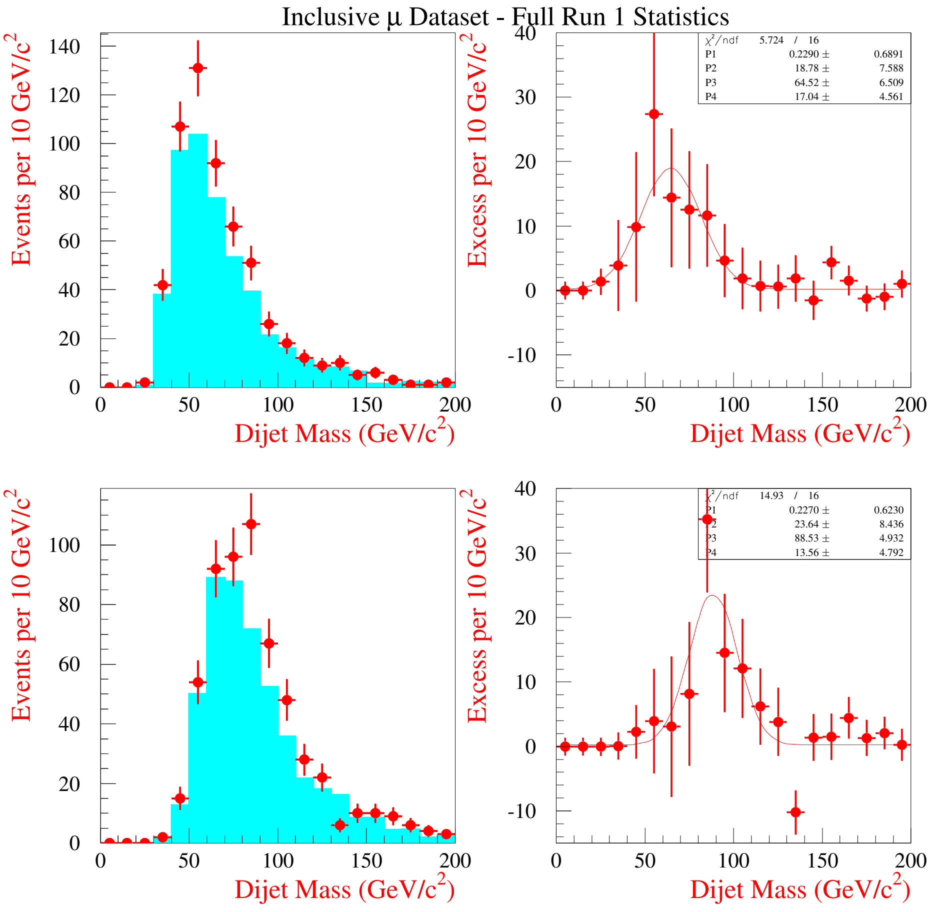

Figure 6.5: *Results of the counting experiment with a* 10 *GeV*$/c^2$ *binning. Top: mass distributions (left) and excesses in each bin of the mass histogram (right), for the mass computed with the standard QDJSCO jet energies; bottom: the same for the mass computed after the application of the improved jet corrections described in Ch. 5.*

6.5 Estimations of the Systematic Error of the Counting Method

The visual impact of the mass plots shown above gives us confidence in the smallness of any systematic error that may affect the counting method. In fact, the observed and predicted events away from the Z pole are in excellent agreement, whatever bin size one chooses to evaluate the tag probability; but we need an estimate for the possible systematic uncertainties related to the extrapolation of the tag probability from the Normalization Zone to the Signal Zone. We have various ways to do that: in the following section we describe each of them.

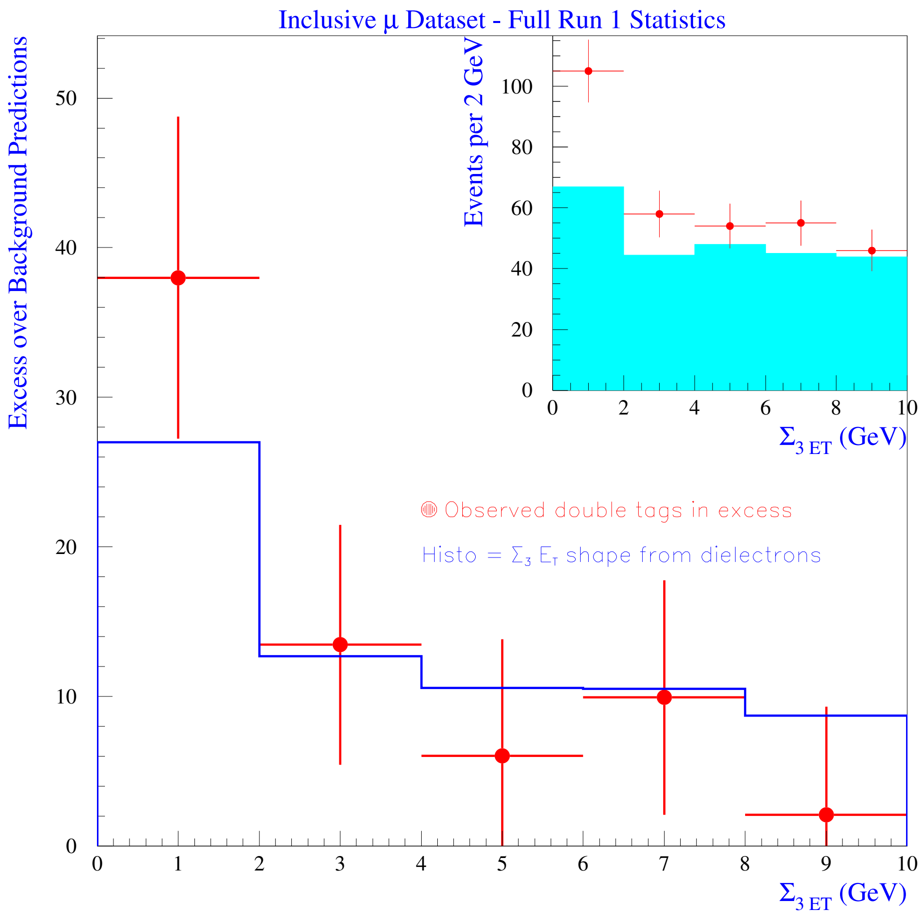

Figure 6.6: *The excess observed in the* $70 \div 110\ GeV/c^2$ *bin is distributed according to expectations in the* $\Sigma_3 E_T$ *variable. The inset shows the total (++) distribution (points with error bars) compared to background expectations (full histogram).*

6.5.1 Flatness of the Tag Probability

The first thing to demonstrate is the flatness of the tag probability with respect to the kinematic variables used for the extrapolation. Here we shall define the tag probability inclusively as the ratio of events having two tagged jets divided by the number of events with a tagged jet and one taggable jet. In order to avoid the bias of a signal contamination to the tag probability[14], we can study each variable in the region excluded by the selection cut on the other: therefore, we compute the tag probability as a function of $\Delta\Phi$ when the $\Sigma_3 E_T$ is greater than 10 GeV, where the background should be dominant; and as a function of $\Sigma_3 E_T$ when the

[14]As already noted, the signal has a higher tag probability and must be eliminated from the samples used to estimate a background tag probability.

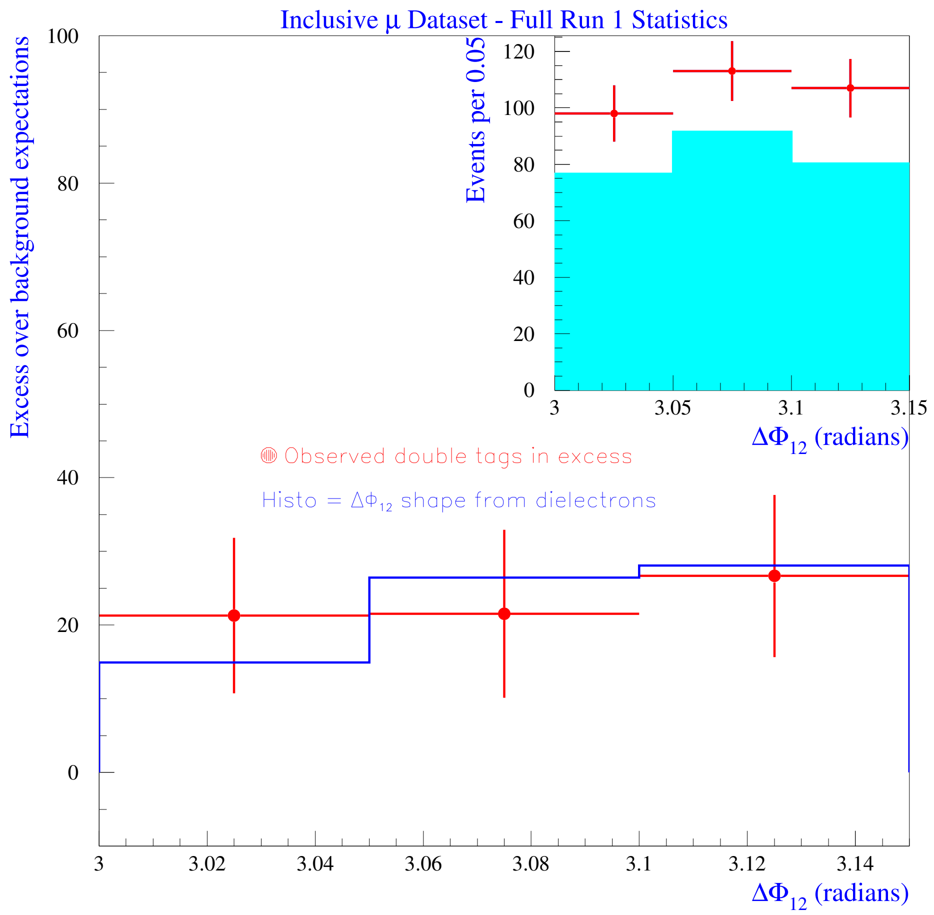

Figure 6.7: *The excess observed in the* $70 \div 110$ *GeV*$/c^2$ *bin is distributed according to expectations in the* $\Delta\Phi$ *variable. The inset shows the (++) distribution (points with error bars) compared to the background expectations (full histogram).*

$\Delta\Phi$ is smaller than 3 radians, again cutting out the higher S/N region. The plots in Fig. 6.8 show that we can confidently think of the tag probability as independent from the kinematic variables in the region where we extrapolate it, the constant fits being both good, with a χ^2 close to 1 per degree of freedom.

It is to be noted that what we have now shown does not by itself warrant against any bias. In principle, events with different dijet mass could have non-flat tag probabilities in $\Delta\Phi$ or $\Sigma_3 E_T$ hidden behind the cumulative distributions shown in Fig. 6.8. The plots can be redone using events with about the same dijet mass; the tag probabilities remain flat, but the statistical power of the subsamples gets too poor for a conclusive proof. We have a better way to estimate the systematics, as will be clear in the following.

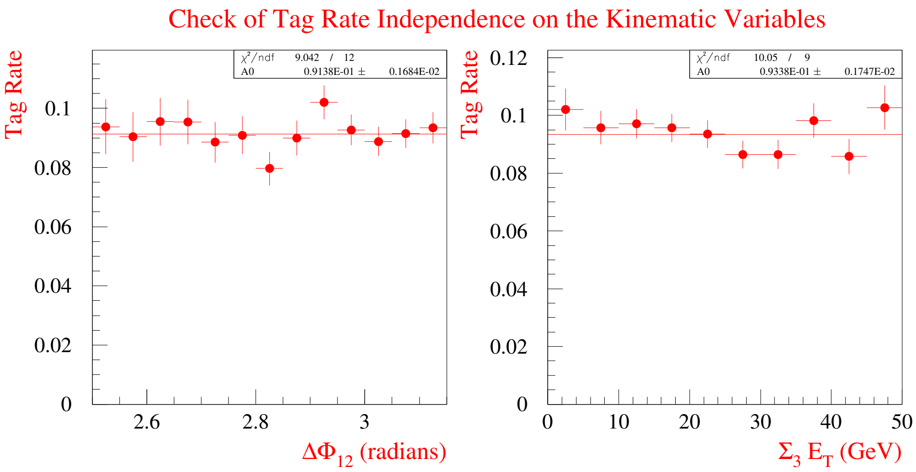

Figure 6.8: *Left: the tag probability is flat as a function of the azimuthal angle between the two leading jets, in events with* $\Sigma_3 E_T > 10$ *GeV; right: the same holds as a function of the* $\Sigma_3 E_T$*, for events with* $\Delta\Phi_1 2 < 3$.

6.5.2 Predicting Single Tags with Zero Tag Events

If the presence of a reconstructed secondary vertex in a jet is indeed uncorrelated with our kinematic variables in events where another jet contains a SECVTX tag, one might guess that the same holds also in a dataset where no other jet is tagged. This may not be true in principle: if one imagines, for instance, that the data with two taggable jets include a process—say, for instance, diffractive charm production—which is absent in the sample with one or more SECVTX tags (due in this case to the low tag probability of charm jets), and if that particular process has a distribution in the two kinematic variables used for the extrapolation that is different from those of the other processes contributing to the dataset[15], then the prediction of events with one tag from events with a taggable jet may be jeopardized, even if the prediction of events with two tags from events with one tag is still working.

We cannot be sure of the sample composition of either the $(+0)$ or the (00)[16] datasets, and studying it in detail would carry us too much off our path, but we can check whether the application of the counting method in the prediction of $(+0)$ events with (00) events makes sense or not. If it does, we can elect the residual discrepancy to a systematic uncertainty due to the extrapolation; if it does not, we do not need to despair, since we have many other ways to verify the correctness of

[15] This may well be the case for the process chosen as an example: charm jets may be present in the dataset selected by the muon trigger, and $c\bar{c}$ events have a lower tag probability than $b\bar{b}$ events; a diffractive production would surely show a much smaller calorimetric activity, and the $\Sigma_3 E_T$ distribution would then be different.

[16] As the reader may have already guessed, we label "(00)" those events having two taggable jets.

our method in the $(+0)$ *versus* $(++)$ case.

It is to be noted that the Z contamination in the $(+0)$ sample is not completely insignificant (we expect to have 117 ± 26 $Z \to b\bar{b}$ decays in the signal region of the $(+0)$ dataset[17] in a background of about 5,200 QCD events), although a similar contamination in the (00) sample from $Z \to c\bar{c}$ and $W \to c\bar{s}$ decays, falling more or less in the same dijet mass bins, will cancel out a good part of it[18]. All in all, we do not expect to be able to discern any discrepancy from the background predictions due to Z contamination, because of the smallness of the S/N ratio in the $(+0)$ sample.

The results of this counting experiment are shown in Fig. 6.9. The prediction is in excellent agreement with the observed number of events over the whole mass range accessible with the dataset, so that by dividing the observed by the predicted spectrum and fitting to a constant we get an average ratio $R = 0.986 \pm 0.016$, which again confirms the absence of any sizable bias due to the extrapolation of the tag probability.

6.5.3 Events with Three Jets

Our extrapolation of the tag probability in the variable $\Sigma_3 E_T$ may in principle be affected by many biasing factors. We need to check whether the amount of radiation outside the leading jets has an impact of any kind on the tag probability. A nice way to do that is to define a new variable, $f_4 = \Sigma_4 E_T$, evaluated as the sum of uncorrected cluster energy from the fourth to the n-th $R = 0.7$ cluster found in the event. This variable is defined only for events having at least three jets[19], but it has in principle the same physical meaning as our extrapolation variable $\Sigma_3 E_T$, and we can therefore choose events with at least a third $E_T > 10\ GeV$ jet to check whether a tag probability extrapolation would work for this variable. The big advantage of this game is that we do not expect to have a significant amount of Z events in the subsample of events with a third $E_T > 10\ GeV$ jet, and thus we can compare the background with its estimate without having to worry too much about anything else.

We select events from our SECVTX dataset having three or more jets with $E_T > 10\ GeV$, and perform a counting experiment exactly as we do with the normal dataset. We of course find no evidence of a signal (we expect only about four $Z \to b\bar{b}$

[17] This number results from the number of Monte Carlo events passing the cuts defining the $(+0)$ sample, corrected for the scale factor according to the following prescription: $N'_{+0} = N_{+0} \times (1 - (f-1)\frac{\mathcal{P}}{\mathcal{P}_0})$, where $\mathcal{P}_0$ denotes the probability that a jet is taggable but not tagged, $\mathcal{P}$ the usual SECVTX tagging probability per jet, and f is the relevant scale factor. The efficiency of the kinematic cuts defining the Signal Zone on the $Z \to b\bar{b}$ signal has been obtained in Sec. 4.2.

[18] Once all known effects are accounted for in the relative Monte Carlo samples, we expect the presence in the Signal Zone of 201 ± 37 $W \to c\bar{s}$ and 21 ± 4 $Z \to c\bar{c}$ decays having two taggable jets (see also Table 6.3), plus 36 ± 12 from the $Z \to b\bar{b}$ process. The ratio between $(+0)$ and (00) events is roughly equal to 25% around the Z mass, and therefore the sum of these electroweak processes is expected to cancel out about half of the excess from $Z \to b\bar{b}$ events in the $(+0)$ sample.

[19] For $N_j = 3$ one defines $f = 0$, while f cannot be defined for $N_j < 3$.

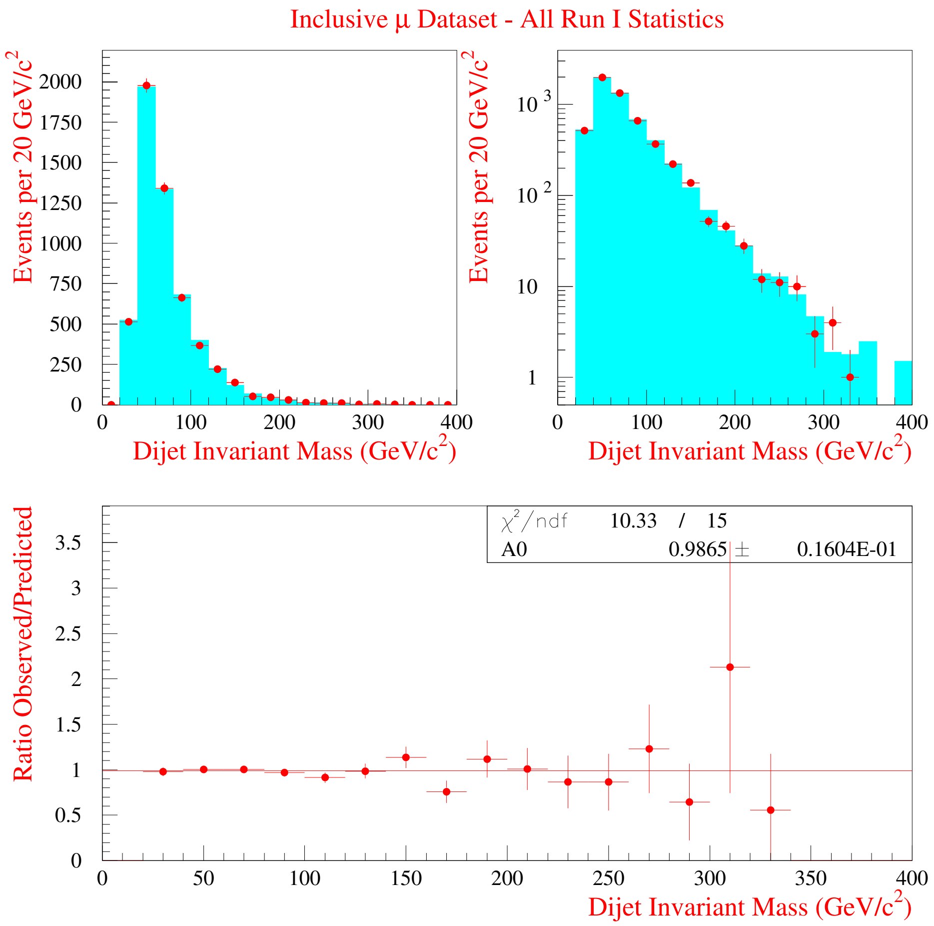

Figure 6.9: *Top: the mass spectrum for events with one tag and one taggable jet (points with error bars) is compared with the background prediction obtained from double taggable events (histogram); on the right, a logarithmic plot unveils the perfect agreement in the event rate over three orders of magnitude. Bottom: the ratio between observed and predicted events is shown to be constant and very close to 1.*

events with two tagged jets, a third jet, $\Delta\Phi_{12} > 3$, and $f_4 < 10\ GeV$)[20], and find a very good agreement between the observed number of events and background predictions. The ratio of these is flat in invariant mass (see Fig. 6.10), and is well fit by a constant equal to $R = 1.030 \pm 0.124$. Albeit with large errors, we get an indication of a systematic error at the 3% level.

We can also check if the same extrapolation has a different impact on the counting method when one tries to predict the number of (+0) events with events having

[20] The efficiency of the combined ($E_T^3 > 10\ GeV$, $f_4 < 10\ GeV$) calorimetric cut is as usual derived from $Z \to e^+e^-$ data: in Z events with $\Delta\Phi_{12} > 3$ radians, only 2.8% of them fulfill the two requirements.

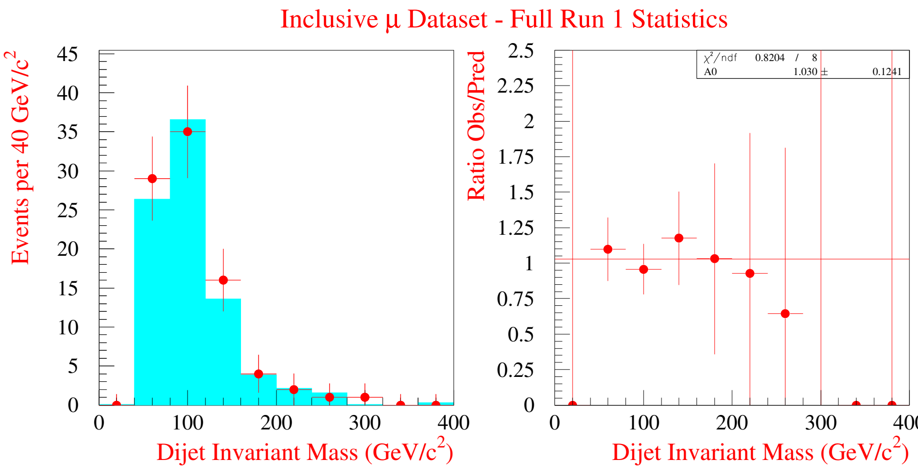

Figure 6.10: *Left: the mass spectrum of events with two tags, a third jet, and* $\Delta\Phi > 3$, $f_4 < 10$ *GeV (points with error bars) is predicted with excellent accuracy by the counting method, extrapolating the tag probability from the high* f_4*, low* $\Delta\Phi$ *region. Right: the ratio of observed events divided by the background prediction is flat in invariant mass, and well fit by a constant not far from 1.0.*

two taggable jets. The two datasets have higher statistics and they have a negligible signal contamination. The prediction is again in excellent agreement with the observed mass spectrum of single tags, and from a constant fit to the ratio observed/predicted we get this time $R = 0.971 \pm 0.043$: again, the systematics can be estimated at the 3% level.

6.5.4 Other Estimates

Probably the most straightforward way—although not the most powerful—to check for a systematic error in the background prediction is represented by fitting to a constant the ratio of observed and predicted events for the (++) sample where we extract an excess. Of course, the presence of the Z signal has to be sidestepped by excluding from the fit those bins where the Z signal contributes. By excluding the third bin in the counting experiment (see Table 6.1), for instance, we obtain a good fit ($\chi^2/dof = 0.43$ for six d.o.f.) and a ratio $R = 1.081 \pm 0.076$. Performing the counting experiment with a 10 GeV binning, as shown in Fig. 6.5, and again excluding the region $70 \div 110$ GeV/c^2, we obtain instead $R = 0.951 \pm 0.072$ ($\chi^2/dof = 0.87$, 26 d.o.f.); a 20 GeV/c^2 binning yields $R = 1.020 \pm 0.074$ ($\chi^2/dof = 0.88/14$ d.o.f.). These checks, although vitiated by the large statistical uncertainties on the fitted value of R, seem to point again to a very small systematic bias of the counting method: the differences in the values of R are to be attributed to statistical

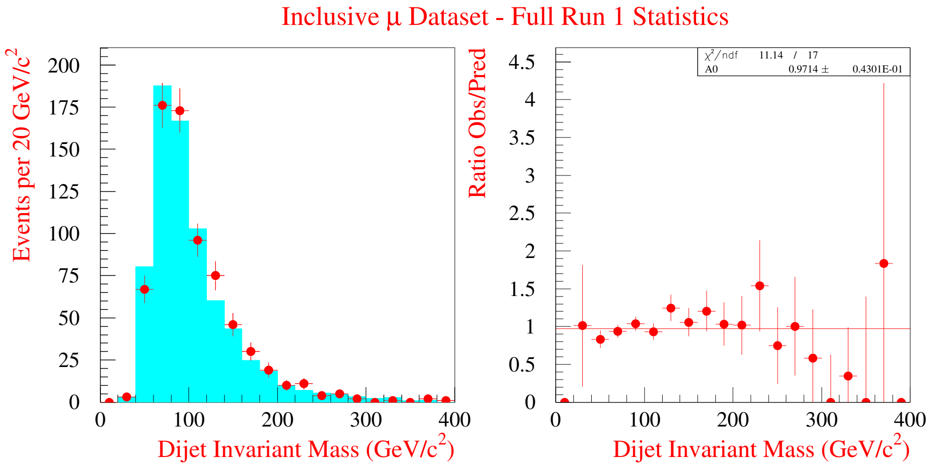

Figure 6.11: *Left: the mass spectrum for events with one tag, a taggable jet, and a third jet over* 10 GeV*,* $\Delta\Phi > 3$ *and* $f_4 < 10$ GeV *(points with error bars) is well predicted by the counting method, extrapolating the tag probability from the high* f_4*, low* $\Delta\Phi$ *region. Right: the ratio of observed events divided by the background prediction is flat in invariant mass, and well fit by a constant very close to 1.0.*

fluctuations.

6.5.5 Other Systematics

Up to now, we have discussed various methods to look for a possible bias due to the extrapolations of the tagging probability from the Normalization Zone to the Signal Zone. The last systematic error to consider is the one due to the choice of the cuts defining the Normalization Zone: we chose these boundaries rather arbitrarily, although moved by the just cause of excluding events featuring a totally different jet multiplicity and radiation topology, which would influence unpredictably the tag probability. To estimate a systematic error due to the arbitrary choice we can change the boundaries of the Normalization Zone, observing the variations in background prediction that we get for the search bin $70 \div 110$ GeV/c^2.

We vary the outer $\Sigma_3 E_T$ boundary between 30 and 70 GeV and the $\Delta\Phi$ boundary between 2.2 and 2.8 radians. We obtain four combinations of the cuts, to which correspond four different background predictions that we compare to our best estimate. Of course, all of the four considered zones include those $(++)$ and $(+0)$ events falling in the narrowest region ($\Sigma_3 E_T < 30$ GeV, $\Delta\Phi > 2.8$ radians), and therefore the four computations of the tag probability are correlated with each other; but, since the statistics in the larger zones is more than twice as big, the correlation is not very important.

The results of these studies are listed in Table 6.2: we find that the prediction does not shift appreciably by changing either of the two cuts, the maximum difference being a positive shift of 1.8% with respect to our result. Although these evaluations are dominated by the statistical errors in the background prediction, we can elect 1.8% as our best guess of the systematic error due to the choice of the Normalization Zone.

$\Sigma_3 E_T$ boundary	$\Delta\Phi$ boundary	Prediction	Shift from (50,2.5)	Shift %
50 GeV	2.5 radians	248.49 ± 8.95	-	-
30 GeV	2.5 radians	252.93 ± 9.96	+4.44	+1.8%
70 GeV	2.5 radians	246.48 ± 8.68	-2.00	−0.8%
50 GeV	2.2 radians	244.94 ± 8.71	-3.55	−1.4%
50 GeV	2.8 radians	246.45 ± 9.78	-2.03	−0.8%

Table 6.2: *Variations of the background prediction in the mass bin* $70 \div 110\ GeV/c^2$ *with the position of the boundaries defining the Normalization Zone.*

6.6 Other Background Contaminations

Table 6.3 summarizes the expected contaminations from other electroweak processes to the muon dataset; we have neglected the light quark decays of W and Z bosons, since their contribution is expected to be undetectably small. Apart from these two processes, the double tagged data may contain other backgrounds (for instance, $t\bar{t}$ decays to dilepton pairs, where a muon triggers the event and both b-quark jets get tagged by SECVTX, may pass the selection criteria with relative ease), but their contribution is negligibly small once the two kinematic cuts are applied[21].

The $W \to c\bar{s}$ decay contributes more than the $Z \to c\bar{c}$ decay to the SECVTX sample, due to the tenfold higher cross section of that process ($\sigma_W \times B(W \to c\bar{s}) = 7.86 \pm 0.46\ nb$ as compared to $\sigma_Z \times B(Z \to c\bar{c}) = 0.906 \pm 0.088\ nb$)[22]; however, the s-quark jet is almost never tagged by SECVTX, as one should expect, with the result that the $W \to c\bar{s}$ process disappears completely in the double tagged sample; the charm decay of the Z boson is also expected to yield practically no events when

[21]We expect less than one event from $t\bar{t}$ production. Other processes that cannot be filed as "QCD background" are even smaller. In particular, $Z \to \tau^+\tau^-$ decays fail to produce two SECVTX tags, since the requirement of a tight muon implies that one of the tau leptons has decayed leptonically, and has not produced a jet in the calorimeter (the lifetime of τ leptons grants them a nonvanishing tag probability only when they decay hadronically into three or more charged tracks).

[22]We use for this estimate the PDG average[36] $\Gamma(Z \to c\bar{c})/\Gamma(Z \to h) = 0.177 \pm 0.008$ and the numbers already given in Ch. 3; as for the $W \to c\bar{s}$ decay, the quoted number has been derived by scaling up the cross section for decays to electron-neutrino pairs measured at CDF ($\sigma_W \times B(W \to e\nu) = 2.49 \pm 0.12\ nb$, run 1a[52]) using the relative widths[36] $B(W \to e\nu) = 0.1074 \pm 0.0033$, $B(W \to had) = 0.678 \pm 0.010$, and barely dividing by two to account for the ratio of $c\bar{s}$ decays over the total hadronic decays.

Sample	$W \to c\bar{s}$ MC (PYTHIA)	Events in 103 pb^{-1}	$Z \to c\bar{c}$ MC (PYTHIA)	Events in 103 pb^{-1}
Initial	500,000	$(810 \pm 79)10^3$	500,000	$(93 \pm 9)10^3$
Trigger	651	1054 ± 113	775	145 ± 15
SECVTX	76	123 ± 19	145	27 ± 3.5
2 SECVTX tags	0	0.0 ± 1.6	4	0.8 ± 0.4
Final sel.		0.0 ± 0.5		0.2 ± 0.2

Table 6.3: *Number of events selected by each of the cuts described in the text for the $Z \to c\bar{c}$ and the $W \to c\bar{s}$ simulations, and expected number of events in the total dataset. The number of events expected after the kinematic cuts ($\Sigma_3 E_T < 10$ GeV, $\Delta\Phi_{12} > 3$ radians) are estimated using the efficiency obtained from $Z \to e^+e^-$ data.*

two tags are required. To estimate the number of events in the final selection, we have assumed that the two studied background processes pass the two kinematic requirements with the same efficiency as the $Z \to b\bar{b}$ signal; this is a good first-order approximation, given that the color flux in the initial and final state of these processes is the same as that of the signal.

Chapter 7

Likelihood Fits to the Mass Spectrum

We describe in this chapter the unbinned likelihood fitting technique we have devised to interpret the dijet mass distribution of the selected dataset as the sum of signal and background contributions. After a brief introduction, we describe in Sec. 7.2 and Sec. 7.3 some preliminary fits we perform in order to identify a good parametrization of the background shape. We describe the likelihood fit in Sec. 7.4, and in Sec. 7.5 some checks of the correctness of the method. In Sec. 7.6 we evaluate the systematic uncertainties on the number of fitted signal events, and in Sec. 7.8 we describe an alternative method to fit the mass distribution. After a brief summary of the likelihood approach in Sec. 7.9, we discuss in Sec. 7.10 a toy Monte Carlo method to estimate the significance of the fitted signal.

7.1 Introduction

We have seen in the previous chapter that a simple counting experiment allows us to pinpoint a significant excess of events in our selected dataset. As we have pointed out in the discussion of those results, however, the counting experiment underestimates by construction the size of the signal, because of the contamination of $Z \to b\bar{b}$ events in the samples used to determine the background normalization: that contamination prevents us from using that result for a cross section measurement[1]. The easiest way out is to look for the signal directly in the dijet mass spectrum, by means of an unbinned likelihood fit of its shape to the sum of its two components: the Z signal and the QCD background. We will then be able to use the number of fitted events for the computation of the cross section for $Z \to b\bar{b}$ production at the Tevatron.

As far as the signal shape is concerned, we have shown in Ch. 5 that after the application of our improved jet corrections the $Z \to b\bar{b}$ signal should give a Gaussian peak at 90 GeV/c^2, with a width of 12.3 GeV/c^2. We therefore assume a

[1] As a matter of fact, the biases in the counting experiment—due to the amount of signal present both in the (+0) sample and in the Normalization Zone—could be accounted for with an iterative procedure. The large uncertainties on the selection efficiencies, however, get too amplified and make the result useless.

Gaussian shape, but allow its parameters to float in the fit, thus avoiding the typical systematic uncertainties related to the Monte Carlo description of the signal, the detector simulation, *et similia*. The background shape is instead determined from an enriched background sample: the procedure relies on some assumptions, whose systematic effects will be taken into account. An alternative fit technique will also be used to check the robustness of the results.

We recall here some of the definitions and notations already defined in the previous chapter, since we will make wide use of them in the following. By "(+0)" we indicate the sample with one SECVTX tagged jet and one taggable jet, as far as the two leading jets are considered; similarly, "(++)" indicates the sample with two SECVTX tagged jets. We define *Signal Zone* and *Normalization Zone* as the subsets defined by the following requirements:

- Signal Zone:
$$\Delta\Phi > 3\, radians \qquad\qquad \Sigma_3 E_T < 10\, GeV$$
- Normalization Zone: *not* in Signal Zone and:
$$\Delta\Phi > 2.5\, radians \qquad\qquad \Sigma_3 E_T < 50\, GeV$$

7.2 The Background Shape

We will consider the (+0) data in the Normalization Zone as our background sample. We first fit the mass distribution of those events by using an unbinned likelihood fitting technique with the following definition:

$$\mathcal{L}^{(+0)}_{NormZone} = Poisson(N^{(+0)}_{NormZone},\, n^{(+0)}_{NormZone}) \cdot \prod_{i=1}^{N^{(+0)}_{NormZone}} P_{bgr}(m_i) \qquad (7.1)$$

where:

- $Poisson(x,\lambda) \equiv \frac{e^{-\lambda}.\lambda^x}{x!}$ is the Poisson constraint of x around the expected value λ;
- $N^{(+0)}_{NormZone}$ is the observed number of (+0) events in the Normalization Zone;
- $n^{(+0)}_{NormZone}$ is the fitted number of (+0) events in the Normalization Zone;
- $P_{bgr}(m)$ is the mass shape distribution of the background events, normalized to unity. We use the following parametrization with three free parameters in the fit:
$$P_{bgr}(m) = \frac{1}{\lambda} \cdot e^{-m/\lambda} \otimes N(\mu,\sigma) \qquad (7.2)$$
where "$\otimes$" is the convolution operation.
$N(\mu,\sigma)$ is, of course, a Gaussian distribution.

The above parametrization of the background shape has been chosen because it describes very well the mass shape. In the fit we ignore any $Z \rightarrow b\bar{b}$ contamination of the $(+0)$ sample in the Normalization Zone; we will check this hypothesis later on. The fit results, shown in Fig. 7.1, are:

$$\begin{aligned} n^{(+0)}_{NormZone} &= 28,053 \pm 167 \quad (N^{(+0)}_{NormZone} = 28,053) \\ \sigma &= 9.72 \pm 0.14 \, GeV/c^2 \\ \lambda &= 32.96 \pm 0.31 \, GeV/c^2 \\ \mu &= 51.69 \pm 0.18 \, GeV/c^2 \end{aligned}$$

For reasons that will appear clear later, we report here also the covariance matrix of the three background shape parameters (in the order σ, λ, μ):

$$\begin{pmatrix} 0.0200 & -0.0196 & 0.0160 \\ -0.0196 & 0.0949 & -0.0357 \\ 0.0160 & -0.0357 & 0.0323 \end{pmatrix} \tag{7.3}$$

7.3 Parametrization of the Tag Probability Curve

To fit the signal in the mass spectrum of $(++)$ events in the Signal Zone we need to know the background mass shape for this sample. We cannot use directly the background shape $P_{bgr}(m)$ determined above, because it has been extracted from single tagged events, and—as we have already discussed in Ch. 6—the tag probability is not flat as a function of the dijet mass. However, if we determine the shape of the tag probability $P_{tag}(m)$, we can use for the $(++)$ sample the following background shape:

$$P'_{bkg}(m) = \frac{P_{bkg}(m) \cdot P_{tag}(m)}{\int P_{bkg}(\xi) \cdot P_{tag}(\xi) \cdot d\xi} \tag{7.4}$$

where the denominator takes care of the normalization.

To determine the curve of tag probability we compute the ratio between double tagged and single tagged events in the Normalization Zone, and then fit the distribution with the following functional form:

$$P_{tag}(m) = \left\{ \begin{array}{ll} p_1 \cdot Freq(\frac{m-p_2}{p_3}) & \text{if } m < p_5 \\ p_1 \cdot Freq(\frac{m-p_2}{p_3}) \cdot e^{-\frac{(m-p_5)}{p_4}} & \text{if } m > p_5 \end{array} \right\} \tag{7.5}$$

where $Freq$ is the well-known CERN library function[50]. The five free parameters returned by the fit are:

$$\begin{aligned} p_1 &= 0.115 \pm 0.007 \\ p_2 &= 39.62 \pm 3.32 \, GeV/c^2 \\ p_3 &= 24.50 \pm 6.55 \, GeV/c^2 \\ p_4 &= 167.66 \pm 36.87 \, GeV/c^2 \\ p_5 &= 75.00 \pm 1.06 \, GeV/c^2 \end{aligned}$$

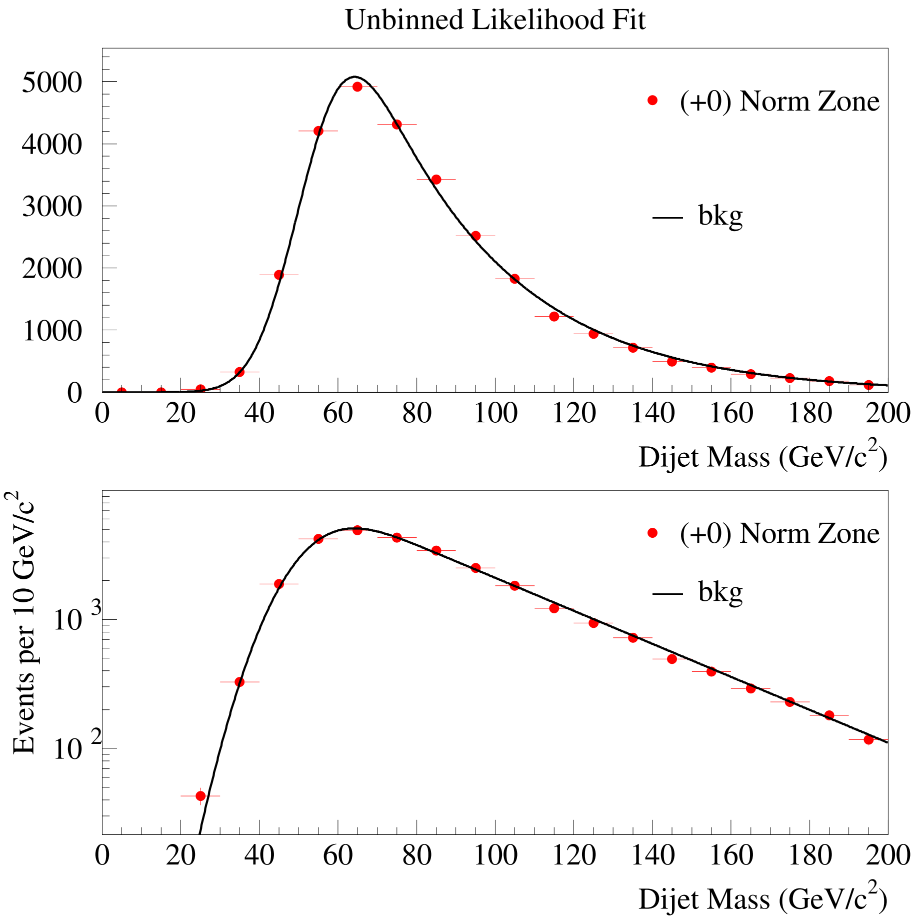

Figure 7.1: *Unbinned likelihood fit of the dijet mass distribution in the (+0) sample in the Normalization Zone, in linear scale (upper plot) and in logarithmic scale (lower plot). Here we are fitting for the background shape, by assuming no signal contamination.*

The χ^2/ndf of the fit (Fig. 7.2) is 0.48. For reasons that will appear clear later, we report here also the covariance matrix of the five parameters:

$$\begin{pmatrix} 0.0000539 & 0.00764 & 0.0364 & -0.222 & -0.000802 \\ 0.00764 & 11.0 & -5.78 & -35.1 & -0.190 \\ 0.0364 & -5.78 & 42.9 & -142.0 & -0.373 \\ -0.222 & -35.1 & -142.0 & 1360 & -0.0398 \\ -0.000802 & -0.190 & -0.373 & -0.0398 & 1.11 \end{pmatrix} \tag{7.6}$$

The reader may notice that, as is clear in Eq. 7.4, what really matters here is the *shape* of the tag probability curve, not its absolute value. For this reason, although we have to assume that the tag probability of the background in the Normalization Zone is the same as the one in the Signal Zone, the correlation with the counting experiment is small, the latter being strongly dependent on the *absolute* value of the

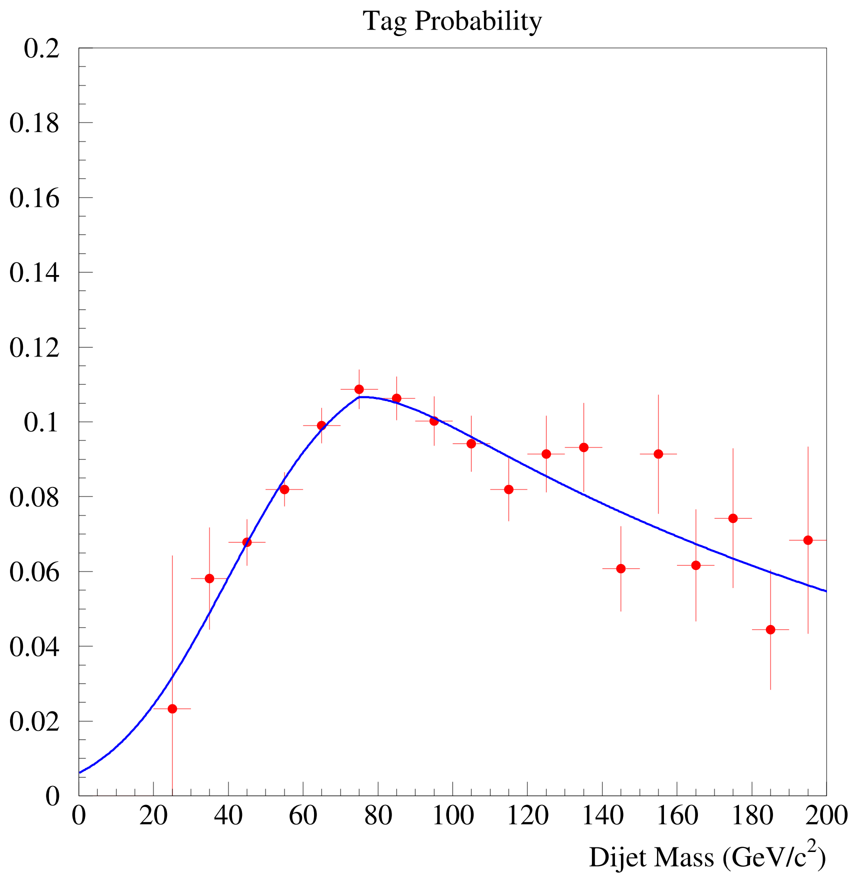

Figure 7.2: χ^2 *fit of the double tag probability as a function of the dijet mass. The probability is calculated, bin by bin, as the ratio of double tagged events (++) over single tagged events (+0), in the Normalization Zone.*

tag probability in each bin, and much less on its overall shape.

7.4 The Likelihood Fit of the Signal Sample

After having dealt with the fits described above, we are ready to perform a two-component unbinned likelihood fit of the dijet mass distribution of the (++) sample in the Signal Zone. We define the likelihood as follows:

$$\mathcal{L}^{(++)}_{SignalZone} = Poisson(\, N^{(++)}_{SignalZone},\, (n^{(++)}_{sig,\,SignalZone} + n^{(++)}_{bgr,\,SignalZone})\,)\times$$
$$\prod_{i=1}^{N^{(++)}_{SignalZone}} \frac{n^{(++)}_{sig,\,SignalZone}\cdot P'_{sig}(m_i) + n^{(++)}_{bgr,\,SignalZone}\cdot P'_{bgr}(m_i)}{n^{(++)}_{sig,\,SignalZone} + n^{(++)}_{bgr,\,SignalZone}} \tag{7.7}$$

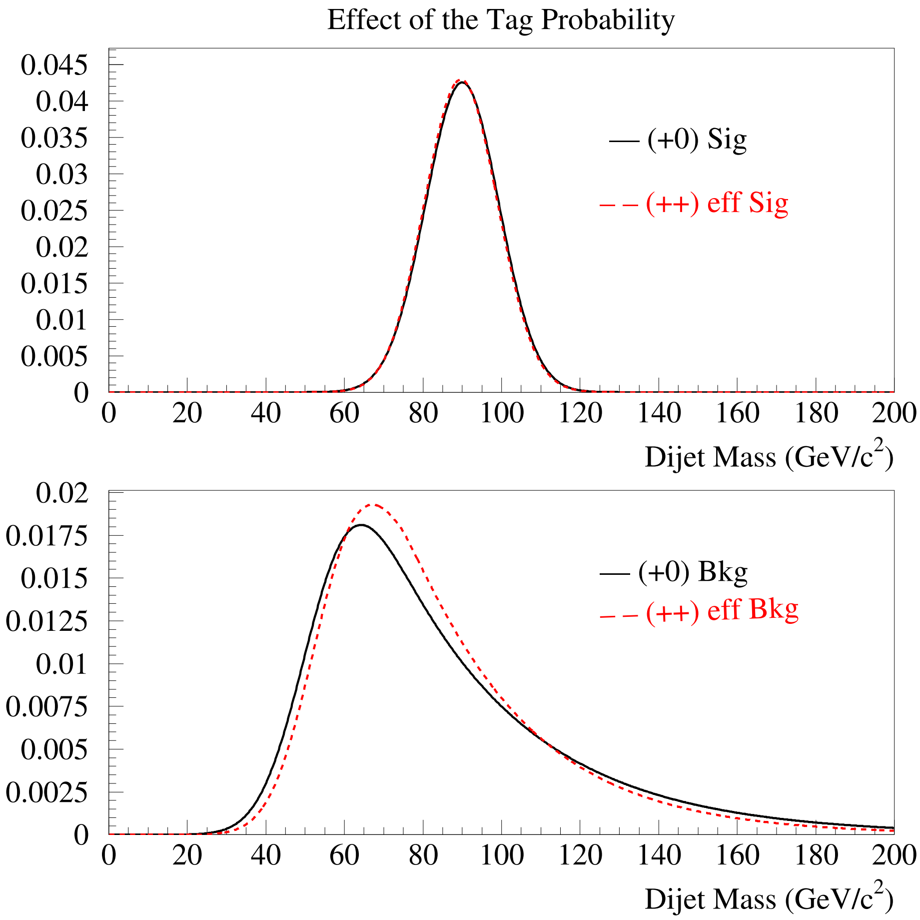

Figure 7.3: *The signal shape (upper plot) and the background shape (lower plot) are shown before (full line) and after (dashed line) the effect of the tag probability. Notice that the signal shape remains almost unchanged, whereas the background shape is eroded in the low and high mass regions, where the tag probability is smaller.*

where:

- $N^{(++)}_{Signal\,Zone}$ is the observed number of events in the (++) sample in the Signal Zone;
- $n^{(++)}_{sig,\,Signal\,Zone}$, $n^{(++)}_{bgr,\,Signal\,Zone}$ are respectively the number of signal and background events in the (++) sample in the Signal Zone;
- $P'_{bgr}(m)$ is the normalized background mass shape of the events in the (++) sample. This shape is determined from the background shape of the (+0) sample, by using the tag probability curve as shown in Eq. 7.4; the effect of the tag probability curve on the background shape is shown in the lower plot of Fig. 7.3;
- $P'_{sig}(m)$ is the normalized signal mass shape of the events in the (++) sample.

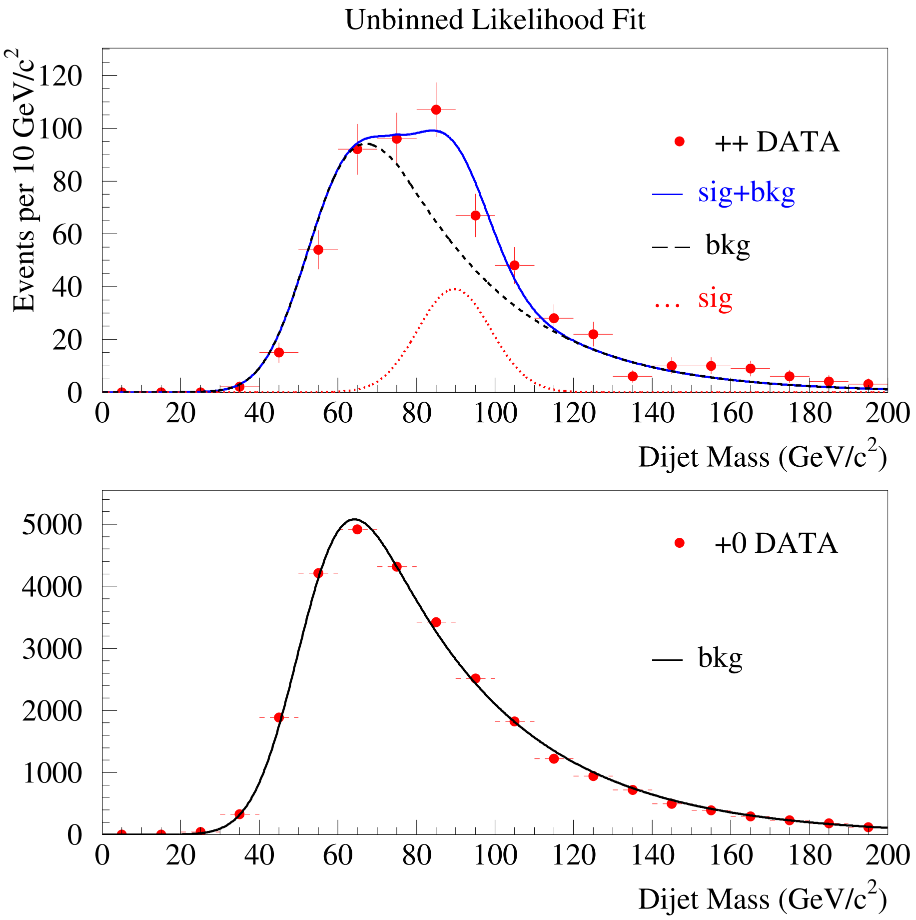

Figure 7.4: *Unbinned likelihood fit of the dijet mass distribution in the (++) sample in the Signal Zone (upper plot). The background shape in this fit has been determined by fitting the mass distribution in the (+0) sample in the Normalization Zone (lower plot), and then taking into account the effect of the double tag probability curve.*

We assume, as the Monte Carlo suggests, a Gaussian distribution $N(M_Z, \sigma_Z)$ as signal shape in the (+0) sample, $P_{sig}(m)$, and we apply to it the tag probability curve[2] as in Eq. 7.4 to get the signal distribution in the (++) sample. Notice, however, that for typical Z mass resolutions of the order of 10 GeV/c^2, the probability curve is almost flat in the mass region of interest for the Z, and therefore the effect of the tag probability on the signal shape is small. Therefore $P'_{sig}(m)$ is still almost perfectly describable as a Gaussian distribution. This is shown in the upper plot of Fig. 7.3.

By minimizing the $-\log \mathcal{L}$ with respect to the number of signal and background events, and to the mass and width of the signal shape, but keeping frozen the

[2]We point out that the nice Gaussian shapes obtained in Ch. 5 for the $Z \to b\bar{b}$ decay signal are in principle relevant only to signal events unbiased by a tag requirement: in fact, we did not use SECVTX as a selection tool for the samples used to study the jet energy.

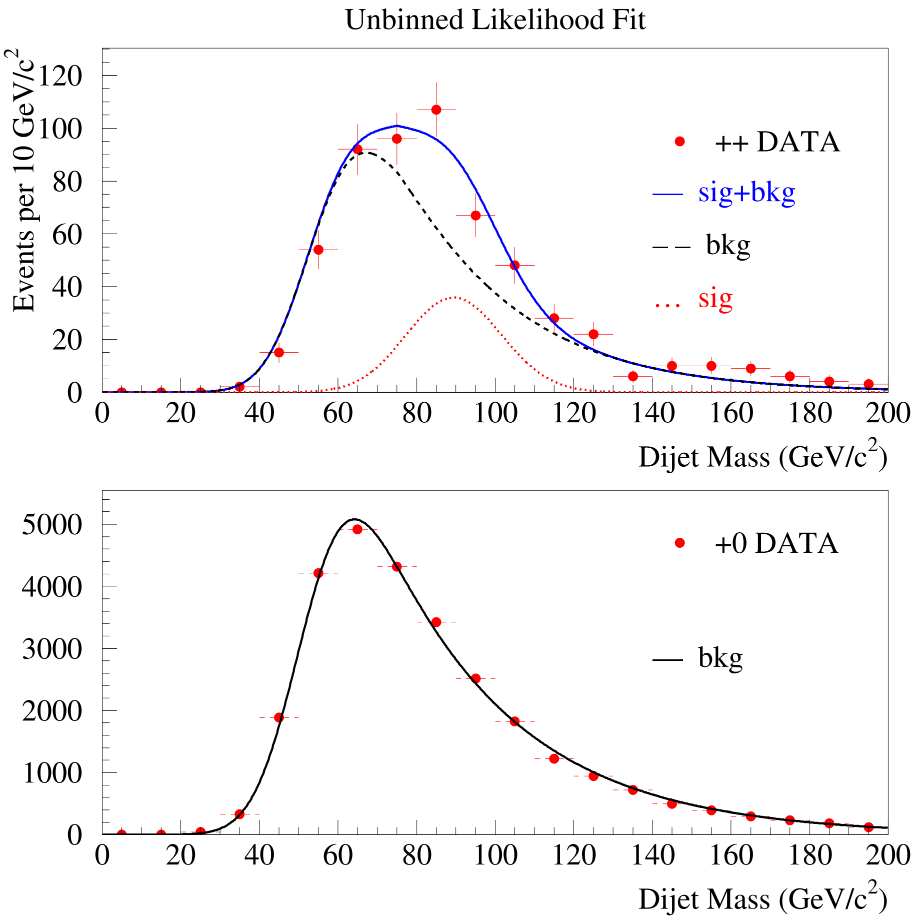

Figure 7.5: *Unbinned likelihood fit of the dijet mass distribution in the (++) sample in the Signal Zone (upper plot), with both signal and background shapes fixed. The background shape was determined by fitting the mass distribution in the (+0) sample in the Normalization Zone (lower plot). The mean* $M_Z = 90\ GeV/c^2$, *and the sigma* $\sigma_Z = 12.3\ GeV/c^2$ *of the signal shape (Gaussian) correspond the PYTHIA Monte Carlo expectation.*

background shape parameters and the tag probability curve parameters, we obtain[3]:

$$\begin{aligned} \mathbf{n^{(++)}_{sig,\ SignalZone}} &= \mathbf{91 \pm 30} \\ n^{(++)}_{bkg,\ SignalZone} &= 488 \pm 36 \\ M_Z &= 90.0 \pm 2.4\ GeV/c^2 \\ \sigma_Z &= 9.4 \pm 3.5\ GeV/c^2 \end{aligned} \tag{7.8}$$

The fit results are shown in the upper plot of Fig. 7.4. In the lower plot is shown the fit of the (+0) sample in the Normalization Zone, from which we determined the

[3]The fit is performed in the region $0 \div 200\ GeV/c^2$; 579 of the 588 double tagged events of the Signal Zone fall in this mass interval.

background shape. Notice that the fit is unbinned, and therefore the bin choice in these plots does not affect the results at all. These are in very satisfactory agreement with the Monte Carlo predictions for the mass and width of the signal[4].

7.5 Checks of the Fitting Procedure

First of all, let us show what is the result of the fit of the (++) sample in the Signal Zone if we fix the signal shape as well as the background. This means fixing the parameters M_Z and σ_Z of the Gaussian (see Eq. 7.7) to the values observed in the Monte Carlo, and allowing the fit to determine only the number of signal and background events:

$$\begin{aligned}
\mathbf{n}^{(++)}_{\mathbf{sig,\,SignalZone}} &= \mathbf{109 \pm 24} \\
n^{(++)}_{bkg,\,SignalZone} &= 470 \pm 30 \\
M_Z &= 90.0\ GeV/c^2\ \text{, fixed} \\
\sigma_Z &= 12.3\ GeV/c^2\ \text{, fixed}
\end{aligned} \tag{7.9}$$

This fit, shown in Fig.7.5, corresponds to the "usual" procedure one adopts when one wishes to obtain a mass measurement from a fit using Monte Carlo templates. In the present instance, we already know the value of the true mass of the searched signal, hence we need only one template, which turns out to be well described by a Gaussian. Although the number of events obtained in the fit is in agreement with the previous result, we prefer to allow to float both M_Z and σ_Z rather than fixing them as above, in order to avoid any systematics related to the Monte Carlo description of the signal.

To determine the background shape, as we have seen, we fit the (+0) sample in the Normalization Zone by assuming no $Z \to b\bar{b}$ signal contamination is present there. Monte Carlo simulations predict as a matter of fact a small contamination, of the order of three hundred events, out of more than 28,000 events with a single SECVTX tag populating the Normalization Zone. It is worth checking both if the fit is sensitive to such a small contamination and—if it is the case—what is the effect on the fit of the (++) sample in the Signal Zone. To do that, we perform three different fits:

a) We fit as usual the (+0) sample in the Normalization Zone, without signal, as usual; then, after the fit has converged, we fix the background shape parameters, but allow the signal shape and normalization in the (+0) (Normalization Zone) to float. Of course, in order to do that, we change the likelihood fit (Eq. 7.1), to add the signal contribution. The likelihood looks like the one for the (++) sample in the Signal Zone (see Eq. 7.7), with the only important

[4]The expected value of M_Z from PYTHIA is 90 GeV/c^2, while the width is expected to be 12.3 GeV/c^2. Instead, HERWIG predicts respectively 90.7 and 11.7 GeV/c^2 for these quantities.

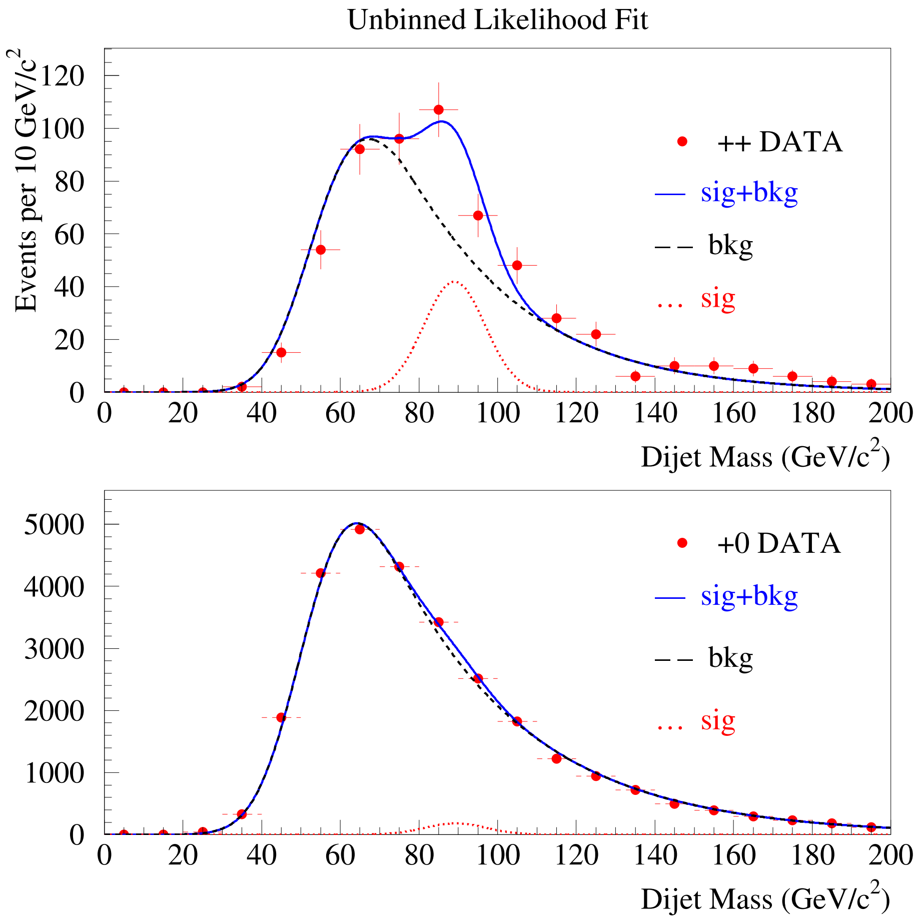

Figure 7.6: *Simultaneous unbinned likelihood fit of the dijet mass distribution of the (++) sample in the Signal Zone (upper plot) and of the (+0) sample in the Normalization Zone (lower plot). The background shape is fixed in the fit, and it has been determined separately by fitting the (+0) sample (Normalization Zone) without signal contamination.*

difference that the tag probability curve is *not* applied in this case.
The fit gives the following results:

$$\begin{aligned} n^{(+0)}_{sig,\, NormZone} &= 347 \pm 128 \\ n^{(+0)}_{bgr,\, NormZone} &= 27,706 \pm 209 \\ M_Z &= 88.7 \pm 2.6\ GeV/c^2 \\ \sigma_Z &= 7.3 \pm 2.4\ GeV/c^2 \end{aligned} \tag{7.10}$$

It is worth noting that we do not fit the (+0) sample in the Normalization Zone with both background and signal shapes free to float at the same time, because the statistical fluctuations of that sample are comparable with the expected signal contamination, hence the Gaussian shape is driven from the former effect rather than from the latter.

b) We fit the $(+0)$ sample in the Normalization Zone, without signal, as usual; then, after the fit has converged, we fix the background shape parameters, as in the previous case; now, however, we fit simultaneously for the signal both the $(+0)$ sample in the Normalization Zone and the $(++)$ sample in the Signal Zone.
The results, shown in Fig. 7.6, are:

$$\begin{aligned} n^{(+0)}_{sig,\,NormZone} &= 358 \pm 120 \\ n^{(+0)}_{bgr,\,NormZone} &= 27,695 \pm 204 \\ \mathbf{n^{(++)}_{sig,\,SignalZone}} &= \mathbf{82 \pm 23} \\ n^{(++)}_{bgr,\,SignalZone} &= 497 \pm 31 \\ M_Z &= 89.2 \pm 1.7\ GeV/c^2 \\ \sigma_Z &= 7.8 \pm 2.1\ GeV/c^2 \end{aligned} \tag{7.11}$$

Remarkably, the number of Z events is beautifully consistent with expectations in both these tests, and so is the mass value. The width of the signal comes out instead too small by about 2σ.

c) we fit the background shape from the $(+0)$ sample in the Normalization Zone, but we *exclude from the fit the mass window* $75 \div 105\ GeV/c^2$. Then we fit only the $(++)$ sample in the Signal Zone, with background shape fixed. The results are:

$$\begin{aligned} \mathbf{n^{(++)}_{sig,\,SignalZone}} &= \mathbf{98 \pm 30} \\ n^{(++)}_{bgr,\,SignalZone} &= 481 \pm 36 \\ M_Z &= 89.9 \pm 2.4\ GeV/c^2 \\ \sigma_Z &= 9.5 \pm 3.4\ GeV/c^2 \end{aligned} \tag{7.12}$$

The conclusions we can draw from the above results are that the fit is certainly able to extract the small contamination of signal in the $(+0)$ sample in the Normalization Zone, and the effect of this contamination in the number of signal events in the $(++)$ sample in the Signal Zone, $n^{(++)}_{sig,\,SignalZone}$, is well below the statistical uncertainty. However, if we want to quote a number for these systematics on the number determined in Eq. 7.8, we can conservatively take the largest difference between the latter and the above numbers:

$$(82 - 91) = -9 \qquad (98 - 91) = +7 \tag{7.13}$$

7.6 Other Systematics

In the fit (Eq. 7.7) of the $(++)$ sample in the Signal Zone, both the background shape and the double tag probability curve are fixed. Since these two functions were determined with independent fits, we need to study how their uncertainties may affect the main results of the fit, *id est* $n^{(++)}_{sig,\,SignalZone}$. To do that we have to exploit our determination of the covariance matrix, obtained from the fit of the curve under consideration (that is, Eq. 7.3 for the background shape, and Eq. 7.6 for the tag probability), because the covariance matrix contains the information on the errors of each parameter and the correlation between them. Let us explain the method in some detail for the double tag probability curve: for the background shape the same procedure is applied.

First, the fit of the (++) sample in the Signal Zone is done the usual way, that is, with free signal shape and fixed background shape. Then, the parameters that describe the tag probability curve are allowed to float in the fit, within their multivariate Gaussian constraints, and the fit is repeated again. The new fitted $n^{(++)}_{sig,\,SignalZone}$ will have a larger fit error than before: the quadratic difference of those fit errors can be taken as the systematics on $n^{(++)}_{sig,\,SignalZone}$ due to the uncertainty on our knowledge of the tag probability curve, as pictured by its covariance matrix.

Before reporting the results, we write explicitly the multivariate Gaussian constraint:

$$\begin{aligned} -\log \mathcal{L}^{(++)}_{SignalZone} \;\; &\to \;\; -\log \mathcal{L}^{(++)}_{SignalZone} + \\ &+\frac{1}{2}\cdot(\,\vec{\tilde{\mathbf{p}}}-\vec{\mathbf{p}}\,)^T\cdot \mathbf{COV}^{-1}\cdot(\,\vec{\tilde{\mathbf{p}}}-\vec{\mathbf{p}}\,) \end{aligned} \tag{7.14}$$

where:

- $\mathcal{L}^{(++)}_{SignalZone}$ is the usual likelihood, see Eq. 7.7;
- $\vec{\tilde{\mathbf{p}}}$ are the five free parameters which described the tag probability curve (see Eq. 7.5) and $\vec{\mathbf{p}}$ are their "central" values (Eq. 7.6);
- **COV** is the 5×5 covariance matrix of the double tag probability fit, reported in Eq. 7.6.

Here are the results, to be compared with Eq. 7.8:

1) Double tag probability curve:

$$\begin{aligned} \mathbf{n}^{(++)}_{\mathbf{sig,\,SignalZone}} &= \mathbf{80 \pm 34} \\ n^{(++)}_{bkg,\,SignalZone} &= 499 \pm 40 \\ M_Z &= 89.5 \pm 2.6\ GeV/c^2 \\ \sigma_Z &= 8.2 \pm 4.4\ GeV/c^2 \\ \\ \tilde{p}_1 &= 0.115 \pm 0.006 \\ \tilde{p}_2 &= 42.68 \pm 3.67\ GeV/c^2 \\ \tilde{p}_3 &= 22.60 \pm 5.20\ GeV/c^2 \\ \tilde{p}_4 &= 190.61 \pm 26.43\ GeV/c^2 \\ \tilde{p}_5 &= 74.95 \pm 1.06\ GeV/c^2 \end{aligned} \tag{7.15}$$

The systematic on $n^{(++)}_{sig,\ SignalZone}$ due to the uncertainty of the double tag probability curve can then be estimated by:

$$\sqrt{34^2 - 30^2} = 16. \tag{7.16}$$

2) Background shape:

$$\begin{aligned}
\mathbf{n}^{(++)}_{\mathbf{sig,\ SignalZone}} &= \mathbf{91 \pm 30} \\
n^{(++)}_{bkg,\ SignalZone} &= 488 \pm 36 \\
M_Z &= 89.9 \pm 2.4\ GeV/c^2 \\
\sigma_Z &= 9.3 \pm 3.6\ GeV/c^2 \\
\\
\sigma &= 9.71 \pm 0.14\ GeV/c^2 \\
\lambda &= 33.1 \pm 0.30\ GeV/c^2 \\
\mu &= 51.69 \pm 0.18\ GeV/c^2
\end{aligned} \tag{7.17}$$

The systematic on $n^{(++)}_{sig,\ SignalZone}$ due to the uncertainty of the background shape is given by the quadratic difference between the returned error on $n^{(++)}_{sig,\ SignalZone}$, and the one in the "central fit" (Eq. 7.8), where we now need to appeal, in order to appreciate a difference between them, to the first decimal figure of those numbers:

$$\sqrt{29.9^2 - 29.5^2} = 4.9. \tag{7.18}$$

The systematic error is small due to the fact that the background shape is precisely determined by virtue of the very high statistics of the (+0) sample in the Normalization Zone.

Let us conclude with a remark on another possible systematic effect. Potentially, this could be due to the assumption of the same shape of the background in the Normalization and Signal Zones. However, from the studies described in Sec. 6.5, we know that the relative systematic error due to this bias cannot be larger than 3%. Besides being tiny, in comparison to the statistical and the other systematic errors, we believe that it is already included in the large systematic due to the tag probability curve. In fact, although very different in nature, both effects change the background shape, and the above variation of the tag probability curve corresponds to more than a "one sigma" systematic. This is due to the fact that the variation of the likelihood function between the "central fit", Eq. 7.8, and the fit with the tag probability parameters free, Eq. 7.15, is:

$$-2\cdot\log\mathcal{L}^{(++)}_{SignalZone} + 2\cdot\log\mathcal{L}^{(++)}_{SignalZone}(\text{tag prob. floating}) = 5400 - 5397 = 3. \tag{7.19}$$

This corresponds to a $\sqrt{3} \approx 1.7$ sigma variation. In other words, we have reason to believe that the quoted systematic error on the tag probability is conservative, and therefore we can safely assume that it includes the much smaller possible bias in the extrapolation of the background shape from the Normalization Zone to the Signal Zone.

7.7 Significance of the Signal

The correct way to estimate the statistical significance of the fitted number of signal events, $n^{(++)}_{sig,\, Signal Zone}$, in the fit (Eq. 7.8), defined as the probability that a background fluctuation fakes a signal contribution in the mass distribution of (++) events in the Signal Zone equal to or higher than the one we found, is to perform a large number of pseudo-experiments with a toy Monte Carlo technique. The idea is to generate mass values for (++) events using only the background shape, and then fit the sample exactly as we do on the real data, *id est* with a free signal shape. The significance is then given by the fraction of pseudo-experiments where the fitted number of signal events, $n^{(++)}_{sig,\, Signal Zone}$, is greater than 91, the number that we get from the real data. With this procedure, one can take into accont the various systematics—and the statistical fluctuation of the sample size—by extracting the number of events from a Poisson distribution around the observed number of real events, and the systematics of the background shape and the double tag probability curve, by drawing the parameters which describe those shapes from multivariate Gaussian distributions, where the covariance matrices are those reported in Eq. 7.3.

The toy Monte Carlo approach is described in Sec. 7.10.2. Here we wish to use an alternative approach, much faster although rough and without systematics: it can only give us an upper bound on the significance of the shape of the mass distribution for our signal sample. We fit the (++) sample in the Signal Zone using only the background shape: this means to use the likelihood of Eq. 7.7 with $n^{(++)}_{sig,\, Signal Zone}$ fixed to zero. We can then compare the likelihood value thus obtained with the one obtained before, when the signal was allowed to float in the fit:

$$\begin{aligned} \text{Significance} &= \sqrt{-2 \cdot \log \mathcal{L}^{(++)}_{Signal Zone}(\text{only bkg}) + 2 \cdot \log \mathcal{L}^{(++)}_{Signal Zone}} \\ &= \sqrt{5416 - 5400} \simeq 4. \end{aligned} \tag{7.20}$$

Almost identical values are obtained if we perform a simultaneous fit of (+0) events in the Normalization Zone and (++) events in Signal Zone, again without signal, and with a free background shape in the fit. The result is shown in Fig. 7.7.

7.8 An Alternative Approach

A more straightforward and simple method to fit the signal in the (++) sample in the Signal Zone is to the consider the (++) sample in the Normalization Zone as our background-enriched sample, instead of the (+0) sample in the Normalization Zone as we have done so far. The advantage is twofold: first, we do not need any more the double tag probability curve; second, by fitting simultaneously the two samples, (++) in Signal and Normalization Zones, we may allow the background shape to float as well: the result will take naturally into account the uncertainty of the background shape. Unfortunately, there is a much lower statistics of the (++) sample in the Normalization Zone with respect to the one we had by using the (+0) sample in the Normalization Zone; besides, the relative signal contamination is much higher

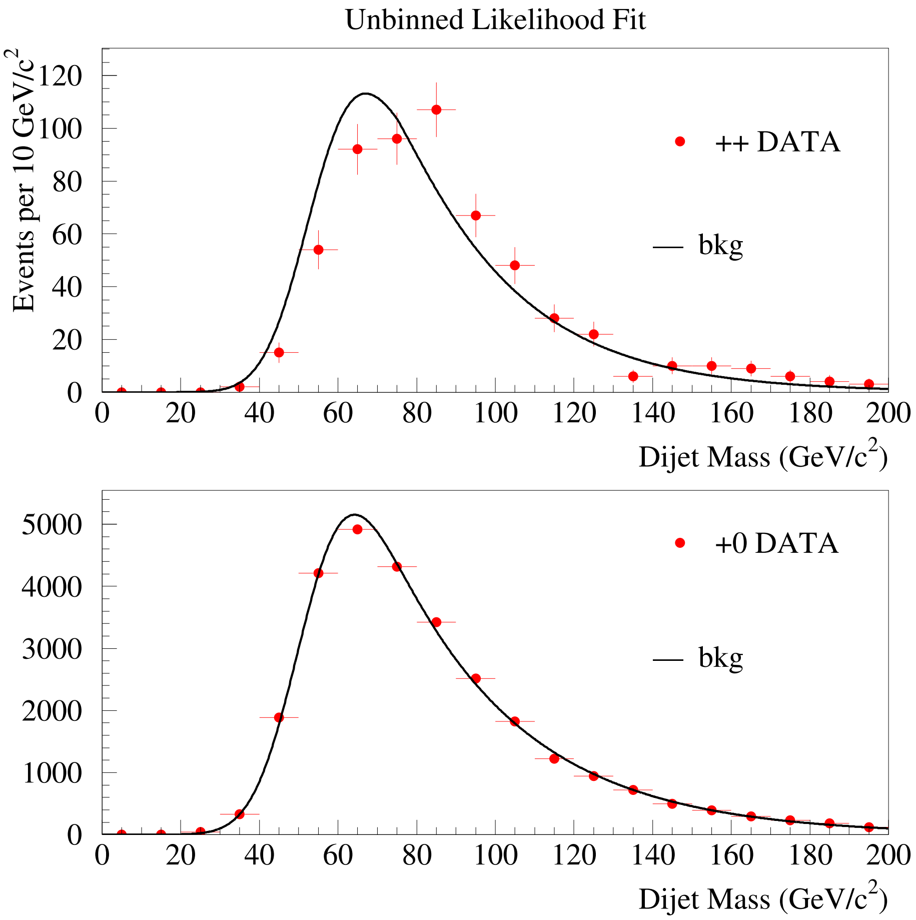

Figure 7.7: *Simultaneous unbinned likelihood fit of the dijet mass distribution of the (++) sample in the Signal Zone (upper plot) and the (+0) sample in the Normalization Zone (lower plot), when no allowance is made for a contribution from the $Z \to b\bar{b}$ signal.*

in the former than in the latter, and this could in principle get the fitter into trouble when it tries to separate the signal from the background. It turns out that this is not the case, and the results are very close to those reported in Eq. 7.8. Of course, it appears clear that the two results are strongly correlated to each other; however, the two fitting techniques are quite different: that allows us to conclude that our result is robust, being relatively unaffected by the fitting procedure.

The likelihood can be easily obtained from what has been said in the previous sections. However, for the convenience of the reader, we write explicitly below all its pieces again:

$$\mathcal{L}^{(++)} = \mathcal{L}^{(++)}_{SignalZone} \times \mathcal{L}^{(++)}_{NormZone}$$
$$\mathcal{L}^{(++)}_{SignalZone} = Poisson(\, N^{(++)}_{SignalZone},\, (n^{(++)}_{sig,\, SignalZone} + n^{(++)}_{bgr,\, SignalZone})\,)\times$$

$$\prod_{i=1}^{N^{(++)}_{SignalZone}} \frac{n^{(++)}_{sig,\,SignalZone} \cdot P_{sig}(m_i) + n^{(++)}_{bgr,\,SignalZone} \cdot P_{bgr}(m_i)}{n^{(++)}_{sig,\,SignalZone} + n^{(++)}_{bgr,\,SignalZone}}$$

$$\mathcal{L}^{(++)}_{NormZone} = Poisson(\, N^{(++)}_{NormZone},\, (n^{(++)}_{sig,\,NormZone} + n^{(++)}_{bgr,\,NormZone})\,)\times$$

$$\prod_{i=1}^{N^{(++)}_{NormZone}} \frac{n^{(++)}_{sig,\,NormZone} \cdot P_{sig}(m_i) + n^{(++)}_{bgr,\,NormZone} \cdot P_{bgr}(m_i)}{n^{(++)}_{sig,\,NormZone} + n^{(++)}_{bgr,\,NormZone}} \tag{7.21}$$

where for $P_{sig}(m)$ the Gaussian distribution $N(M_Z, \sigma_Z)$ can be taken, and $P_{bgr}(m)$ is the function written in Eq. 7.2, with three free parameters. Notice that we use now the same functional form we used before for the single tagged events, but we allow the shape parameters to float in the fit, and then adjust for the effect of the second SECVTX tag requirement.

The results of this fit, shown in Fig. 7.8, are:

$$\begin{aligned}
\mathbf{n^{(++)}_{sig,\,SignalZone}} &= \mathbf{96 \pm 31} \\
n^{(++)}_{bgr,\,SignalZone} &= 483 \pm 37 \\
n^{(++)}_{sig,\,NormZone} &= 70 \pm 66 \\
n^{(++)}_{bgr,\,NormZone} &= 2537 \pm 82 \\
M_Z &= 88.6 \pm 2.3\ GeV/c^2 \\
\sigma_Z &= 9.2 \pm 2.5\ GeV/c^2 \\
\\
\sigma &= 9.62 \pm 0.49\ GeV/c^2 \\
\lambda &= 29.6 \pm 1.06\ GeV/c^2 \\
\mu &= 54.89 \pm 0.82\ GeV/c^2
\end{aligned} \tag{7.22}$$

Notice that the returned fit error on $n^{(++)}_{sig,\,SignalZone}$ (that is, the statistical error) is now larger than in the "central fit" (Eq. 7.8): the quadratic difference of the two values, $\sqrt{31^2 - 30^2} \approx 8$ can be seen as the price in statistical uncertainty one must pay if the (++) sample in the Normalization Zone is used as a background enriched sample rather than the (+0) sample in the Normalization Zone. *A posteriori*, this "statistical penalty" turns out to be milder than the "systematic penalties" we paid with our main approach, although in the second approach we have neglected the small systematic of the background shape extrapolation from the Normalization Zone to the Signal Zone. We therefore leave open the possibility, in the near future, to adopt this second approach as the "central" one.

7.9 Summary of the Likelihood Fit Approach

In Table 7.1 we collect the systematics of our fit result on $n^{(++)}_{sig,\,SignalZone}$:
As a final number we quote:

$$\mathbf{n^{(++)}_{sig,\,SignalZone} = 91\ \pm\ 30\,(stat.)\ \pm\ 19\,(syst.)} \tag{7.23}$$

where the systematic uncertainties are added in quadrature.

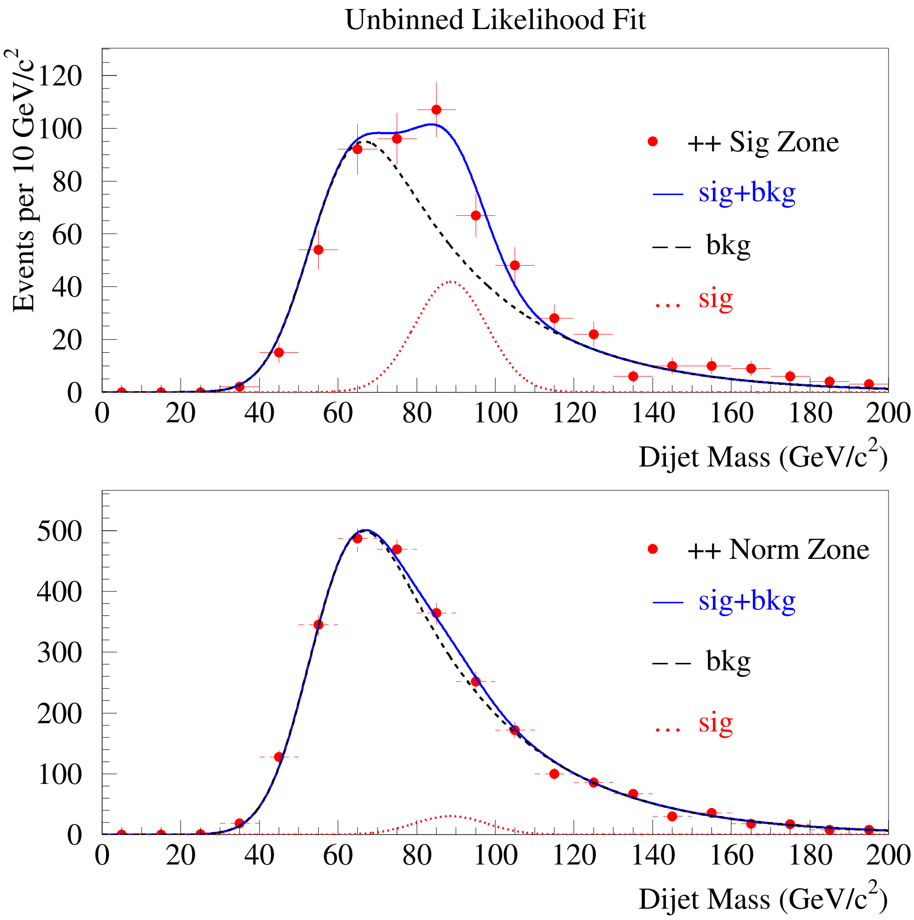

Figure 7.8: *Simultaneous unbinned likelihood fit of the dijet mass distribution of the (++) sample in the Signal Zone (upper plot) and of the (++) sample in the Normalization Zone (lower plot). Signal and Background shapes are allowed to float in the fit.*

Source	Systematic error
Sig. contamination	± 9 events
Tag probability curve	± 16 events
Bgr. shape	± 5 events

Table 7.1: *List of the main systematic uncertainties in the estimation of the number of $Z \to b\bar{b}$ events obtained with the unbinned likelihood fit.*

7.10 A Better Evaluation of the Significance

The evaluation of the significance of the discrepancy between expected and observed mass distributions in the double tagged data presented in Sec. 7.6 can only be regarded as a first-order estimate, since it does not take into account any systematic uncertainty in the fitting procedure (due to our arbitrary parametrization of the tag probability curve and of the background shape); moreover, the extraction of a number of standard deviations as $N_\sigma = \sqrt{(2 \times \Delta(log\mathcal{L}))}$ relies on a Gaussian approximation that may not be valid. To obtain a better estimate of the significance, we have already noted that a toy Monte Carlo approach is mandatory: specifically, we need to generate a large number of mass distributions according to the expected background shape, and fit them as the sum of signal and background, *id est* pretending that they contain a Gaussian signal contamination. The number of returned signal events should on average be zero, in which case the frequency by which the fit returns a number of events greater or equal to 91 tells us about the real significance of our result on the experimental data.

Before we put this project into practice, we need to check the compatibility between the toy Monte Carlo and the likelihood fitting procedure. This will allow us to verify the absence of any biases in the fit itself, and the correctness of our toy generator. These checks are briefly described in the following.

7.10.1 Sanity Checks of the Fitting Procedure

We used a toy Monte Carlo procedure to generate 10,000 mass distributions, each consisting of 579 events—just as the data in our final signal sample, falling in the region $0 \div 200\ GeV/c^2$. 91 masses were randomly chosen from a Gaussian distribution equal to the signal shape expected by the PYTHIA Monte Carlo ($M = 90.0\ GeV/c^2$ and $\sigma = 12.3\ GeV/c^2$, see Ch. 5), while the 488 events representing the background were randomly chosen from a mass distribution equal to a background shape obtained by fluctuating the parameters of the fitted background shape with multivariate constraints within their covariance matrix (Eq. 7.6), and applying to them the tag probability curve—again allowed to vary within its covariance matrix (Eq. 7.6). Each set of 579 events was then fit with our likelihood procedure, exactly the same way as the real fit on the experimental data. This procedure allows us to verify that the number of events returned by the fit is not systematically smaller or larger than the amount of signal in the real data, due to problems of the toy generation or other unknown biases. The most straightforward way to check these biases consists in plotting the distribution of the "*pull*", defined as follows:

$$pull = (N_{fit} - N_{true})/\sigma(N_{fit}). \tag{7.24}$$

From the central limit theorem, we expect the distribution of pulls to be normal, *id est* resemble a Gaussian with mean value equal to zero and sigma equal to one. A mean value significantly different from zero would imply the presence of a bias (either in the likelihood fitting procedure or in the toy Monte Carlo), while a width different from one would suggest that our treatment of errors in the fit is wrong. As can be seen in Fig. 7.9, the distribution of pulls is really well described by a Gaussian with the expected characteristics.

A second straightforward check can be performed by generating a single toy Monte Carlo distribution, amounting to 400 times the statistics of our final dataset, by taking exactly 400×91 masses distributed as the Monte Carlo signal shape and

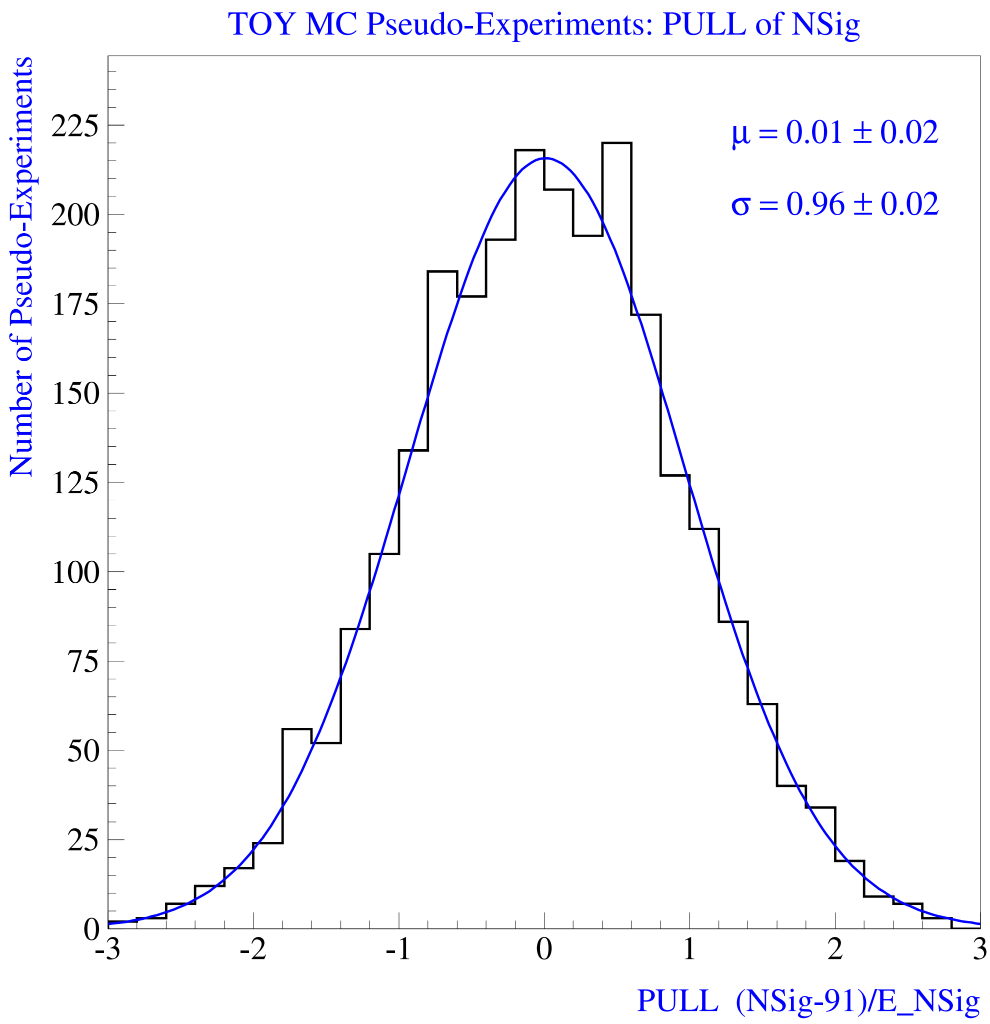

Figure 7.9: *Distribution of pulls in the number of signal events fitted in the toy Monte Carlo samples. See the text for details.*

exactly 400×488 masses distributed as the expected background—this time taken from the central fits of the background shape and the tag probability curve. This distribution is then fit exactly as the real data, as shown in Fig. 7.10, and the number of signal events, mass and width returned by the fitter are compared to the input numbers. These show a very good agreement indeed:

$$\begin{array}{rl} N_{sig}(fit) = 36,309 \pm 746 & N_{sig}(input) = 36,400 \\ N_{bgr}(fit) = 195,291 \pm 833 & N_{bgr}(input) = 195,100 \\ M_Z(fit) = 90.0 \pm 0.2\ GeV/c^2 & N_Z(input) = 90.0\ GeV/c^2 \\ \sigma_M(fit) = 12.3 \pm 0.2\ GeV/c^2 & \sigma_M(input) = 12.3\ GeV/c^2 \end{array}$$

giving a further blow to any speculation of a fitting bias.

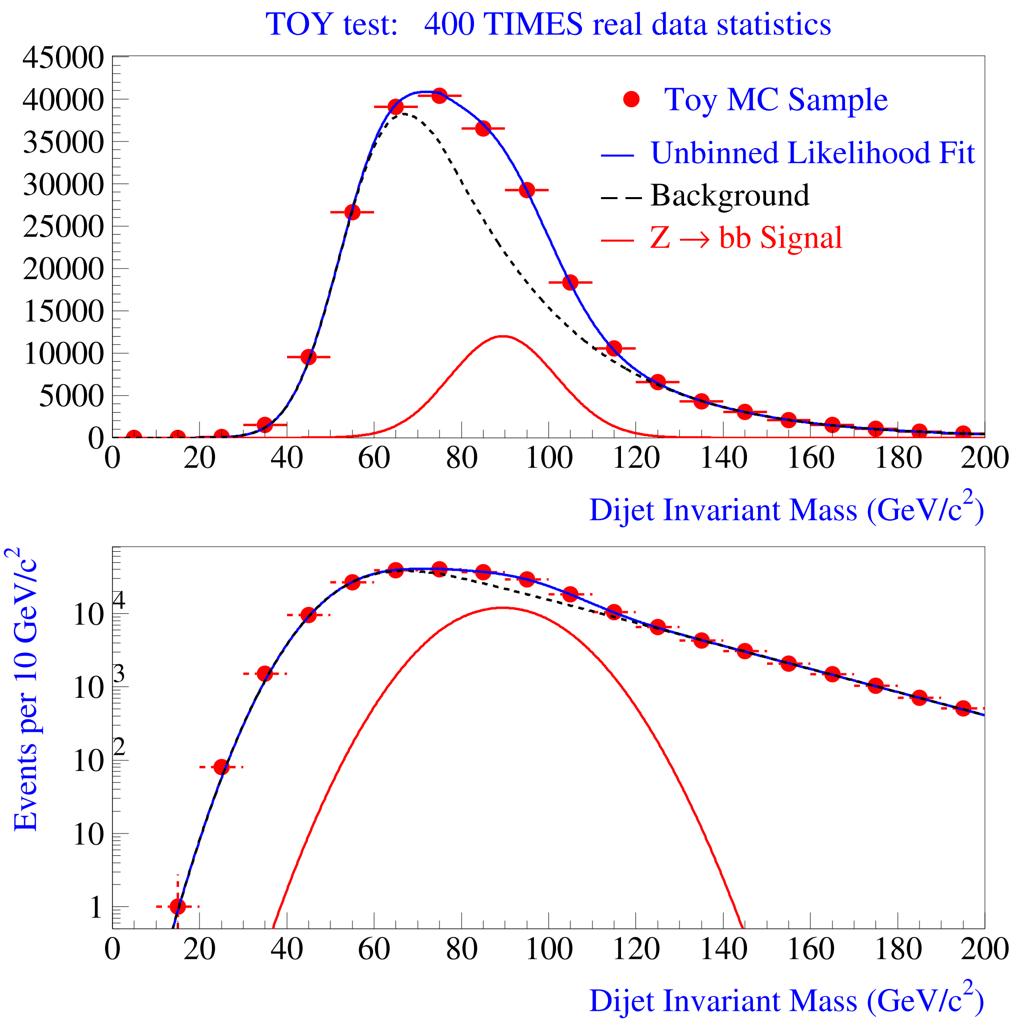

Figure 7.10: *Results of the fit to a toy Monte Carlo spectrum of 400 times the statistics of the real data.*

7.10.2 Significance of the Mass Distribution

Armed with some confidence in our tools, we can now generate 20,000 mass distributions of pure background with our toy Monte Carlo, again fluctuating with multivariate constraints its shape according to the covariance matrices of the tag probability curve and background shape parameters. Each distribution is then fit as the sum of the background and a free Gaussian shape, as we did on the real data. The frequency distribution of the number of signal events found by the fitter in the samples is shown in Fig. 7.11. One notices various features in this distribution. First of all, it is symmetric around zero, something we would have expected, given that the number of events in each sample is well above the limit of Gaussian

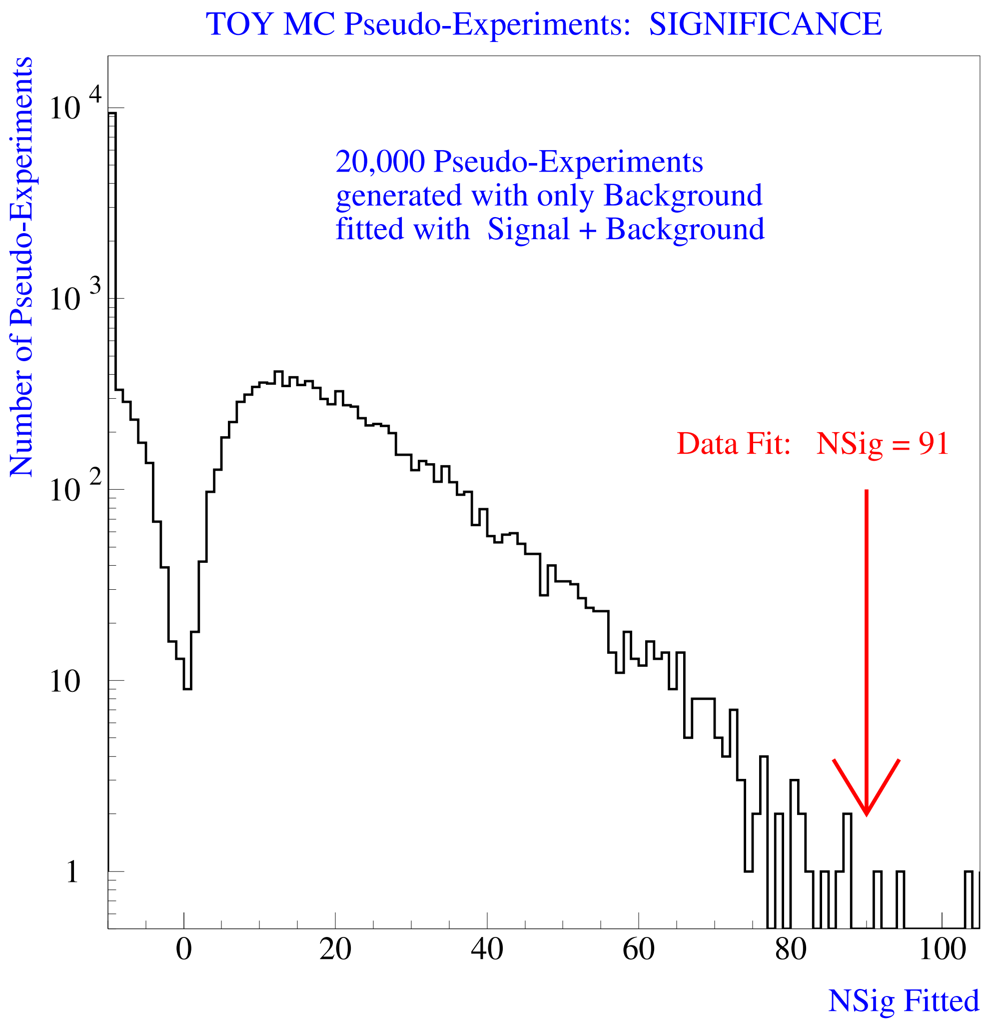

Figure 7.11: *Distribution of the number of fitted signal events in background mass distributions generated with a toy Monte Carlo. See the text for details.*

statistics[5]. Secondly, the distribution shows a dip where zero signal events are fit: the fitter *prefers* to use the Gaussian degrees of freedom to pick a small signal here or there, be it positive or negative with respect to the background shape[6]. Thirdly, the distribution sets in an exponential fall off for a large number of fitted signal events—again in agreement with what we would intuitively expect. We can then count the number of times that a fluctuation of the background allows us to fit a number of events equal to or greater than 91: this happens only three times out of $20,000$. This corresponds to a single-sided Gaussian fluctuation of

[5]If the distributions had a small number of entries, upper fluctuations would prevail, given that lower fluctuations (fitted as a negative number of signal events) are bound to remain greater or equal to zero.

[6]We indeed checked that the dip disappears if we fix the Gaussian shape parameters to a given value, underlining their statistical origin. Nonetheless, we are bound to use a free Gaussian for an evaluation of the statistical significance of the signal.

$$significance = 3.61\sigma, \tag{7.25}$$

which is our final estimate for the significance of our bump.

The number now obtained may be regarded as a conservative estimate, for two reasons. First of all, it is based on a pure comparison of *shapes*: no normalization factor enters the calculation. A significance number based on normalization information has been extracted in Ch. 6, but we cannot really combine the two numbers together, since they are based on the same data and on similar assumptions; nevertheless, the present estimate is certainly an underestimation of the significance. Secondly, we must note that the $3/20,000 = 0.015\%$ probability of obtaining a signal of 91 or more events from our fits is evaluated disregarding completely the shape and location of the signal: the Gaussian we use to parametrize a signal has three free parameters in the fit, and we are only using the first of them for our significance. This would be a fair technique for a search of a new process of unknown mass and experimental width, but in our case the Z boson cannot exactly show up anywhere in our spectrum, nor can it show an arbitrary width: the fact that mass and width turn out to give a very good agreement to our expectations cannot go unclaimed[7]. In conclusion, we have good reasons to declare that our 3.6σ significance, although far from the 5σ limit that is usually taken as the lower bound in a claim for a new observation, must be regarded as really conclusive evidence of the presence of $Z \to b\bar{b}$ decays in our dataset.

We will use the number of fitted signal events we have obtained in the present chapter for a preliminary determination of the $Z \to b\bar{b}$ cross section times branching ratio. That is the subject of the next chapter.

[7]In fact, the denominator in 3/20,000 would increase if we required to fit the gaussian with fixed parameters, corresponding to an increase of the value of the curve displayed in Fig. 7.11 for low values of N_{sig}.

Chapter 8

Calculation of the Z Cross Section

In the present chapter we give a preliminary estimation of the cross section times branching ratio of the $Z \to b\bar{b}$ process. After a brief introduction, we discuss in Sec.8.2 a rather complicated but unbiased method to estimate the efficiency of the kinematic selection cuts, based on the merging of a real $Z \to e^+e^-$ event with a $b\bar{b}$ final state from Monte Carlo. We use the results of that procedure to estimate the $Z \to b\bar{b}$ cross section in Sec.8.3.

8.1 Introduction

In the present chapter we present a preliminary measurement of the Z boson production cross section in proton-antiproton collisions at the center-of-mass energy of $\sqrt{s} = 1.8\ TeV$.

The preliminary nature of this measurement stems from various facts, which we list in the following:

- At the time of writing, a new (and improved) evaluation of the integrated luminosity of the dynamically prescaled triggers is under way at CDF; a careful work has been performed by D. Hennessy[64] to use our best knowledge of the $p\bar{p}$ inelastic cross section and the details of the various triggers to reduce the systematic uncertainty on this quantity, which of course limits the accuracy of all cross section measurements performed by CDF. The present knowledge of the effective luminosity of our experimental dataset is therefore going to become obsolete very soon, and the reduction of the luminosity error will make advantageous the use of more careful techniques to evaluate the cross section itself: as the reader may remember, in Sec.3.2.1 we discussed an approximate evaluation of the combined integrated luminosity of the various triggers contributing to the inclusive muon dataset. That approximate method is justified if the systematic uncertainty in the luminosity of any single trigger stream is 8%, but will become too sketchy when a better knowledge of that quantity is achieved (a total uncertainty of about 4% is estimated in the new determination): in the latter case, separate Monte Carlo simulations—one per each contributing trigger—need to be merged together according to the relative contributions of the various triggers.

- The SECVTX scale factor f, introduced in Sec.2.6.4 and used in all our estimates of the amount of $Z \to b\bar{b}$ signal in the muon dataset, is not an official number yet. The CDF Collaboration is currently reviewing its determination, in view of a possible new publication of the $t\bar{t}$ cross section—which is as much dependent on the value of f as is our Z cross section, given that both analyses make heavy use of SECVTX tagging. Because of the preliminary nature of the current value of f, a Z cross section measurement is preliminary itself.

- A determination of the Z cross section at the Tevatron from the sample of $Z \to b\bar{b}$ decays extracted in the present work must of course be regarded just as a check of consistency, given that the statistical error alone is going[1] to amount to more than four times the total uncertainty of the σ_Z measurement obtained in the analysis of e^+e^- decays[54]. We do not consider it the most important part of the present analysis, which consists instead in having demonstrated the possibility of extracting a $Z \to b\bar{b}$ signal in Tevatron data. We plan to use the present work as a starting point for a new analysis, aimed at an optimization of the selection strategy for a cross section measurement rather than a proof of the presence of $Z \to b\bar{b}$ events in the dataset.

These issues make the present determination of the Z cross section taste just like a warming-up exercise. The number we will obtain here is correct, but also certainly preliminary and subject to future improvements. Because of that, we will not try to evaluate the small contributions to the systematic uncertainty coming from the choice of parton distribution functions and fragmentation model. They are certainly irrelevant in our measurement.

8.2 Efficiency of the Selection Cuts

The acceptance of the kinematic requirements for the $Z \to b\bar{b}$ process is not as easy to obtain as are the acceptances for the muon identification cuts (either at trigger level or the additional offline criteria) and the double SECVTX tagging requirement. The two variables we have used for the selection of our final dataset are in fact highly dependent on the characteristics of the transverse energy depositions in the calorimeters, as well as on the amount and pattern of initial and final state radiation for the process under study. While we can confidently rely on PYTHIA or HERWIG to simulate correctly the radiation from the $b\bar{b}$ singlet produced in the Z decay, the initial state radiation carries a larger uncertainty, since it depends on the choice of parton distribution functions used for the generation of the process. By far, though, the most difficult effects to factor in are the underlying event and the presence of multiple interactions in the same bunch crossing that produced the selected events. We shall briefly discuss these effects in what follows.

[1]See Eq.7.23: the relative statistical uncertainty is 33% alone.

Multiple interactions occurring during the same bunch crossing are not normally included in a Monte Carlo simulation; their impact on the quantities measured in the detector is usually negligible, their largest effect being an increase, proportional to the instantaneous luminosity, in the total energy measured in the calorimeters[2]; the standard CDF jet energy corrections, already described in Sec.2.6.1, take the effect of multiple interactions into account by subtracting from the corrected transverse energy of the jets in a given event a fixed value that depends on the number of good primary vertices reconstructed by the tracking: this number is the average E_T deposition due to these additional interactions inside the $\eta\phi$ area spanned by the jet cones. In our case, the definition of the $\Sigma_3 E_T$ is highly dependent on the instantaneous luminosity, since the number and energy of the additional clusters in the calorimeters is a function of it, and the sum is willfully made on raw jet energies. To estimate the efficiency of a cut on the $\Sigma_3 E_T$ for the signal we then have to take into account its shift due to the luminosity. In Appendix A.3.3 we discuss a treatment of this problem that relies on adding to the energy of the calorimeter towers in Monte Carlo events the energy measured in real minimum bias events, taken at the same instantaneous luminosity one wishes to simulate: by stating the obvious, *id est* that the characteristics of a hard scattering cannot depend on those of any additional interaction occurred in the same bunch crossing and vice-versa, one can overlay incoherently the two effects, and obtain a calorimeter response to a Monte Carlo event added up of the typical contribution it could have received at a given luminosity. The method works very well, and we could use it without further ado here for the purpose of the measurement of the $\Sigma_3 E_T$ efficiency. But—as is demonstrated in Sec.4.2 by comparing $\Sigma_3 E_T$ distributions for real $Z \to e^+e^-$ data and Monte Carlo $Z \to e^+e^-$ data mixed according to the recipe now outlined—the multiple interactions are not the only effect we must account for. The other is the underlying event.

What is—or should be[3]—called *underlying event* is the amount of detectable radiation coming from the spectator partons in the two colliding hadrons. The spectators are sometimes involved in a hard or semihard collision themselves, and in that case one speaks of a *double* or a *multiple* parton scattering[65]. In any case, the Monte Carlo generators are not able to simulate with sufficient accuracy these

[2]This happens when the data are collected by triggers requiring the presence of objects that cannot be faked—or whose rate is not increased—by the simultaneous presence of a handful of additional minimum bias interactions: this is the case of most triggering objects like electrons and muons, photons, and so on; the datasets collected by these triggers do not show any luminosity-dependent bias in the variables defining the triggering object. Conversely, care must be taken with ΣE_T triggers, whose rate increases with instantaneous luminosity due to the contributions from "pileup" events; another example is the multijet trigger, a trigger requiring the presence of four energy clusters above a given threshold: above a certain luminosity, the cross section for a single hard scattering giving four jets is surpassed by the combined cross section of two independent dijet events.

[3]As a matter of fact, the term "underlying event" is frequently and—at least in our opinion—incorrectly used to describe the cumulative effect of hadron remnants, multiple parton scatterings in the same $p\bar{p}$ collision, and pileup events, i.e. multiple satellite interactions in the same bunch crossing. We prefer to separate the latter effect from the others, since their proper treatment in a Monte Carlo simulation is a completely different problem.

effects. Use is sometimes made of some tunable parameters to reproduce the observed charged track multiplicity and the track P_T spectrum for a given process; but different processes will *a priori* yield a different amount of radiation from these mechanisms, due to the different initial states involved. For our purpose, we could tune these parameters to reproduce the $\Sigma_3 E_T$ distribution in the experimental dataset of $Z \to e^+e^-$ decays with a Monte Carlo of the same process: the same initial state configuration is involved in a $Z \to b\bar{b}$ event, and therefore one could export the same tuned parameters for the simulation of the signal.

The method we actually use is less straightforward, but has the strong advantage of not involving the tuning of arbitrary quantities, and relying on a Monte Carlo generator only where it is believed to give the right answer without any fine tuning or arbitrary input. We start from the observation that initial and final states are color-disconnected in the Drell-Yan process, and therefore the gluon radiation in the initial state is not dependent on what happens to the final state (and conversely). We do have a very clean sample of events where we can study the effect of initial state radiation, multiple interactions, and underlying event: it is the $Z \to e^+e^-$ dataset—by now well known to the reader—, a virtually clean sample of Z production and decay to e^+e^- pairs. For a precise determination of the acceptance of our data selection we can use those events, provided that we add the effect of the final state radiation, which is present in the $Z \to b\bar{b}$ decay. Monte Carlo generators such as HERWIG or PYTHIA perform very well in the simulation of final state radiation: we can then try and join data and simulation to provide the best understanding of our process.

In order to join the initial state of a real event with a final state from Monte Carlo, we need to merge their calorimetric banks and use a certain number of technicalities. The various ingredients are described in what follows:

1. A list of electron and positron trimomenta and event vertex z coordinates from the 6722 events in the $Z \to e^+e^-$ dataset is needed by the Monte Carlo, to generate outgoing b quarks with the same momenta of the final state leptons.

2. We use HERWIG for the generation of the b quark final state, since its features allow us to perform our task in the easiest way. Version 5_9 of HERWIG allows the user to completely switch off the initial state radiation, but unfortunately that version is not supported by the CDF standard offline package, and cannot therefore be interfaced with the CDF detector simulation program, QFL. To overcome this problem we had to implement the needed option into the older 5_6 version. Checks were made to make sure that, after switching off the initial state radiation, the results were in agreement with expectations, and that proton and antiproton remnants were not generating any final state particle with appreciable transverse momentum.

3. HERWIG needs to be modified in order to generate a final state in a given momentum configuration. We developed a routine to read the electron and positron momenta from an ASCII file, and produce a b quark pair with those

momenta in a color-singlet configuration. That also required us to force HERWIG to violate momentum conservation, since in a collision with no initial state radiation the final state momenta should balance in the transverse plane, unlike the input momenta, which come from a real $Z \to e^+e^-$ event that may receive a significant contribution from initial state radiation.

4. The final state particles produced by the fragmentation of the b quarks have to be attached to a vertex with the same z coordinate of the corresponding $Z \to e^+e^-$ event, if we wish to match the overall topology of the simulated event to the real $Z \to e^+e^-$ topology. We therefore modified the detector simulation, QFL, where the z coordinate is normally chosen randomly according to the observed distribution in the real data, to force the event vertex be the one in the input ASCII list.

5. Events generated from the $Z \to b\bar{b}$ decay must have a muon passing our trigger and offline criteria in order to allow us an unbiased determination of acceptance of the kinematic cuts. In fact, the presence of the muon may indeed give a small difference in the $\Sigma_3 E_T$ distribution, since a jet containing a muon satisfies particular angular constraints (the central muon chambers extend only up to $|\eta_d| = 0.6$ in rapidity) that are not necessarily shared by the electrons from the reconstructed $Z \to e^+e^-$ decays; moreover, the muon—accompanied by its neutrino—normally carries away a respectable portion of the energy of the jet it belongs to[4]. This may sometimes cause the muon jet to become the third in the uncorrected-energy-ordered list, causing a bias not present in normal hadronic b-quark decays. To overcome this bias, we needed to implement an iterative procedure: after the decay of the b quarks had taken place, a routine read the data bank containing the information for the generated particles, and discarded the event if there was no generated muon with characteristics (P_T , η) that made it likely to pass the offline cuts. If the event was discarded, the same electron trimomenta were used again for another $b\bar{b}$ decay simulation. This loop may go on forever in situations where both input electrons have trimomenta pointing far away from the muon chambers. Whenever that problem arose, the dielectron event was discarded, and the following one used as input.

6. We developed a routine to modify the bank containing the amount of energy deposited in each calorimeter tower, totally removing the energy belonging to the electron and positron clusters in the real $Z \to e^+e^-$ events.

7. The last tool necessary for our task was a merging algorithm that allowed us to add the contents of the calorimeter banks from the initial state of the $Z \to e^+e^-$ event to those of the simulated final state $b\bar{b}$ pair. The merged banks were then used to reconstruct the event and obtain a new list of clusters.

The final result of this rather complicated process is a sample of hybrid (*id est*, partly simulated) $Z \to b\bar{b}$ events that do not suffer from any of the problems of

[4]Compare to the results of Ch.5.

normal simulated processes: their calorimetric banks are as similar as they may get to those a real Z signal would show. The price to pay is a rather low statistics: of the already small initial sample of 6722 events, 758 get discarded by the iterative muon filtering procedure; offline, we have to apply our tight muon cuts and require that the leading jets are both taggable[5], ending with a dataset of only 1075 events. The sample is however large enough to allow a sufficient determination of the acceptance of the kinematic selection cuts: we must not forget that our signal amounts to less than 100 events, and the statistical error due to the uncertainty in the likelihood fit will dominate by far any other uncertainty.

To estimate the efficiency of the kinematical cuts we cannot straightforwardly count how many events pass the selection criteria $\Sigma_3 E_T < 10\ GeV$, $\Delta\Phi_{12} > 3$. There is in fact one last bias to deal with: the $Z \to e^+e^-$ decays, which we have been using as a blueprint for the generation of the mixed events, have been collected by a trigger different from the one that collected our inclusive muon data. The high E_T electron trigger was not dynamically prescaled, and thence its instantaneous luminosity distribution is quite different from that of the SECVTX data (Fig.8.1). The instantaneous luminosity affects the $\Sigma_3 E_T$ distribution, biasing our acceptance estimates.

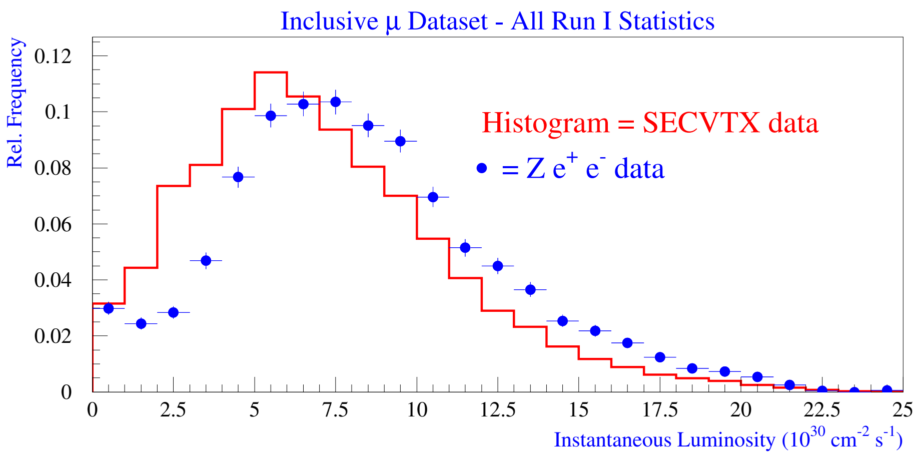

Figure 8.1: *Instantaneous luminosity distribution of the SECVTX data and the* $Z \to e^+e^-$ *dataset.*

To deal with this bias we must invert our former selection chain: we have to first estimate the efficiency of the $\Delta\Phi_{12}$ cut, which is not dependent on instantaneous luminosity[6], and then unfold the bias to the $\Sigma_3 E_T$ distribution given by the higher instantaneous luminosity of the $Z \to e^+e^-$ data. The cut on $\Delta\Phi_{12}$ selects 541 events

[5]Requiring both jets to be tagged by SECVTX would reduce too much the sample, and we already know, from the studies described in Sec.6.5, that there is no sizable correlation between SECVTX tagging and the kinematic variables.

[6]In fact, the kinematic of the final state partons belonging to the Drell-Yan process is totally insensitive to any additional interaction occurred in the same bunch crossing.

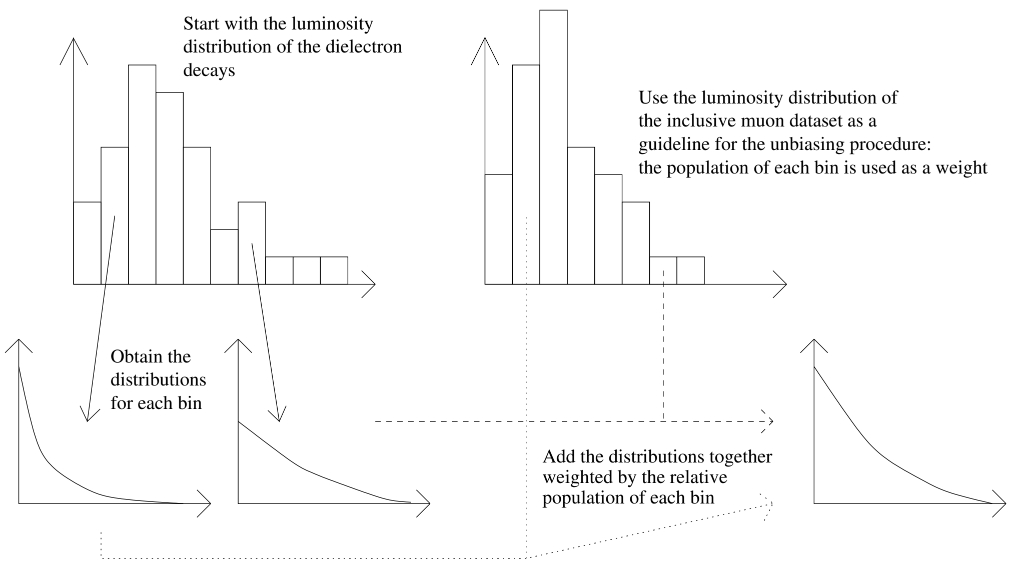

Figure 8.2: *Graphical explanation of the unfolding procedure of the luminosity bias to the $\Sigma_3 E_T$ cut acceptance.*

out of 1075, with an efficiency $\epsilon_{\Delta\Phi} = 50.3 \pm 2.7\%$ (statistical error only): we now see that our prediction of degradations of the order of 10% in the acceptance (Sec.4.2.2), due to the effect of final state radiation, was not too bad: the efficiency has degraded by roughly 13% from the value (58.8±1.3%) we measured in plain $Z \to e^+e^-$ decays[7]. We can take as an estimate of the systematic error on the efficiency the difference with this number, so that our final estimate is $\epsilon_{\Delta\Phi} = 50.3 \pm 2.7 \pm 8.5\%$.

To correct the $\Sigma_3 E_T$ distribution for the luminosity bias we divide the luminosity spectrum of $Z \to e^+e^-$ data accepted by the $\Delta\Phi_{12}$ cut into 25 bins, obtaining 25 $\Sigma_3 E_T$ spectra from the corresponding simulated $Z \to b\bar{b}$ events falling in each bin (after the cut $\Delta\Phi_{12} > 3$). We can then weight these $\Sigma_3 E_T$ distributions for each subsample according to the luminosity distribution in the inclusive muon data (see Fig.8.2). This procedure unfolds the bias and is virtually systematics-free. The acceptance of the cut $\Sigma_3 E_T < 10$ GeV on $Z \to b\bar{b}$ decays already selected by the $\Delta\Phi > 3$ requirement can now be estimated to be $\epsilon_{\Sigma E_T} = 43.2 \pm 4.1\%$, where the statistical uncertainty has been obtained by propagating the relative uncertainties in all the $\Sigma_3 E_T$ shapes merged together and in the luminosity distribution used for the weighting procedure. A systematic uncertainty to this number could in principle be given by the difference with the plain Monte Carlo prediction, but we know the latter cannot be correct: the difference is huge (21.4%), and the sources of the mismatch are those well-known effects we have fought so hard to insert in our tuned Monte Carlo. We believe that a sounder choice is to assign to that number, as a systematic

[7]Of course, the numbers we obtain from the merging procedure are still dependent on the modeling of final state radiation performed by HERWIG; we can guess from the numbers above that using PYTHIA instead would not change too much the results, though.

error, the difference with the efficiency of the cut on $Z \to e^+e^-$ data, before any luminosity correction ($46.3 \pm 1.5\%$). Our final estimate is then $43.2 \pm 4.1 \pm 3.1\%$.

8.3 Results

We show here once again all the numbers we are about to use for our preliminary estimation of the Z boson production cross section:

- the number of fitted $Z \to b\bar{b}$ events in our final dataset has been estimated to be $N_{obs} = 91 \pm 30 \pm 19$ events (Eq.7.23, Ch.7);
- the integrated luminosity of the experimental dataset amounts to $\int \mathcal{L}dt = 103 \pm 8\,pb^{-1}$ (Sec.3.2.1);
- the combined acceptance of the selection leading to the double SECVTX tagged data is $\epsilon_{2\,SECVTX} = 0.176 \pm 0.033\%$, where the error is totally dominated by the uncertainty in the SECVTX scale factor (Sec.2.6.4);
- the combined acceptance of the kinematic requirements discussed above is $\epsilon_{KIN} = 21.7 \pm 4.9 \pm 5.5\%$. The systematic uncertainty is taken as the difference between the quoted number and the combined acceptance of the two kinematic requirements on $Z \to e^+e^-$ data (27.2%, see Sec.4.2).

These numbers are all we need for a computation of the $Z \to b\bar{b}$ cross section. We can use the following expression:

$$\sigma_Z \times BR(Z \to b\bar{b}) = \frac{N_{obs}}{\int \mathcal{L}dt \times \epsilon_{2\,SECVTX}\epsilon_{KIN}}. \tag{8.1}$$

Combining all uncertainties together we finally obtain:

$$\sigma_Z \times BR(Z \to b\bar{b}) = 2.3 \pm 1.3\,nb. \tag{8.2}$$

A final comment is necessary. The value we obtain for the cross section of the searched process is twice as big as what was expected from previous measurements at the Tevatron in other channels (e^+e^- , $\mu^+\mu^-$). Nevertheless, the uncertainties on its determination are really large, and the result is within one statistical deviation from the expected value; the systematic uncertainties have been evaluated conservatively, but the dominant error is due to the statistical uncertainty in the likelihood fit.

The present determination, albeit very preliminary, may be regarded as additional evidence of the measurability of the $Z \to b\bar{b}$ decay at the Tevatron collider: the error is about seven times larger than the most precise determination of the cross section of Z production, obtained from one of the cleanest final states available at the Tevatron (two high P_T electrons)[54], but it has been obtained from a dataset featuring a starting S/N ratio more than four orders of magnitude smaller and a starting acceptance almost two orders of magnitude smaller. We therefore believe that the present result is a success.

Chapter 9

Conclusions and Prospects

In this chapter we summarize the results of the present work, and describe their application for the physics that the new CDF II detector will do in run 2. In Sec. 9.1 we give our results, and in Sec. 9.2 we describe the possibility of collecting a large dataset of $Z \to b\bar{b}$ events with dedicated triggers in run 2, which will probably be one of the decisive factors in the reduction of the systematic uncertainty of the top quark mass measurement.

9.1 $Z \to b\bar{b}$ in Run 1

9.1.1 Introduction

Audaces fortuna iuvat, goes a say much older than the land where the Tevatron sits. We have undertaken a search for the $Z \to b\bar{b}$ decay in hadron collider data, and we have found it. The search was audacious because the probability of failure seemed high at the beginning: the enormity of background rate, and the lack of a dedicated trigger, indeed look like difficult factors to reckon with. Besides, a small success rate is commonplace in high energy physics when a new phenomenon is sought, but if one looks for a particle as deeply studied as the Z boson, one has better find it, or else his efforts will sound rather seamless and amount to a big waste of time. Luckily, this was not so.

9.1.2 Results of the Present Study

The importance of detecting a $Z \to b\bar{b}$ signal in our run 1 data has been stressed in Ch. 1.1: to summarize it briefly here, we remind the reader that this process has been isolated here for the first time in a proton-antiproton experiment. But this may not be a justification alone, given the thorough knowledge we already gathered on the properties of this particle. The real importance of the present search lies in yielding a proof that such decays can really be isolated, and thus be used to provide a calibration line for the jet energy measurement if the statistics is high enough. In run 2 we will hopefully be able to use this tool to reduce the jet energy scale systematic uncertainty in the top mass determination to a level comparable to—or, optimistically, smaller than—the expected statistical uncertainty, which for a dataset of 2 fb^{-1} is expected to be smaller than 1 GeV/c^2. If no new physics will be found by CDF and D0 in run 2, the reduction of systematic uncertainties in the top mass measurement will become one of the critical factors in determining if an extension of data taking over the expected running time is worth the effort.

We have therefore sought the $Z \to b\bar{b}$ decay in a sample of 5.4 million events collected by triggers requiring the presence of a central muon candidate. This choice was dictated by the lack of a trigger collecting dijet events with an energy threshold small enough to be acceptant for hadronic Z decays: a dataset with leptons is then the best starting point to collect b-quark decays, and the muon sample warrants to be more efficient for our signal, due to the relative ease with which muons can be identified inside a high E_T jet, when compared to the tight requirements that must be applied to electron candidates to obtain a similar dataset (see App. A).

The signal-over-noise ratio at trigger level was predicted to be smaller than 10^{-3}. This very small number needed a drastic increase in order to allow the identification of a bump in the mass distribution. To achieve that, a tight selection based on increasing the $b\bar{b}$ purity of the sample was performed by requiring that the two leading jets in the events contained a set of tracks coming from a common secondary vertex. A kinematic selection was then performed by using two variables capable of discriminating the electroweak process from the QCD backoground of direct $b\bar{b}$ production. At the end of the selection 588 events were left, with a signal-to-noise ratio of roughly 1/5 in the Z mass region, *id est* 500 times more than what it was at trigger level.

A detailed study of the energy resolution for b-quark jets was performed with a Monte Carlo generation of $Z \to b\bar{b}$ events that passed all our selection requirements, using their particular characteristics to design an improved set of energy corrections; these allowed a relative improvement in the $b\bar{b}$ mass resolution of 50%. This proved fruitful for the identification of the Z signal in the mass distribution of the selected sample, and further increased the peak S/N ratio to a value close to 1/3.

In order to demonstrate the presence of the signal in the selected sample, we devised a way to estimate the absolute number of background events as a function of the dijet invariant mass. The very good precision of the method allowed the extraction of an excess of 70 ± 20 events, inconsistent with the hypothesis of an upper Poisson fluctuation of the background by 3.2 σ. The method also allowed us to check that the improved jet corrections we had devised worked exactly as expected for the Z signal.

Finally, we devised an unbinned likelihood procedure to fit the mass distribution of the selected dataset to the sum of background and signal contributions. This enabled us to check the consistency of the signal shape and size to the expectations from Monte Carlo simulations. The number of fitted events was used to measure the cross section of Z production at the Tevatron: we found $\sigma_Z \times BR(Z \to b\bar{b}) = 2.3 \pm 1.3$ nb. This measurement is in agreement to the previous Tevatron measurements of the Z boson cross section, obtained from leptonic decays.

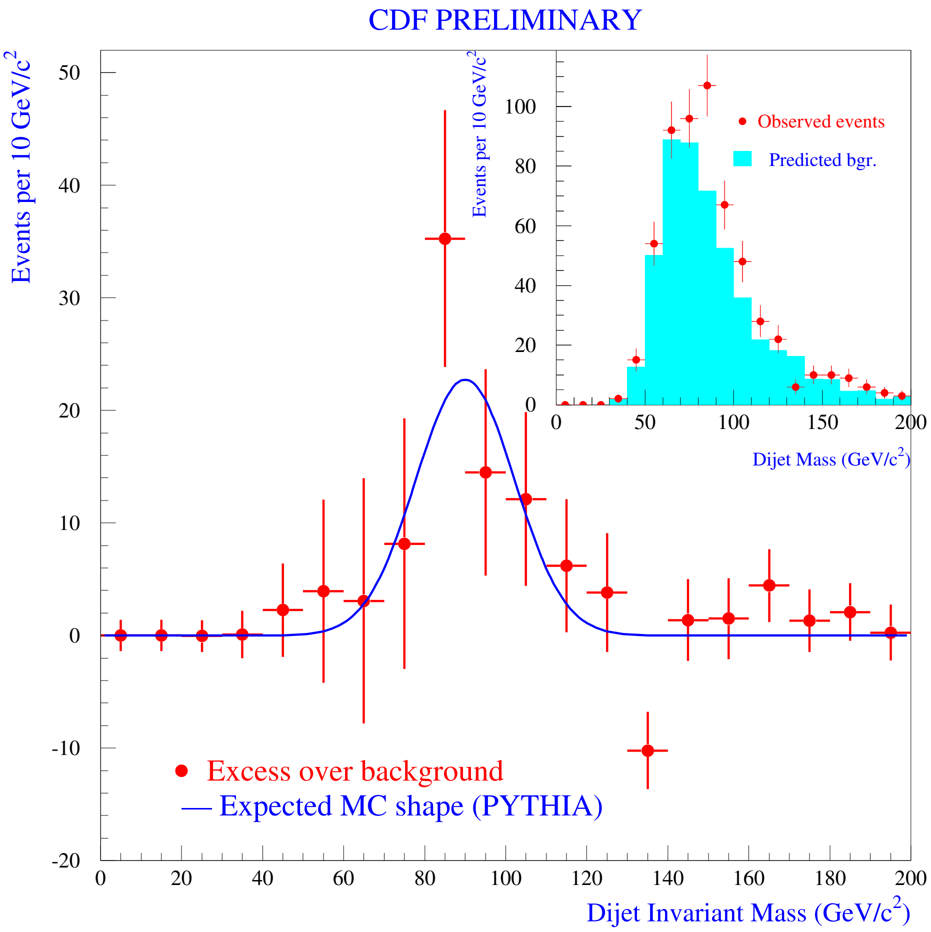

Figure 9.1: *Results of the counting experiment: the excess over background predictions is distributed as expected from $Z \to b\bar{b}$ decays (blue line, PYTHIA Monte Carlo, arbitrary normalization). Inset: the points represent the observed events, the full histogram is the absolute background prediction computed with the counting method.*

9.2 $Z \to b\bar{b}$ in Run 2

9.2.1 Introduction to Run 2

The Tevatron run 2 will start in the year 2000. The physics potential of the CDF and D0 detectors will be boosted with respect to run 1 by the improvements achieved both in the accelerator and in the design of the detectors. A new accelerator, the Main Injector, has been successfully built to replace the main ring as injecting machine for the Tevatron synchrotron: this will allow the Tevatron to operate with higher intensity beams and with a bunch spacing reduced from the 3.5 μs of run 1 down to 396 ns and finally 132 ns, with a corresponding increase of the number of orbiting bunches from six up to 108: these factors will increase the instantaneous $p\bar{p}$ luminosity of at least an order of magnitude, *id est* up to $\mathcal{L} \sim 2 \times 10^{32}\ cm^{-2}s^{-1}$, while keeping the average number of interactions per bunch crossing at a manageable level[1] (see Fig. 9.2), preventing a massive overlap of simultaneous hard scatterings which could make physics analysis very difficult[66].

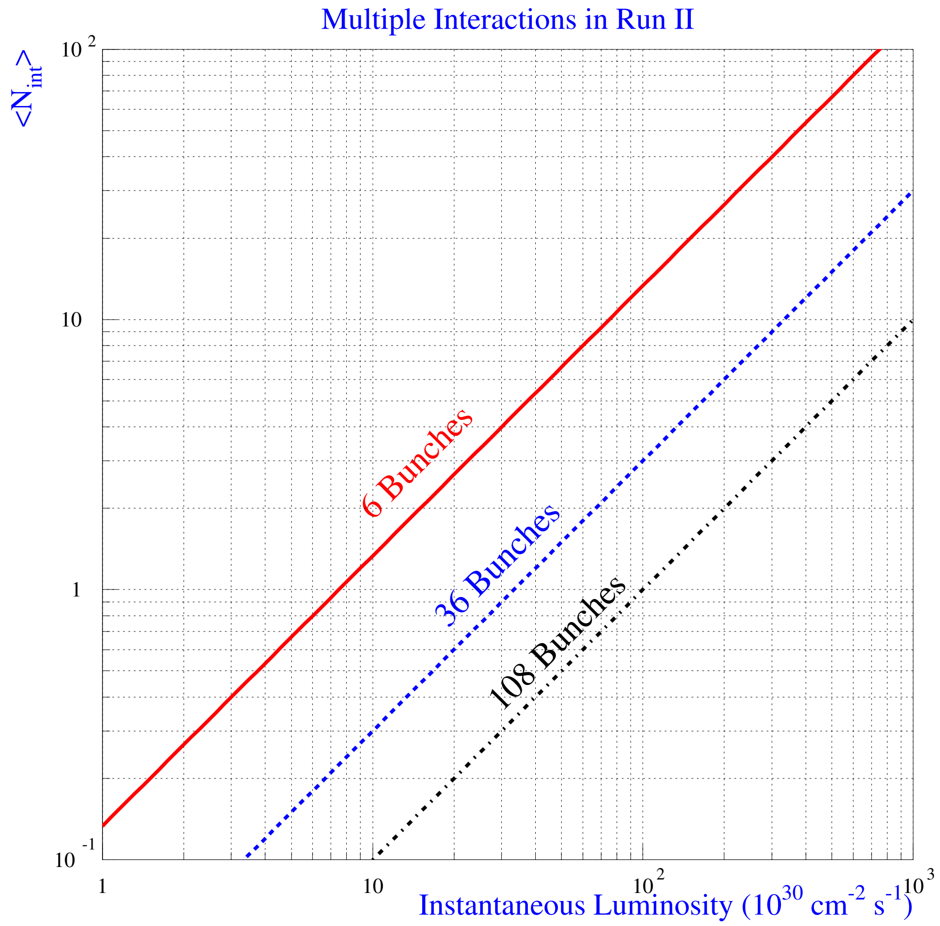

Figure 9.2: *Average number of interactions per crossing as a function of instantaneous luminosity for Tevatron run 2.*

The Tevatron energy will also increase, from 900 to 1000 GeV per beam: this will in turn increase the production rate of heavy objects such as top quark pairs[2].

[1]Run 1 experience suggests that the simultaneous interactions in each bunch crossing must not exceed the number of five: higher rates would make it difficult to interpret certain characteristics of the event, such as the calorimetry, due to the limited coverage of central tracking in rapidity. It would become very difficult to correctly associate the appearance of a photon candidate in the calorimeter or a significant amount of missing transverse energy to a particular hard scattering producing other physics objects: both these signatures are exceedingly important for exotic physics searches that CDF and D0 intend to perform in run 2.

[2]The cross section for $t\bar{t}$ production increases by 40% when the c.m. energy is raised to 2 TeV.

The decrease in bunch spacing demands a massive upgrade of the data acquisition and new strategies for the trigger system for the CDF detector. To allow a reasonable 5.5 μs decision time for the Level 1 trigger, all front-end electronics will need to be pipelined with a buffering of 42 beam crossings. Informations from the calorimeters, the central tracking and the muon chambers will be used to decide if a particular bunch crossing deserves further investigation. The Level 2 input rate is designed to be lower than 50 kHz, and a massive reduction in rate, down to about 300 Hz, will be required to keep the deadtime lower than 10% if the average decision time of the Level 2 triggers is to be kept at about 20 μs. Level 3 will then be responsible of a further selection to fit the $30 \div 50$ Hz budget of mass storage.

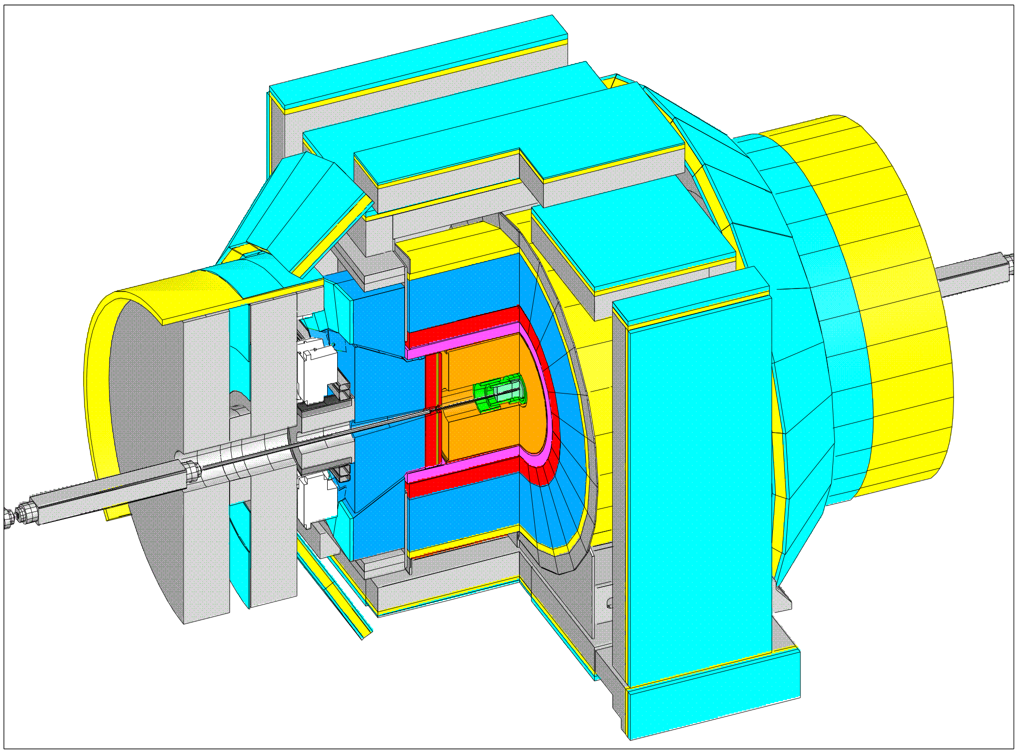

Figure 9.3: *Isometric view of the CDF II detector.*

The CDF II detector will not only operate with a new DAQ system. Its whole inner tracking system will be completely upgraded by the insertion of a set of very redundant and high precision devices. The SVX II, a new three-barrel, double-sided silicon vertex detector, will provide many more hits in the $r\phi$ plane per track, securing performances similar or better than those of the old detector and suffering no degradation in performance in a much higher radiation environment. Outside of it, three additional layers of silicon, named ISL (for Intermediate Silicon Layers) will provide valuable additional information. Finally, the old CTC will be replaced by the COT, a new tracking chamber of high redundancy and performance. The old Endplug calorimeters will be replaced with a new scintillator-based system, and the muon system will also be upgraded to increase the coverage in the $\eta\phi$ regions not instrumented during run 1. A time-of-flight detector is also under consideration in the space between the COT and the magnet coils. The CDF II detector is shown in Fig. 9.3.

9.2.2 Design of a $Z \to b\bar{b}$ Trigger

The accept rates quoted above for the three trigger levels are the main guidelines for the design of any trigger for CDF in run 2: depending on the relevance of the acquisition of a particular dataset, a larger or smaller portion of the total bandwidth at each level can be devoted to that purpose.

As was discussed in Ch. 1, the importance of the identification of a $Z \to b\bar{b}$ signal in run 1 lays in the prospects it may open for the precision measurement of the jet energy scale for b quarks, a subject of crucial importance in the determination of the mass of the top quark and hopefully of new particles decaying to b-quark jets. Armed with the knowledge of the tools necessary for the reduction of background rate in the search for the $Z \to b\bar{b}$ decay, CDF is currently designing a trigger strategy for the collection of a significant amount of this signal.

The collection of $Z \to b\bar{b}$ decays with lepton triggers has been proven possible in the present work, but on the other hand it has also shown its two main limits, namely a very low collection efficiency and a strong bias due to the semileptonic decay. As far as the low efficiency issue is concerned, we have learned in Ch. 3 that a muon trigger with a combined Level 3 cross section[3] of about 50 nb had a collection efficiency of only $\sim 1.5\%$ for the $Z \to b\bar{b}$ decay: increasing the efficiency is therefore not viable by just lowering the muon momentum threshold, since the mass storage limitations give an upper bound of 30 to 50 Hz—corresponding to less than 2 μb at $\mathcal{L} = 2 \times 10^{32}\ cm^{-2}s^{-1}$—for the combined trigger cross section at Level 3: for the sake of collecting a calibration dataset, $\sim 3\%$ of the acceptance is already a large number[4]. Besides, the bias to the b-jet energy in a $Z \to b\bar{b}$ event collected with a muon trigger would make it very difficult to understand the energy scale for the b-quark jets in a $t\bar{t}$ event collected by a W trigger: the calorimeter response is in fact very different in the two cases.

A much better strategy is viable in run 2, since we will profit from two new important hardware tools: the reconstruction of charged tracks at Level 1 by a very fast online tracking algorithm, XFT[66], and the measurement of the impact parameter of the tracks at Level 2 by SVT, the Silicon Vertex Tracker[67].

Two different approaches have been proposed to collect a significant amount of $Z \to b\bar{b}$ decays in run 2 without drawing on the tight trigger budgets. We briefly describe each in the following; the rates quoted are relevant for operations at the design run 2 luminosity of $2 \times 10^{32}\ cm^{-2}s^{-1}$ in the first study, and the slightly lower value of $1.4 \times 10^{32}\ cm^{-2}s^{-1}$ in the second.

In the study by M. Shochet and R. Culbertson[68], $Z \to b\bar{b}$ events are collected

[3] The inclusive muon dataset corresponding to the total run 1 luminosity of 100 pb^{-1} amounts in fact to 5 million events.

[4] Of course a large muon dataset is very appealing for b physics, but its collection cannot be justified by our arguments alone.

at Level 1 by the combined requirement of two tracks with transverse momentum greater than 5 GeV and two calorimeter towers with E_T above the same threshold; these requirements are reasonably efficient for the signal and give an expected rate of 466 Hz, *id est* less than 1% of the expected Level 1 accept rate. At Level 2, SVT information is used to require the presence of two tracks with $P_T > 2\ GeV/c$, separated in azimuth $\Delta\phi_{12} > \pi/2$, each with impact parameter larger than 150 μm: this allows a huge reduction of background, and effectively selects heavy flavor decays; the decay characteristics of the Z signal are exploited by requiring two central calorimeter clusters with $\Delta\phi > \pi - 0.5$ and imposing a veto on a third cluster in the calorimeter. The expected rate of such events is less than 9 Hz, slightly high but still reasonable if compared to the total Level 2 bandwidth (300 Hz). At Level 3, the full reconstruction of the event by software algorithms allows to select events containing two jets tagged by SECVTX, with a total expected rate of 0.1 Hz; the expected amount of $Z \to b\bar{b}$ events that would be thus collected in the expected run 2 dataset (2 fb^{-1}) is 34,000: a sample large enough to allow for a precise calibration of the b-jet energy scale. As we see, this trigger chain exploits those signal characteristics we have proven to be most effective in the reduction of the QCD background.

The second detailed study has been carried out by S. Leone and C. Bigongiari[69]. At Level 1 they propose a trigger based on track informations alone: one track with $P_T > 6\ GeV/c$ and another with $P_T > 4\ GeV/c$, with $\Delta\phi > 150°$, both having rapidity $|\eta| < 1$; the expected rate is high (5.1 kHz) but the absence of a calorimetric requirement makes this trigger of possible use to other analyses. At Level 2, the same requirements of the Shochet/Culbertson study are applied on the two tracks, but the tighter cut already applied at Level 1 on the tracks opening angle is enforced; no calorimetric cuts are applied, and the total rate is expected to be about 8.5 Hz. At Level 3, a b-quark tagging is required to two jets: both SECVTX and another algorithm are considered for this purpose. The latter is based on computing the combined probability of all tracks in the jet being consistent with originating from the primary vertex, and requiring this probability to be small ($< 5\%$); this cut has a higher acceptance for b decays than the full reconstruction of the secondary vertex by the SECVTX algorithm, and the rate is expected to be smaller than 1 Hz in both cases. They estimate that these triggers may be able to collect more than 20,000 $Z \to b\bar{b}$ decays, again enabling detailed studies of the energy scale for b jets.

While these results are preliminary, and there exists as of yet no detailed prediction of the decrease in systematic uncertainty a $Z \to b\bar{b}$ dataset may provide on the top mass measurement, we can look at the future of CDF top physics with optimism.

Acknowledgements

There is a large amount of people I need to acknowledge for the help, the suggestions, the support, and the encouragement they offered me during the three years I spent on the present work. I feel now afraid to forget some of them, but I need to make a try anyways.

First of all, I need to thank (and chances are that I will need to feed with a big seafood dinner in Venezia soon) all the members of the CDF-CMS group at Padova University.

Giovanni, my official advisor, needs infinite thanks for his amazing availability to discuss with me innumerable aspects of the analysis, for giving me practical consulence on technical issues, for solving problems I could not solve alone, and for just being there when I needed him, be it a cigarette, a candy, or a plot to discuss. I hope I will be allowed to work with him again in the future. Dario, the real boss of our group, was not very useful for the physics analysis, but I wish to thank him for the invaluable support he never denied me, allowing me to increase my knowledge attending schools and conferences during these three years of Ph.D. studies, and of course for having advised and encouraged me to continue to work in his group as a graduate student in the first place. Although he would deny it if you asked him, he always believed I had the right mixture of attributes to succeed in particle physics, and I appreciated and now miss his continuous teasing me, which was just his way of showing an affection that is truly reciprocated. Luca, my unofficial advisor, also needs a lot of thanks for his help on technical issues. I need to acknowledge him for having been the source of the most important ideas necessary to bring the $Z \to b\bar{b}$ search to a success. I was helped by the interaction and the advices of Luisa, who I thank and regret for her being unable to follow the last months of analysis, due to health problems. A lot of thanks goes to Maurizio, that helped me in various statistical and technical issues, and cheered me up with his fine sense of humor; for the same reason I need to acknowledge Jeff. I also wish to thank Andrea for his technical support on some issues and Nicola for the innumerable cigarettes I forced him to offer me.

Among the "crew" of post-docs, graduate students, undergraduates, and assorted animals of the CDF group that populated the scene at our headquarters in the Physics Department and at Fermilab during the last three years, I must pick Alberto and Matteo as the first to thank. Alberto is extraordinary both as a physicist and as a helping hand when you need one (he does first, asks questions later), and he substantially contributed to many aspects of the search for a $Z \to b\bar{b}$ signal. Matteo was just invaluable to help me—already rather dumb as a VMS user—start navigating in the UNIX sea, and was always willing to lose hours to solve my tiniest software problems, be it color-coding in lego plows under PAW or a particularly tough UNIX command. After these two, but just because of the extraordinary nature of the former, I can only list the names of all those who contributed to my mental sanity in these years. Annachiara, Chiara, Martina one and Martina two, Laura, Patrizia, Andrea, Devis, Gigi, Roberto, Alessandro, Carlo, Massimo, Luca, Antonio, Andrea "l'altissimo", Alvise, Giuliano, Giacomo, Fabio, Mapo...

Special thanks to Jennifer, who did a very professional work with the correction of the grammar, spelling, and style of the thesis. I learned a lot by just comparing her corrections to my drafts.

In acknowledging the people who helped me at Fermilab I will start with the VIPs. I have to particularly thank Bob and Mark, the Electroweak group conveners, for their invaluable and constant support in and out of the meetings. Many thanks also are due to Al and Franco, the spokespersons of CDF, who also supported me in various occasions and provided me with crucial reference letters that opened me the doors of a post-doc position at Harvard. Infinite thanks to Paolo, not so much for any particular help with my work, as for not having carried out his threats of wrecking my brand new Subaru Legacy, which is currently resting in the warmth of his garage during the cold winter nights. I also wish to thank him for unexplicably believing I'm a cut above average.

And now for the not-yet VIPs. First of all I wish to thank Gino, who has shown to be a very good friend: I enjoyed his company a lot in the desert land of the Tevatron. And I want to thank Marta for being a ray of sunshine. I must thank Fotis for the cigarettes and for many explainations on the taggers and, lately, on the CMX. Thanks also to all the other friends: Giusy, Laura, Maria, Robyn, Rocio, Simona, Gervasio, Juan, and may the others forgive me for being unable to cite them all.

Of course I must thank my lovely wife for having always been close and forgiving, and for her devotion in deciding to change her life to follow me in the U.S.. This thesis is dedicated to her.

Appendix A

The Search for $Z \to b\bar{b}$ Decays in the Inclusive Electron Dataset

We document in the present Appendix the studies we have made on a dataset containing electron candidates. This dataset turned out to be less advantageous to spot $Z \to b\bar{b}$ decays than the inclusive muon dataset, as described in App. B; however, it was the first one we analyzed in the search for the signal. We choose to present our studies on the electron dataset as an independent work, although it largely overlaps what has already been described in Ch. 3 to Ch. 9, because the small differences in the selection strategy, and the less refined methods to extract the signal, allow the reader—in our opinion—to view the matter from a different perspective.

A.1 Introduction

What is normally called "Inclusive Electron Dataset" is a data sample collected with triggers requiring the presence of an isolated electromagnetic deposit in the calorimeter, compatible with a CTC track pointing at it and (for most of the dataset) with signals in the CES chambers whose lateral profile resembled those typical of electron showers.

The run 1a dataset consists in $1,874,289$ events collected in the stream named COMBINED_ELE1_CEM, which is a sum of various Level 3 triggers. The run 1b dataset consists instead in $6,125,629$ events collected in the COMBINED_ELEB_CEM stream, again a sum of many different Level 3 triggers. The details of the composition of these datasets are explained elsewhere[70]; the principal triggers contributing to each dataset are described in Table A.1 and Table A.2. All in all, the total run 1 data amount to very nearly 8 million events.

A.2 Estimation of the Integrated Luminosity

To estimate the integrated luminosity the two datasets correspond to we use the offline utility package LUM_CONTROL, supplying as input the list of files con-

Trigger name	Description
ELE1_CEM_8_6*	A trigger requiring an e.m. deposit with $E_T > 8\ GeV$, a matched track with $P_T > 6\ GeV/c$, and tight ID requirements on the electron candidate.
ELE1_CEM_9_6*	A late version of the former, with a higher threshold on the e.m. E_T and the same P_T and ID cuts; it was active for the last 40% of the data taking.
ELE1_CEM_15_10	This trigger had higher thresholds on E_T and P_T than the former ones, but looser ID requirements; it was active for 89% of the data taking.
ELE1_CEM_6_4*	A low energy trigger, with the same tight identification criteria of the 8 and 9 GeV ones.
ELE4_CEM_6_4*	A late version of the above.
ELE1_CEM_20GEV_NOTRACK*	A trigger dedicated to the collection of W events, used for track-finding efficiency studies.
ELE1_CEM_25GEV_NOTRACK*	A higher E_T threshold version of the former.

Table A.1: *Description of the Level 3 triggers contributing to the 1a COMBINED_ELE1_CEM stream. The asterisks tag triggers that existed in different versions.*

tained in the tapes. For each of the two data taking periods there is a trigger whose contribution is largely dominant in the inclusive electron dataset: namely, CEM_9_SEED_9_SH_7_CFT_9_2 for run 1a and CEM_8_CFT_7_5_XCES for run 1b[1]. The luminosities integrated with these, according to LUM_CONTROL, amount respectively[2] to 18.87 pb^{-1} and 75.29 pb^{-1}. The total integrated luminosities corresponding to our file lists amount instead to 19.87 pb^{-1} (run 1a) and 87.67 pb^{-1} (run 1b)[3]. We may take as our estimate the averages of the two numbers for each data taking, and assign as a systematic error half their difference. That way we get $\int_{1a} \mathcal{L}dt = 19.35 \pm 0.45 \pm 0.7\ pb^{-1}$ and $\int_{1b} \mathcal{L}dt = 81.5 \pm 6.2 \pm 6.5\ pb^{-1}$, which can be combined into a total of $\int \mathcal{L}dt = 101 \pm 9\ pb^{-1}$.

[1]The fraction of events collected by these are 78.04% and 86.60%, respectively[70].

[2]The 1b number is significantly lower than the typical run 1b luminosity quoted in other analyses, because the main Level 2 trigger was prescaled dynamically to keep the Level 2 deadtime at a reasonable level.

[3]The luminosity estimates quoted for run 1b are those relevant for data collected with the SVX′ detector operating properly[48, 49]. In fact, we rely on its functioning for our data selection.

Trigger name	Description
ELEB_CEM_8_6*	The main trigger designed for the collection of central electrons; it required an e.m. deposit with $E_T > 8\ GeV$, a matched track with $P_T > 6\ GeV/c$, and tight ID requirements on the electron candidate.
ELEB_CEM_18_LOOSE*	This trigger had a $E_T > 18\ GeV$ cut on the electron cluster as well as a $P_T > 10\ GeV/c$ cut on the track momentum, but looser identification requirements than the previous one; it was active for the last 40% of the run 1b data taking.
ELEB_CEM_8*	A prescaled monitoring trigger which did not require any track pointing to the e.m. deposit.
ELEB_CEM_5_5*	A low energy trigger, with a threshold at 5 GeV for the E_T and P_T of the electron candidate, plus the usual tight identification requirements.

Table A.2: *Description of the Level 3 triggers contributing to the run 1b COMBINED_ELEB_CEM stream. The asterisks denote triggers that existed in different versions.*

A.2.1 The First Requirements

Even after the rather tight identification criteria for the electron candidates at Level 3, the dataset is still rich with fake electrons: tighter requests have to be applied to the electron identification variables to get a reasonably pure sample. Therefore, after a general cleanup, we apply the cuts listed in Table A.3 to obtain good electron candidates; we also reject electrons originated from identified photon conversions, using the algorithm described in Sec. 2.6.2. This selection reduces the dataset by about a factor of two. We then require the events to have at least one jet tagged with SECVTX, with positive decay length.

The requirement of at least one jet with a secondary vertex found by SECVTX (with positive decay length) with the options discussed in Ch. 3 selects $134,548$ events. In this appendix we will refer to this sample as the "SECVTX" dataset: it is the starting point of our search for $Z \to b\bar{b}$ decays with an electron in the final state.

Variable	Cut
Track P_T	$> 6\ GeV/c$
Cluster E_T	$> 7.5\ GeV$
$HAD\_EM_{3\times3}$	< 0.04
L_{shr}	< 0.2
Strip χ^2	< 10
Wire χ^2	< 10
Δx	$< 1.5\ cm$
Δz	$< 3.0\ cm$
SVX hits	not required

Table A.3: *Requirements used for the selection of tight electrons in the SECVTX sample.*

A.3 Other Datasets

A.3.1 The $Z \to b\bar{b}$ Monte Carlo Simulation

We used PYTHIA V5_7 with the MRSD− parametrization of the structure functions to generate $1,000,000$ signal events, choosing the Drell-Yan process and forcing the decay of Z bosons to b-quark pairs. The events were then passed through the usual QQ decay simulation (version 9.0) and the QFL′ detector simulation.

The dataset was then split according to the integrated luminosity of run 1a and run 1b and filtered through the chain leading to the SECVTX selection. The degraded version of SECVTX[37, 38] was used to model CTC track degradation (Sec. 2.6.4), while to include the effect of the Level 2 trigger acceptance for electrons passing the Level 3 requirements (modeled with the electron filter[4]) we used a parametrization of the efficiency curve as a function of the electron E_T obtained for the run 1a trigger by B. Wicklund and K. Byrum[71]; the curve needed to be shifted when used for the run 1b trigger, to account for the lower E_T and P_T thresholds of the latter[5]. The parameters of the efficiency curves are shown in Table A.4. They have an acceptance of about 70% (1a) and 80% (1b) for events containing electron candidates passing the Level 3 identification requirements.

After the whole machinery, 2493 events are left in the dataset (425 from run 1a and 2068 from run 1b). Of these, 596 have two SECVTX tags (104 from run 1a and 492 from run 1b). By applying the corrections for the SECVTX scale factor, as described in Sec. 2.6.4, we obtain a total of

$$N_{\geq 1SVX} = 2919 \pm 444,$$

[4]The Level 3 identification requirements are looser than those listed in Table A.3. For run 1a the wire χ^2 had to be lower than 15, the track residuals in x and z had to be respectively within 3 and 5 cm, and there was no $HAD\_EM_{3\times3}$ requirement; for run 1b the $HAD\_EM_{3\times3}$ fraction had to be lower than 0.125 and the track residuals in x and z were required to be within 3 and 5 (later 10) cm, respectively.

[5]We used the run 1b parameters obtained for the $B^0 - \overline{B^0}$ mixing analysis of M. Paulini *et alii*[72].

Trigger name	A_1	A_2	A_3	A_4	A_5
CEM_9_SEED_9_SH_7_CFT_9_2	0.927	6.75	4.25	9.24	0.622
CEM_8_CFT_7_5_XCES	0.927	6.18	4.20	7.48	0.504

Table A.4: *Parameters of the efficiency curves for the Level 2 electron triggers used in the simulation of run 1a and run 1b data. The efficiency is parametrized as* $\epsilon_{E_T} = A_1 * Freq((E_T - A_2)/A_3) * Freq((E_T - A_4)/A_5)$, where $Freq$ is the well-known CERN library function[50].

$$N_{2SVX} = 921 \pm 167.$$

To obtain the number of events expected from the $Z \to b\bar{b}$ process in the SECVTX selection we need to scale these numbers by the appropriate factors, to account for the integrated luminosity equivalent to our Monte Carlo datasets. Using the cross section derived in Sec. 3.2.1 ($\sigma_Z B(Z \to b\bar{b}) = 1.110 \pm 0.031 \pm 0.090\ nb$) and the estimated integrated luminosity of the inclusive electron dataset, we predict to have collected a total of 327 ± 58 $Z \to b\bar{b}$ decays in the SECVTX selection, 103 ± 21 of whose should have two SECVTX tags with positive decay length.

A.3.2 A QCD Monte Carlo Dataset

For some of our background studies we also perused a large dataset generated with PYTHIA V5_6 by M. Paulini *et alii*[6] for their B-mixing analysis. The dataset comes from a generation of several hundred million events of direct QCD $b\bar{b}$ production, containing electrons required at generator level and filtered using the same cuts described in Table A.3; the events are also filtered by requiring at least one SECVTX tag with the same options we use. These requirements, along with the request that the electron belongs to the tagged jet, reduce the sample to about $100,000$ events. We cannot use this dataset for comparisons to our experimental dataset, since we do not require our events to have the electron in the same jet as the secondary vertex. We can use the QCD Monte Carlo sample only when a tighter selection is made, requiring that the events contain two jets both tagged by SECVTX: in that case the matching of the electron with a SECVTX tag is automatically satisfied.

A.3.3 Adding Minimum Bias to the Monte Carlo Data

In App. C we describe some detailed comparisons of the multiple interactions accompanying the hard interaction in the SECVTX sample with those present in a minimum bias dataset. These studies prove that, in order to model in the Monte Carlo simulations the effect of pileup due to high instantaneous luminosity conditions of data taking, it is sufficient to add the contents of the calorimeter banks of each Monte Carlo event with those of a real minimum bias event, taking care to match the luminosity distribution of the added minimum bias events to that of the dataset one wishes to compare the Monte Carlo events to.

[6] The dataset is described in detail in Ref.[59].

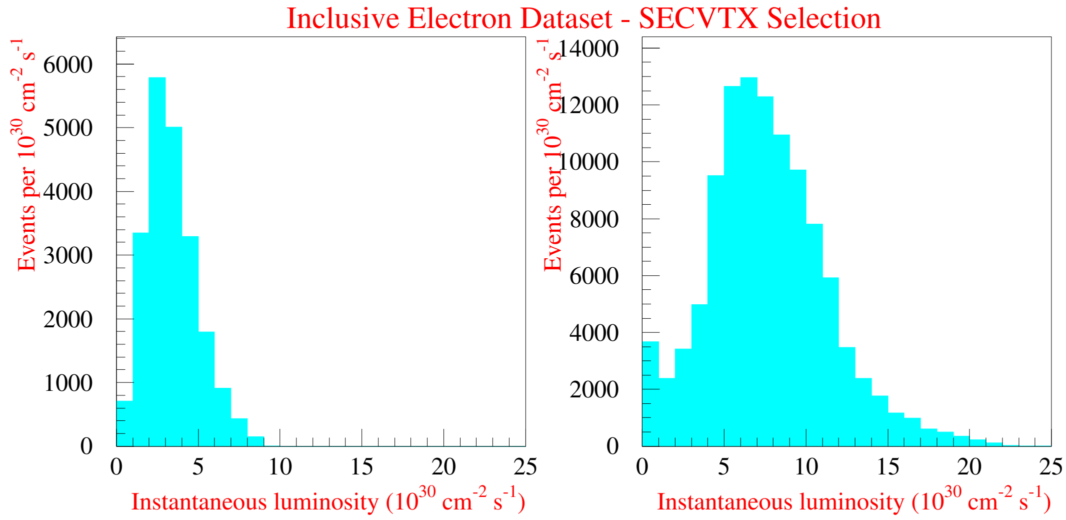

Figure A.1: *Instantaneous luminosity distribution of the SECVTX data from run 1a (left) and run 1b (right). These distributions were used to select suitable samples of minimum bias data for mixing purposes.*

We thus decided to mix the calorimeter banks in our $Z \rightarrow b\bar{b}$ Monte Carlo datasets at the SECVTX level with those of minimum bias data, according to the instantaneous luminosity distributions of the 1a and 1b inclusive electron events at the SECVTX level (see Fig. A.1).

After the mixing, the events were completely reconstructed. This procedure correctly accounts for the effects of multiple interactions in the jet reconstruction; we will deal in Sec. A.5 with the effect of the mixing of calorimetry informations on the kinematic variables we have studied.

The same procedure described above was performed on the additional Monte Carlo samples we use for our studies: QCD $b\bar{b}$ events with two SECVTX tags and $Z \rightarrow e^+e^-$ simulated events.

A.4 Event Selection

The most outstanding difference between simulated Z events and experimental data at the SECVTX level, apart from the invariant mass of the leading jets, is the amount and pattern of their radiation. We have already discussed in detail in Sec. 4.2 this issue; here we just point out that the jet multiplicity distribution itself is very different. If we count jets with uncorrected E_T greater than 15 GeV and pseudorapidity $|\eta_d| < 2.0$, we can compare the jet multiplicity of data and Monte Carlo. We find that Z events almost always possess just two such jets, while the experimental data (that one may safely consider to be pure background at this level) have a much broader multiplicity distribution, both to lower and to higher values

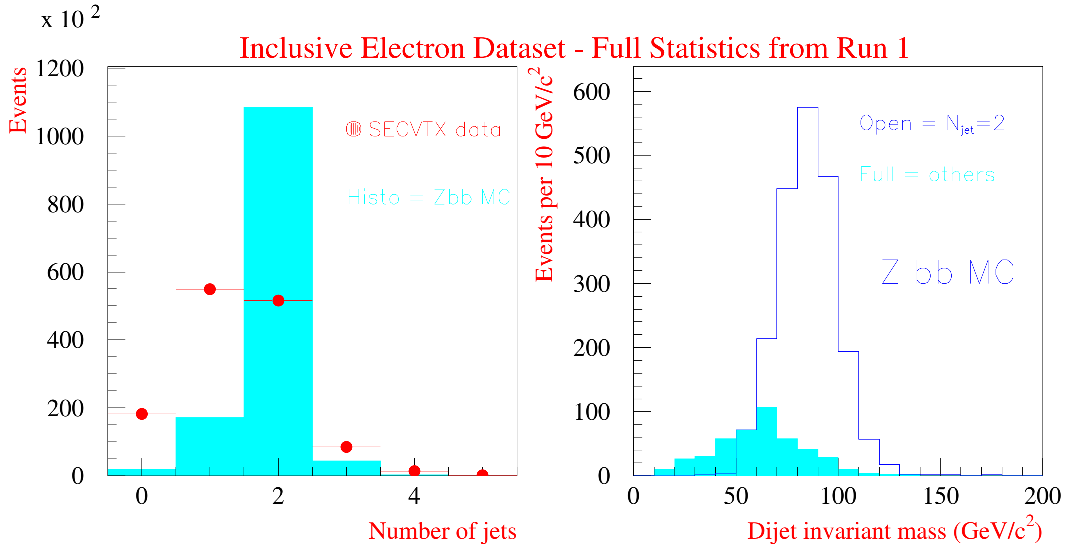

Figure A.2: *Left: distribution of the number of jets for experimental data and for $Z \to b\bar{b}$ events at the SECVTX level. Right: comparison between the invariant mass reconstructed for $Z \to b\bar{b}$ events with exactly two jets and for $Z \to b\bar{b}$ events with a different jet multiplicity. The two highest E_T clusters within $|\eta_d| < 2.0$ are used to evaluate the invariant mass.*

of N_{jet} (Fig. A.2, left). Furthermore, if we take the two highest E_T jets and form an invariant mass with them, after correcting their energy with the standard CDF correction routine (see Sec. 2.6.1), we obtain the distributions shown in the right plot of Fig. A.2. One can notice how only from those Z decays having exactly two jets with the described characteristics one can obtain a reasonable mass peak, a feature of paramount importance for our search. The invariant mass of events with more jets, or events where one of the two leading jets has E_T lower than 15 GeV or high pseudorapidity, hardly makes it to the Z pole.

Following the selection strategy outlined for the muon dataset in Sec. 4.1, we study the double SECVTX tagging rate in the experimental data and in the Monte Carlo sample. We find that only 6085 events out of $134,548$ have a second SECVTX tag in the data (4.5%), as compared to 921 out of 2919 in the simulation (31.6%); when we only consider events with $N_{jet} = 2$, the double tagging fraction becomes $3435/51,931 = 6.6\%$ in the data and $828/2390 = 34.6\%$ in the Monte Carlo[7]. We

[7]The numbers quoted for the double tagging rate before the $N_{jet} = 2$ requirement show that the tagging rate of inclusive electron events is similar to that of inclusive muon events, in agreement to our supposition that the sample composition is similar in the two datasets (sample composition studies of the electron dataset have been performed just as accurately as for the muon dataset as described in Sec. 4.1). Note that the small defect in the double tagging rate of inclusive electrons

therefore choose to restrict our search to events with just two jets[8], both carrying a secondary vertex with positive decay length: the signal to noise ratio in this sample should be close to 1/37.

Sample	run 1a	run 1b	$Z \to b\bar{b}$ MC	Eff.	Events in 101 pb^{-1}	N/S all
Initial			1,000,000		112,100	
Trigger	1,874,289	6,125,629				
SECVTX	21,443	113,105	2,919	0.25%	327.3	411
$N_{jet} = 2$	8,274	43,657	2,390	81.9%	267.9	194
$N_{tags} = 2$	527	2,908	828	34.7%	92.8	37

Table A.5: *Number of events selected by each of the cuts described in the text for the experimental data and for the $Z \to b\bar{b}$ Monte Carlo, efficiency of each cut for the Z signal, expected Z events in the total dataset, and Noise/Signal ratios.*

A.5 Kinematical Tools for the Selection of the Signal

Figure A.3 shows the mass specrum for events with two SECVTX tags. The good agreement with the Monte Carlo simulation of direct $b\bar{b}$ production and with experimental events with only one tag (that contain a fraction of signal about four times lower) is evident, and we still cannot see any hint of the presence in the experimental data of Z decays, that should cluster at a mass[9] of about 87 GeV/c^2.

The smallness of the S/N ratio after double SECVTX tagging forces us to try and use very drastic selection criteria in order to understand if some signal is present in the dataset. One may therefore try to push the selection to the highest value of signal/noise attainable, by selecting $\Delta\Phi > 3$ and picking events with $\Sigma_3 E_T = 0$,

(4.5% with respect to the value quoted in Sec. 4.1: 5.2%) is also not unexpected: as we note in App. B, jets containing a muon candidate must pass the SECVTX jet energy threshold (10 GeV) without the help of the lepton energy in order to be tagged, unlike jets of the electron dataset: electrons leave all their energy in the calorimeter, unlike muons. Because of this difference, the tagging rate of the electron sample is *smaller*, since they contain more jets where almost all the energy is due to the lepton.

[8]The two-jet cut has no effect on the data once the cut on $\Sigma_3 E_T$ is applied (see below, or Sec. 4.2), the latter being much more restrictive than the former. We mention it here since we are describing in this appendix our analysis of the inclusive electron sample just as we did it, step by step. By doing that, rather than just quoting the final results, we can use the description of our work on the electron dataset to give the reader a different perspective, and add information even if it proved unuseful in the end.

[9]The reconstructed mass is expected from Monte Carlo to be lower, on average, than the nominal Z mass, because of the presence in the electron jet of a neutrino from b decay. The effect is actually not so large as that we found in muon events (Ch. 5), due to the competitive effect of the electron cluster: the electron leaves all its energy in the e.m. calorimeter, but the standard jet corrections treat it as a normal hadronic jet, multiplying its energy by a factor of about 1.3. This almost completely cancels the energy lost because of the neutrino.

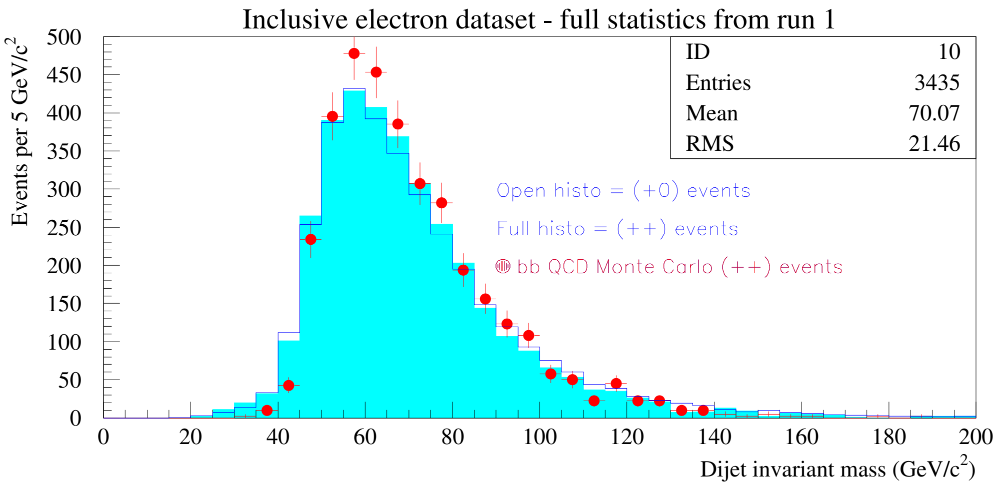

Figure A.3: *Invariant mass of the two jets in the* 3,435 *events having two positive SECVTX tags and* $N_{jet} = 2$ *(full histogram), compared to the invariant mass of simulated* $b\bar{b}$ *QCD events with the same cuts (points with error bars) and to the invariant mass of events having* $N_{jet} = 2$ *but only one SECVTX tagged jet, the second one being inside the SVX acceptance and with two good SVX tracks (empty histogram).*

id est events with no other energy cluster beyond the two tagged jets in the whole calorimeter. Let us take a look at the mass distribution of these few survivors: it is shown in Fig. A.4 (upper left) for double tags and for events with one tagged jet and one taggable jet, a sample that should give a good representation of the background—if, as we believe, the tag rate of Z events is five times higher than that of the QCD background. We indeed note the presence of a suggestive peak at the value where the Z signal is supposed to sit. Loosening up the cut on $\Sigma_3 E_T$ to higher values gives the other spectra shown in Fig. A.4: the shoulder eventually gets wiped out, not unexpectedly, when the background fraction becomes too high. We regard the cluster of events around the Z pole as evidence for the $Z \to b\bar{b}$ signal in our data, and will spend the rest of this appendix to try and estimate its amount and its compatibility to the expectations. To do that we must first of all study the efficiency of the $\Sigma_3 E_T$ cut.

A.6 Studies of the Energy Flow

There are two ways to obtain an estimate for the efficiency of a cut on this variable. The first relies on taking real $Z \to e^+e^-$ events, and substituting the final state electrons with two jets from b quarks by using the Monte Carlo for the modeling of

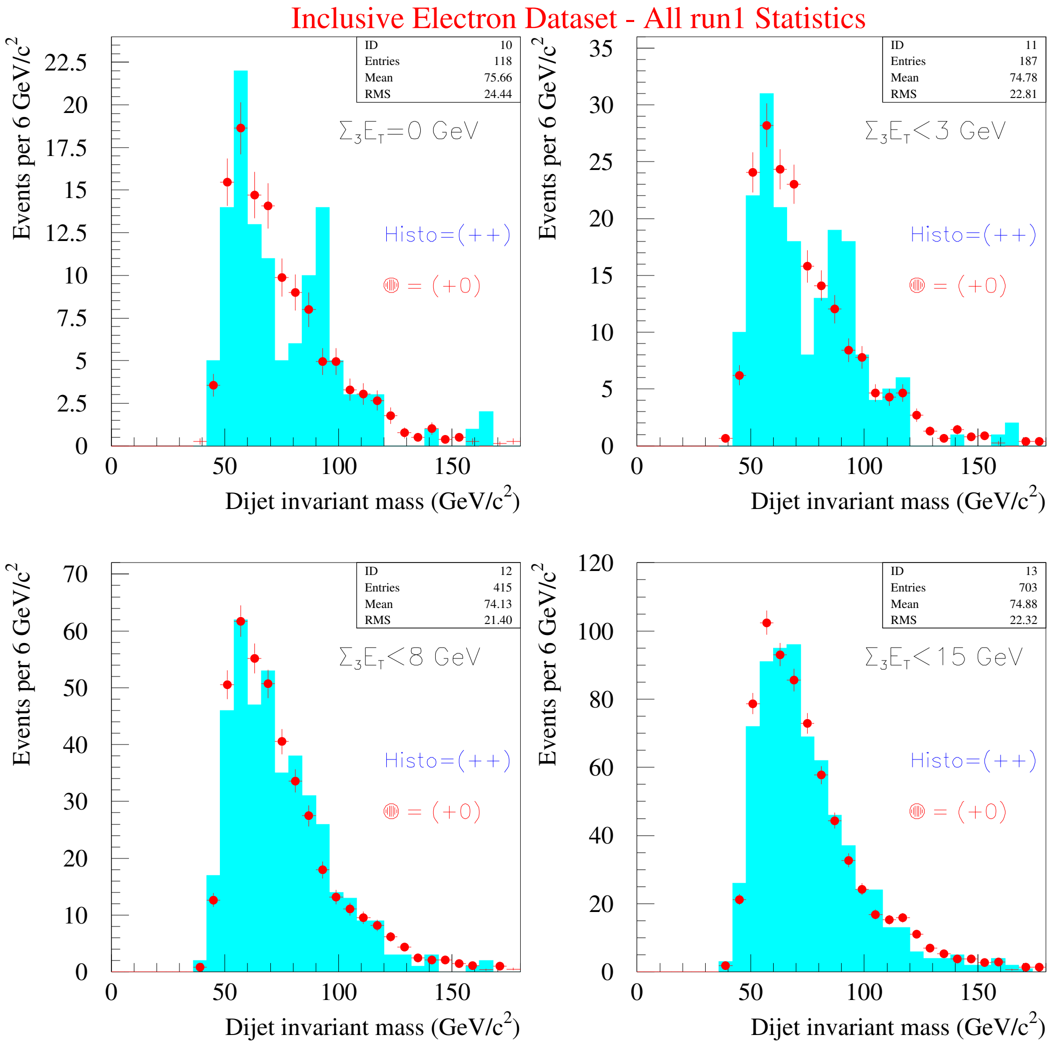

Figure A.4: *The mass spectra of double tagged events (histograms) are compared with those of single tags with the second jet defined as taggable by SECVTX (points with error bars). All the data pass the N_{jet} and $\Delta\Phi_{12}$ requirements, and they also have $\Sigma_3 E_T = 0$ GeV (upper left), < 3 GeV (upper right), < 8 GeV (lower left) and < 15 GeV (lower right).*

the final state. The distribution one gets for $\Sigma_3 E_T$ is then a good approximation to that of the signal, since the Monte Carlo should give a good simulation of final state radiation. We have used that method in Sec. 8.2 for a cross section measurement in the inclusive muon dataset. Here we wish instead to try and model the energy flow in the data from comparisons to the Monte Carlo datasets. This second approach relies upon comparing a QCD direct production Monte Carlo to the experimental data, parametrizing the difference in the $\Sigma_3 E_T$ spectrum with some simple function, and then applying that parametrization to the $Z \to b\bar{b}$ Monte Carlo.

The histogram in the upper right of Fig. A.5 shows the $\Sigma_3 E_T$ distribution for the

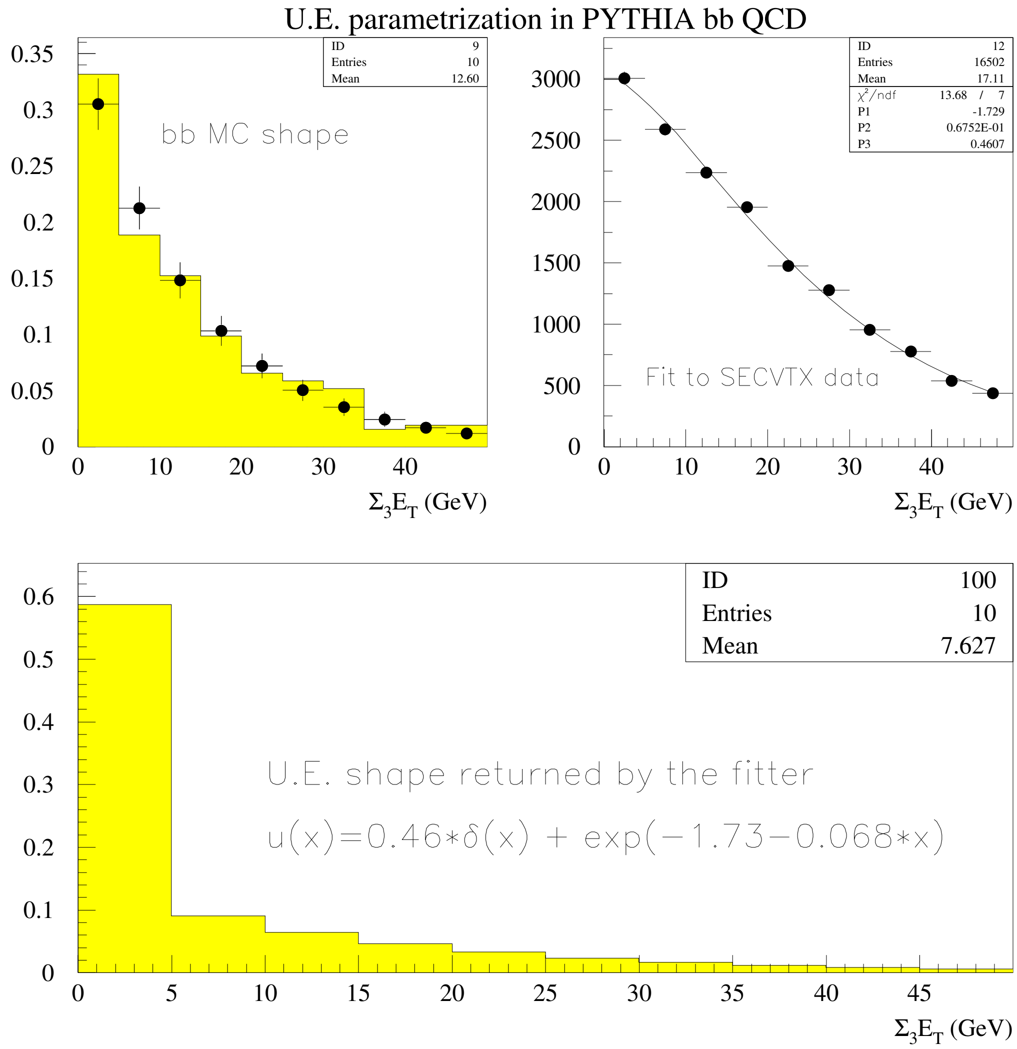

Figure A.5: *Upper left: the histogram shows the $\Sigma_3 E_T$ distribution for PYTHIA direct $b\bar{b}$ events (QCD) with a tight electron and a SECVTX tag, and with $N_{jet} = 2$ and $\Delta\Phi_{12} > 3$. The dots are the exponential fit to the histogram, and they define the $f(x)$ function used in the fit. Right: the dots show the $\Sigma_3 E_T$ distribution for the experimental data passing the same cuts just described for the Monte Carlo, and the line is the fit obtained in the way described in the text. Bottom: the returned shape for $u(x)$ from the fit.*

direct $b\bar{b}$ Monte Carlo after the mixing of minimum bias data described in section A.3.3, for events with two SECVTX tags and passing the requirements $N_{jet} = 2$ and $\Delta\Phi > 3$. The mean value, 12.6 GeV, is much lower than that of the distribution for the experimental data shown on the right (17.1 GeV). If we interpret the difference as a bad modeling of the underlying event[10] in the Monte Carlo, we can try to extract that contribution to the energy flow in these events from the two distributions. Let us call $f(x)$ the $\Sigma_3 E_T$ distribution for the Monte Carlo and $g(x)$ that of the experimental data, and $u(x)$ be the $\Sigma_3 E_T$ contributions from the effects that our Monte Carlo is unable to model. Then the following relation should hold:

$$g(x) = \int_0^x f(t)u(x-t)dt. \tag{A.1}$$

That is to say, a true event with a value x of $\Sigma_3 E_T$ must correspond to a simulated event whose $\Sigma_3 E_T$ was t with probability $u(x-t)$, the latter being the probability distribution that an event gets a $x-t$ contribution to the measured energy flow from the underlying event.

To fit a $u(x)$ distribution one must first assume a functional form for it. We have tried some simple functions to describe it, and found that a reasonably good description of the data was obtained with the use of the following:

$$u(x) = \alpha\delta(x) + exp(\beta + \gamma x). \tag{A.2}$$

That is, we assume that a fraction α of the events will get no contribution at all to the energy flow from underlying event as we measure it with our variable: this both allows to factor in the discrete threshold in the clustering algorithm (a cluster is formed by JETCLU only around calorimeter towers having more than 1 GeV of transverse energy) and allows for zero radiation processes. If we fit the two distributions with this function, we get the results shown in Fig. A.5. There, the $\Sigma_3 E_T$ distribution for the Monte Carlo has been formerly smoothed by an exponential fit (shown by the black dots overlayed to the histogram). The fit to the SECVTX data is all in all satisfactory, the χ^2/dof being lower than 2 notwithstanding the high statistics of that sample.

Prior to using the $u(x)$ form in our $Z \to b\bar{b}$ Monte Carlo, we will check how it scores on $Z \to e^+e^-$ data. We generate with PYTHIA a sample of $Z \to e^+e^-$ events, we add minimum bias to it, according to the recipe already described in Sec. A.3.3, and we compute the $\Sigma_3 E_T$ distribution for events that pass the same selection cuts applied to the SECVTX data: that is, we treat electron clusters as calorimeter clusters, and select events with $N_{jet} = 2$ (the two electron clusters must be therefore accompanied by no jet with $E_T > 15$ GeV and $|\eta_{det}| < 2.0$), with $\Delta\Phi > 3$ between the two "jets". We then obtain the distribution shown in the lower left plot of Fig. A.6. The same cuts, applied to real $Z \to e^+e^-$ events, give the distribution on the lower right on the same figure. We can then use the $u(x)$ parameters returned from the fit to QCD data to get a corrected $\Sigma_3 E_T$ distribution, which can be compared to the experimental data (Fig. A.6, lower right). The agreement is

[10] In what follows we will generically—and rather imprecisely—call "underlying event" all those effects that we believe our Monte Carlo does not model correctly when used with default values for the relevant parameters: underlying event, double/triple parton scattering, initial state radiation. We exclude from this list the effect of multiple interactions (a.k.a. "pileup"), since we believe that effect has been correctly modeled by the mixing procedure.

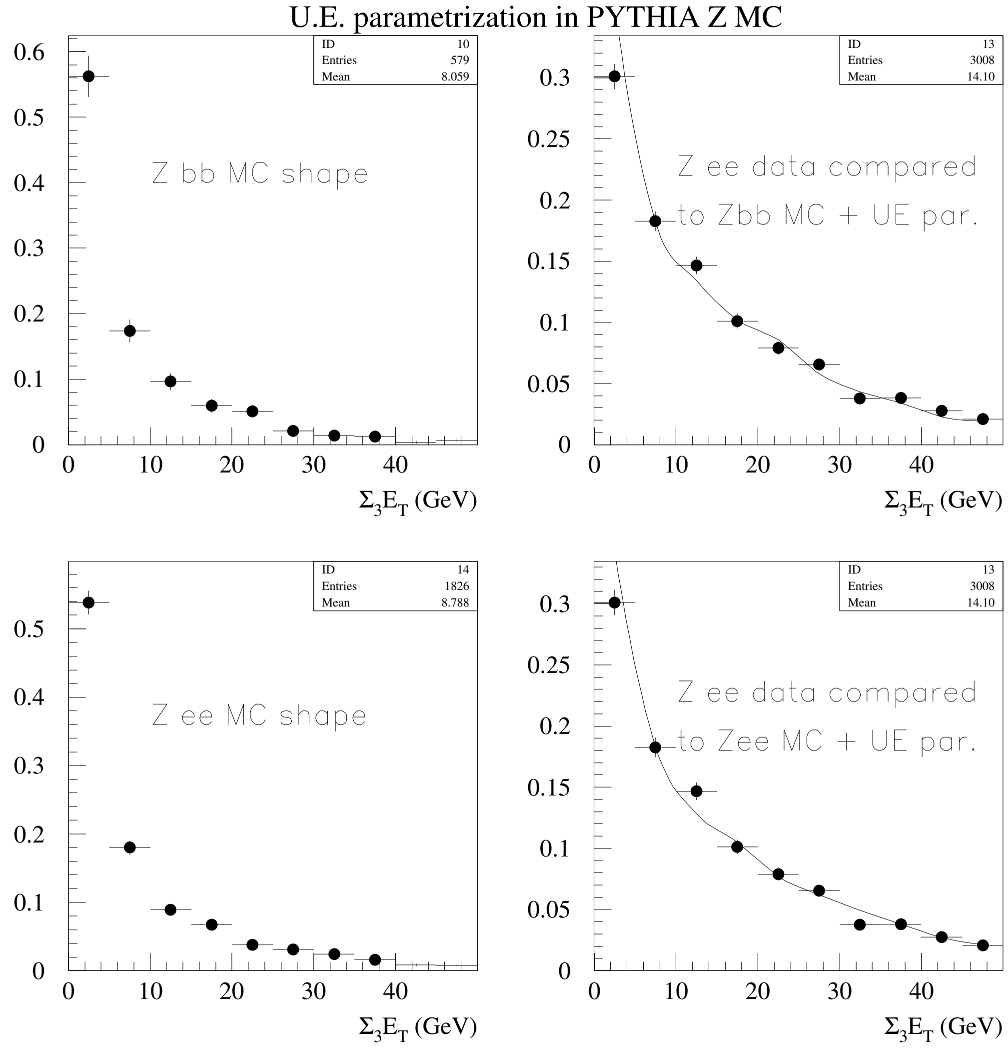

Figure A.6: *Upper left: shape of the $\Sigma_3 E_T$ distribution for PYTHIA $Z \to b\bar{b}$ events with a tight electron and a SECVTX tag, and with $N_{jet} = 2$ and $\Delta\Phi_{12} > 3$ (events have been mixed with minimum bias as explained in Sec. A.3.3). Upper right: comparison between the $\Sigma_3 E_T$ distribution for experimental $Z \to e^+e^-$ data (points) and the distribution returned from the fit (line), once the three parameters defining $u(x)$ have been fixed to the values found in the fits of the QCD events. Bottom: the same distributions for the $Z \to e^+e^-$ Monte Carlo events (left), and the results of the application of $u(x)$ to them (line) compared to experimental $Z \to e^+e^-$ data (right).*

good: this tells us that roughly the same amount of additional radiation is needed in a simulated electroweak production and in a simulated QCD production to get a good match to the data. In other words, what the Monte Carlo is not able to simulate with sufficient accuracy appears to be rather process-independent. We can now apply the same machinery to $Z \to b\bar{b}$ events, selected with the same tools. What we get is shown in the upper right in Fig. A.6. There one notices that no real difference between the obtained $\Sigma_3 E_T$ spectrum and that from experimental $Z \to e^+e^-$ events is evident: final state radiation is what should tell these two datasets apart, which suggests that probably it is not a very important effect in the $Z \to b\bar{b}$ process, once one selects events with two back-to-back jets and no other jet in the calorimeter.

The work just described suggests that for a preliminary evaluation of the efficiency of a cut on the $\Sigma_3 E_T$ one may rely, to first approximation, on the experimental $Z \to e^+e^-$ data. The efficiency of the very tight selection of events with $\Delta\Phi_{12} > 3$ radians and $\Sigma_3 E_T < 3\ GeV$ is only $17.6 \pm 0.7\%$ on the $Z \to e^+e^-$ signal (after the requirement $N_{jet} = 2$): we therefore expect to have collected 16 ± 3 events in the final sample[11].

A.7 Some Simple Evaluations of the Excess

Having obtained some qualitative evidence of the presence of Z decays in the dataset with the use of very restrictive cuts by examining the mass distribution, we now proceed to try and get an estimate for the number of signal events collected there: to that purpose we need a good understanding of our background.

A.7.1 Background Studies

The most natural choice for a background sample is represented by "(+0)" events: we call that way events passing all of our selection cuts ($N_{jet} = 2$, $\Delta\Phi_{12} > 3$, $\Sigma_3 E_T < 3\ GeV$: our "signal region") except for the $N_{tags} = 2$ requirement, that is events where only one of the two jets has a positive SECVTX tag, the other just being inside the SVX acceptance and with enough high P_T tracks to get a reconstructed secondary vertex (a "taggable" jet[12]). Events with these characteristics are as similar as possible to our double tagged data and should contain a much lower fraction of signal, by virtue of the higher probability of tagging the latter. But that is not the only possibility. We can also use events that have both jets defined as taggable (we call these "(00)" events), although this choice selects events with a much lower b quark content: *a posteriori*, the lower b purity does not seem to harm the mass distribution. A third choice relies in picking (++) events that fail the $\Sigma_3 E_T$ cut: they are kinematically very similar to the data in the signal region, but contain a much smaller fraction of signal: therefore, by selecting events with two SECVTX tags, $\Delta\Phi_{12} > 3$ radians, and $3 < \Sigma_3 E_T < 40\ GeV$, we obtain another good background sample[13].

[11] In Sec. A.8.1 we will also need an estimate for the expected number of $Z \to b\bar{b}$ events passing the looser cuts $\Sigma_3 E_T < 10\ GeV$, $\Delta\Phi > 3$: the efficiency is $27.2 \pm 0.8\%$.

[12] In SECVTX jargon, a taggable jet has at least two tracks featuring two hits in the SVX layers and $P_T > 1.5\ GeV/c$, or three or four hits and $P_T > 0.5\ GeV/c$; see also Sec. 2.6.4.

[13] The cut on $\Delta\Phi_{12}$ is necessary because it is correlated to the invariant mass.

In order to obtain a quick and dirty evaluation of the excess of events around the Z mass value, we plot (left side of Fig. A.7) in 25 GeV/c^2 bins the mass spectrum for the (++) events in the signal region , and define our search region as the fourth mass bin ($75 \div 100$ GeV/c^2). There we count 62 events. To estimate the background with the help of (+0) events selected in the signal region, we normalize their mass spectrum outside the search bin to the number of (++) events in the same mass interval ($0 \div 75$ GeV/c^2 + $100 \div 200$ GeV/c^2). We can then read the predicted magnitude of the background in the search bin, and extract an excess. We thus get $N_{bgr} = 41.4 \pm 4.5$, from which $N_{exc} = 20.6 \pm 9.1$ events.

For our second background sample, *id est* events from the (00) sample that pass all the selection criteria, we obtain a mass distribution astonishingly similar to the former one. By means of the same normalization mechanism, we expect in the search bin $N_{bgr} = 43.7 \pm 4.0$ events, which translates to an excess of $N_{exc} = 18.3 \pm 8.9$ events.

The same game can be played with the help of (++) events that fail the $\Sigma_3 E_T$ cut. We predict this time $N_{bgr} = 41.4 \pm 4.8$ events and get an excess exactly equal to the first one: $N_{exc} = 20.6 \pm 9.2$ events. And, again, the background shape agrees impressively well with the former two, as is evident in Fig. A.7.

Since the three background samples are completely uncorrelated, containing as they do different events, we can combine their predictions by simply computing the weighted average of the expected number of events in the search region. We thus obtain $\bar{N}_{bgr} = 42.3 \pm 2.5$ and the excess becomes $N_{exc} = 19.7 \pm 8.3$ events. We should note that all these evaluations are conservative by construction: they neglect the signal present in the background samples in the search bin, and they also neglect Z events falling outside the region $75 \div 100$ GeV/c^2, which are used in the background normalization. These considerations are however not important, in view of the high uncertainty of the background predictions and the approximation of the method itself.

We have various ways to check that the three choices of background just described give a good description of the mass spectrum of QCD events in the signal sample. The three extrapolations we have described involve, respectively:

1. the hypothesis that (+0) and (++) events behave the same way inside the signal region;

2. the hypothesis that (00) and (++) events behave the same way inside the signal region;

3. the hypothesis that (++) events failing the $\Sigma_3 E_T$ cut behave the same way as (++) events that pass the cut, *id est* the belief that the value of $\Sigma_3 E_T$ does not bias the mass spectrum.

We can check each and every one of these extrapolations by predicting:

1. the mass shape of (++) events *failing* the $\Sigma_3 E_T$ cut with (+0) events in the same region;

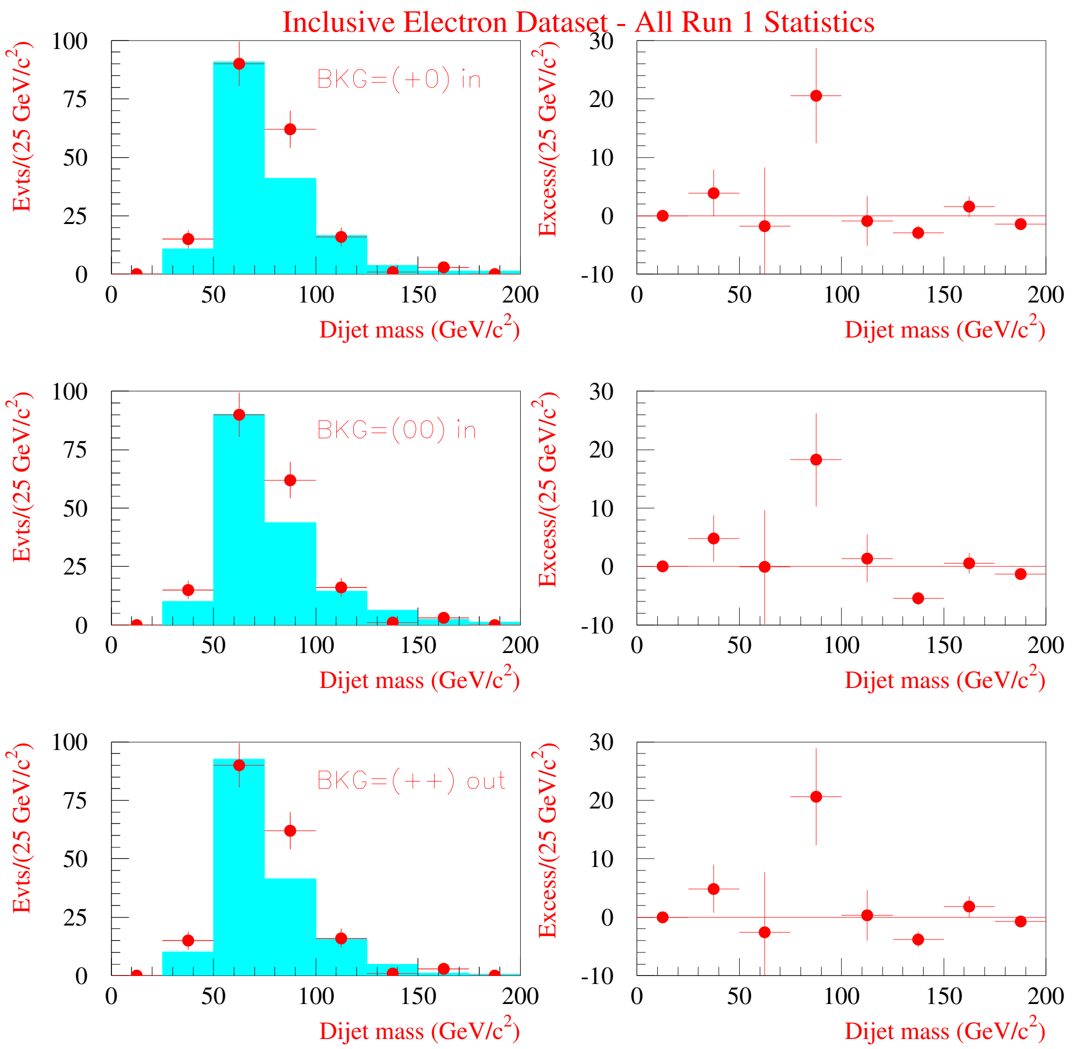

Figure A.7: *On the left side, the mass spectrum of (++) events in the signal region (black dots) is compared to the three choices of background described in the text (full histograms): (+0) events in the Signal Zone, (00) events in the Signal Zone, and (++) events that fail the* $\Sigma_3 E_T$ *cut. On the right side are plotted the differences between (++) events in the Signal Zone and the three background samples, after a normalization of the histograms outside the bin* $75 \div 100$ GeV/c^2.

2. the mass shape of (++) events *failing* the $\Sigma_3 E_T$ cut with (00) events in the same region;

3. both (*a*) (+0) events *passing* the $\Sigma_3 E_T$ cut, and with (+0) events *failing* the cut, and (*b*) (00) events *passing* the $\Sigma_3 E_T$ cut with (00) events *failing* the cut.

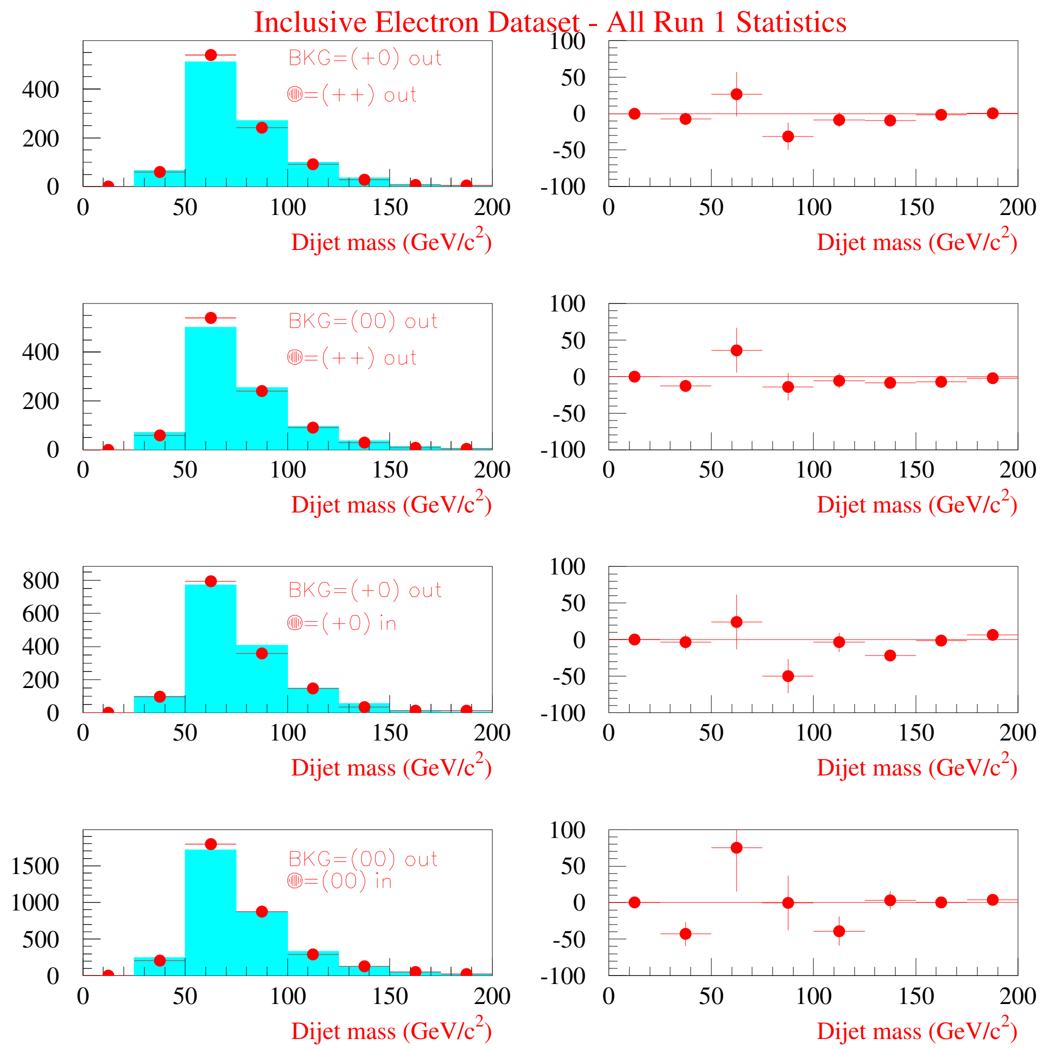

Figure A.8: *The plots on the left show the agreement in the mass spectra between the various samples used for the background predictions described in the text. The plots on the right show the "excesses" of events in each mass bin from the comparisons shown in the plots on the left.*

These four checks (1, 2, 3(a) and 3(b)) are shown graphically in Fig. A.8: the dotted histograms on the left are the spectra to be predicted, and the full histograms are the normalized backgrounds (the normalization is done exactly as described above, *i.e.* excluding the fourth bin). On their right we show the deviations from the predicted backgrounds: these are compatible with originating from statistical fluctuations in the templates, since the distribution of the absolute value of the normalized deviations from the predictions has a mean of 1.08, in very nice agreement to what would be expected (1.0) for a Gaussian distribution (see Table A.6). In the mass bins of our interest, all four predictions are in defect, although compatible with zero in two of the four cases. From the agreement of these spectra we may conclude that the extrapolations are affected by small systematic uncertainties, estimable very conservatively at the 10% level. We can also estimate the systematics due to the presence of signal in the three background samples to another 10%[14]. A $10 \oplus 10 \simeq 15\%$ shift in the three predictions described above corresponds to an increase of the error in the excess over the weighted average background to 10.4 events. Our final result with this method is thus $N_{exc} = 19.7 \pm 10.4$ events.

Check type	$25 \div 50$ GeV/c^2 (% error)	$50 \div 75$ GeV/c^2 (% error)	$75 \div 100$ GeV/c^2 (% error)	$100 \div 125$ GeV/c^2 (% error)
$(++)_{out}$ with $(+0)_{out}$	-12.3 ± 14.0	4.9 ± 5.7	-12.9 ± 7.9	-9.4 ± 11.5
$(++)_{out}$ with $(00)_{out}$	-21.8 ± 14.5	6.7 ± 5.7	-5.7 ± 7.8	-6.1 ± 11.5
$(+0)_{in}$ with $(+0)_{out}$	-3.8 ± 11.0	3.0 ± 4.7	-14.0 ± 6.5	-2.5 ± 9.1
$(00)_{in}$ with $(00)_{out}$	-21.1 ± 8.3	4.2 ± 3.3	0.0 ± 4.3	-13.5 ± 6.9

Table A.6: *Deviations (in percentage w.r.t. the number of events in the sample to be predicted) from the predictions (1), (2), (3a) and (3b) described in the text, for the four central bins of the mass spectra.*

A.7.2 A Counting Experiment Approach

A less crude approach to the problem of estimating the background consists in computing an absolute prediction for the number of (++) events from the number of events just failing the kinematical requirements. If we assume that the signal we try to extract resides mainly in the region of low $\Sigma_3 E_T$ and high $\Delta\Phi_{12}$ of the (++) sample, and is much less significant outside it, then, in the hypothesis that QCD events in the (++) and (+0) samples are kinematically equivalent, we can estimate the tag probability from the ratio of (++) over (+0) events just outside the region where the signal sits, and extrapolate it inside, obtaining a background estimate:

$$N^{++}_{exp,in} = N^{+0}_{obs,in} \times (N^{++}_{obs,out}/N^{+0}_{obs,out}). \tag{A.3}$$

This procedure can be carried out after dividing the two datasets into coarse bins of invariant mass, very much like what has been done formerly: that way we obtain a tag rate as a function of mass, and a prediction that can be checked bin by bin in the mass histogram. An excess around the Z mass will be bias-free evidence for the signal, if away from the Z pole the excesses are compatible with zero.

We define a Signal Zone as the region ($\Sigma_3 E_T < 3\ GeV$, $\Delta\Phi_{12} > 3$) in the plane of these two variables, and a Normalization Zone as the larger region ($\Sigma_3 E_T <$

[14]The number of signal events in each of the three background samples may be computed with the help of Monte Carlo predictions, and is well under the 10% level in the search bin ($75 \div 100\ GeV/c^2$).

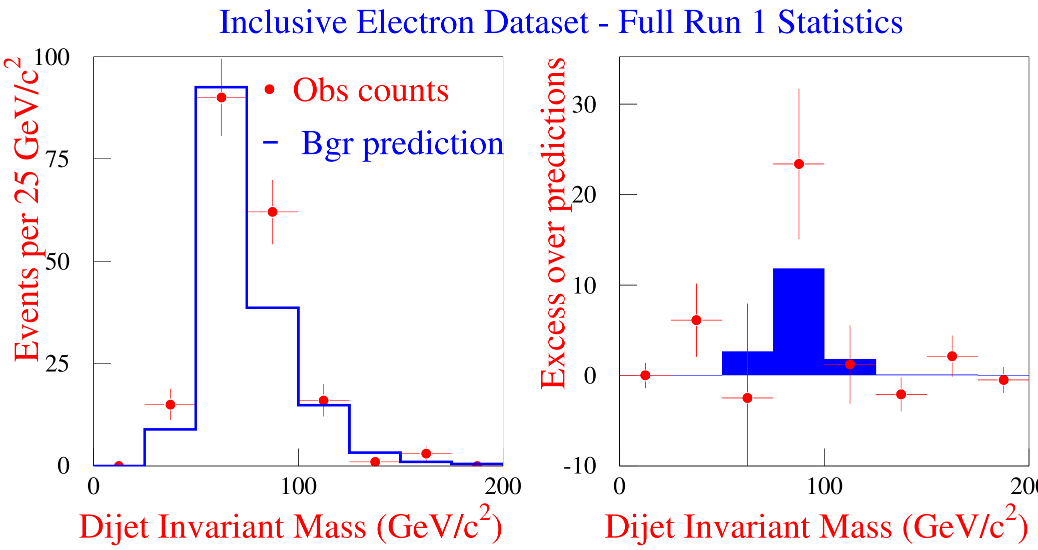

Figure A.9: *Left: comparison between the mass spectrum of (++) events in the signal region ($\Sigma_3 E_T < 3GeV$, $\Delta\Phi > 3$) and the absolute background prediction described in the text. Right: excess of (++) events with respect to the prediction. The full histogram shows the size of the expected signal.*

40 GeV, $\Delta\Phi_{12} > 2.7$) not included in the Signal Zone[15]. We can then compute the tag probability as a function of the invariant mass of the two leading jets in the Normalization Zone, and obtain a background prediction in the Signal Zone which can be compared to the observed number of double tags.

The results of the computation are shown in Fig. A.9 and in Table A.7. One clearly notices that a sizable excess is again present only in the fourth mass bin, where we expect a contribution from Z decay, surrounded by bins where the prediction is in good agreement with the observed number of double tags. By this method we get an excess amounting to $N_{exc} = 23.4 \pm 8.3$ events. The six other bins[16] may then be used to estimate a systematic uncertainty: a constant fit gives an average shift of -0.46 ± 0.30 events per bin. A systematic error due to the arbitrary choice of the Normalization Zone may also be evaluated by varying its boundaries in the intervals $25 \div 50$ GeV and $2.5 \div 2.9$ radians: the predicted background in the search bin ($75 \div 100$ GeV/c^2) varies in the range $38.2 \div 39.0$ events, and we therefore assign another ± 0.4 events uncertainty to the estimate. Our background prediction then becomes $38.6 \pm 2.7 \pm (0.5 \oplus 0.4) = 38.6 \pm 2.8$ events. The probability of a statistical fluctuation of this number to 62 or more events is 0.1%. However, as is evident in Fig. A.9, the expected signal is about two times lower than the observed excess, its

[15] Events in these two regions are not too different kinematically, so we expect the tag probability to be the same—something that can be successfully checked by using (+0) and (00) events. The outer boundaries of the Normalization Zone cannot be pushed too far away, however, otherwise background processes with different tag rates might start to contribute.

[16] The bin $0 \div 25$ GeV/c^2 adds of course no information.

size being estimated to be about 11 ± 3 events in the search bin[17].

Mass interval	(++) events	Prediction	Excess	Probability
$0 \div 25\ GeV/c^2$	0	0.0 ± 0.0	0.0 ± 0.0	–
$25 \div 50\ GeV/c^2$	15	8.9 ± 1.2	6.1 ± 4.1	4.9%
$50 \div 75\ GeV/c^2$	90	92.5 ± 4.3	-2.5 ± 10.4	60.8%
$75 \div 100\ GeV/c^2$	62	38.6 ± 2.7	23.4 ± 8.3	0.088%
$100 \div 125\ GeV/c^2$	16	14.8 ± 1.7	1.2 ± 4.4	41.2%
$125 \div 150\ GeV/c^2$	1	3.2 ± 0.7	-2.1 ± 1.2	94.8%
$150 \div 175\ GeV/c^2$	3	0.9 ± 0.4	2.1 ± 1.8	7.7%
$175 \div 200\ GeV/c^2$	0	0.5 ± 0.3	-0.5 ± 0.3	100.0%

Table A.7: *Number of events, background predictions, excesses, and their Poisson probability (computed as the fraction of times the predicted number of events fluctuates to the number of observed events or more) bin by bin in invariant mass, for (++) events in the Signal Zone, predicted with the counting method described in the text.*

A.8 Mass Distributions

We have already shown in Sec. A.5 that the mass spectrum for double tags after tight cuts on the kinematical selection variables may give some evidence of the presence of a Z signal in the data. In this section we want to try and quantify the possible signal in these spectra by using shape information for the background (given by (+0) events selected the same way as the double tagged sample) and for the expected signal (given by Monte Carlo).

In the last section we have shown that the excess in the data attributable to the Z decay is of the order of 20 events after the selection cuts. In what follows we make an attempt to see if there can be agreement, albeit within the large error bars, between the observed excess obtained with the counting experiment and the signal one is able to fit in the mass distributions.

A.8.1 Fits to the Mass Spectra

We use two components to fit the mass spectrum of double SECVTX tags that pass our kinematic cuts, thus assuming once again that the data are a combination of QCD $b\bar{b}$ production and of Z decays to the same final state. For the signal template we use Monte Carlo events passing all our selection cuts ($N_{jet} = 2$, $\Delta\Phi > 3$, $\Sigma_3 E_T < 3\ GeV$, two SECVTX tags). For the background template we can instead use (+0) events passing the same cuts: the kinematic biases on the mass spectrum introduced by the selection are thus perfectly reproduced. We apply no smoothing to the templates—as is sometimes done in similar cases—since the results do not change appreciably: the templates are already reasonably smooth.

[17]The search bin contains $71 \pm 5\%$ of the Z signal according to the PYTHIA Monte Carlo.

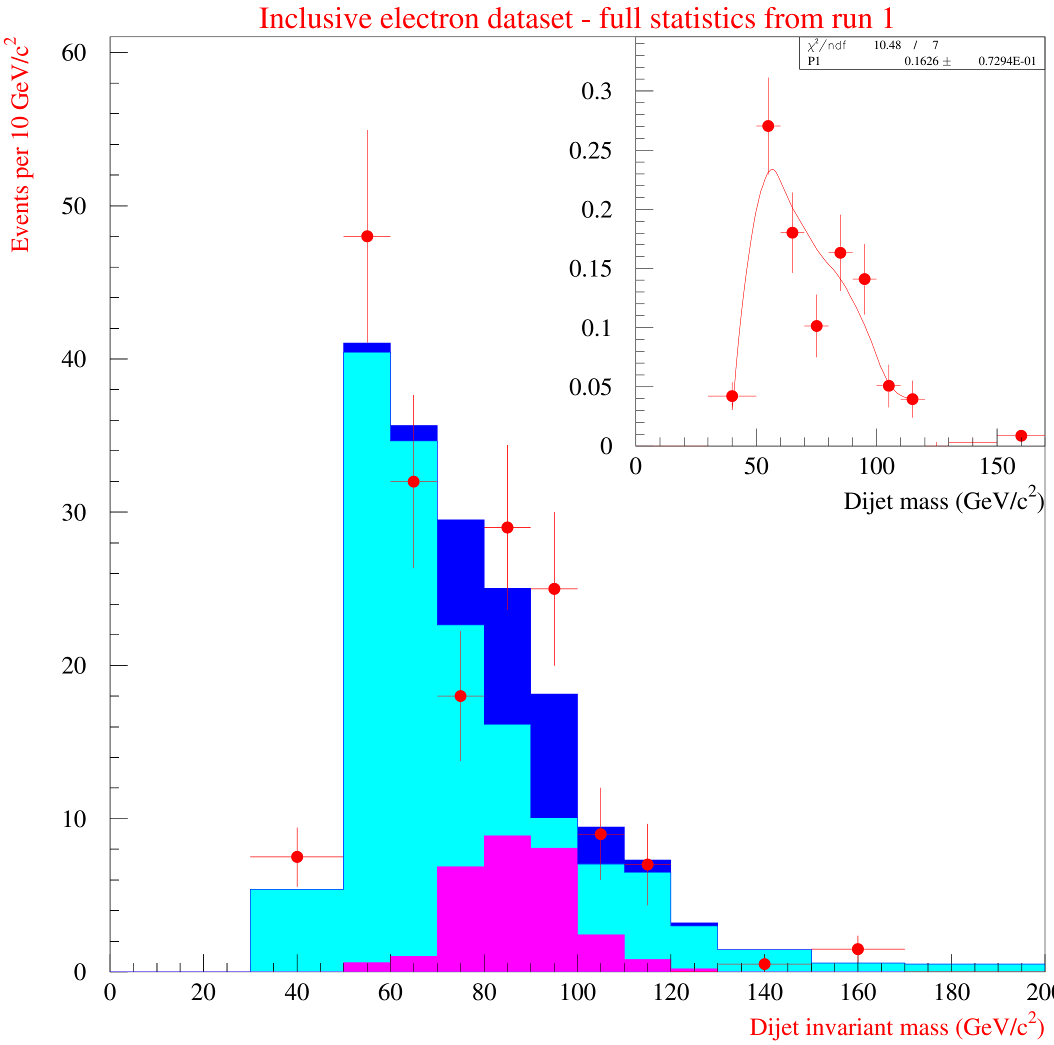

Figure A.10: *Two component fit to the double SECVTX tagged data, selected with* $N_{jet} = 2$, $\Sigma_3 E_T < 3$ *GeV*, $\Delta\Phi > 3$ *(points with error bars). The fit region is* $30 \div 120$ *GeV*/c^2. *The inset shows the fit, the histograms give the results for the template fractions. The background template (lightest histogram) is corrected for the fraction of Z signal accepted in the (+0) dataset, as described in the text.*

It must be pointed out that the choice of (+0) events is not the best one for a background sample: a small but probably not negligible signal contamination (estimable from Monte Carlo to be lower than 5% around the Z mass) must be present there. To sidestep the bias we can however correct the background template to account for the fraction of signal falling in the (+0) selection. To do that, we start with the uncorrected background template, get the fitted number of double tagged Z events and extrapolate it to the signal included in the (+0) selection, trusting the Monte Carlo for their ratio. We then correct the background template and perform the fit again, iterating the whole procedure until convergence is reached.

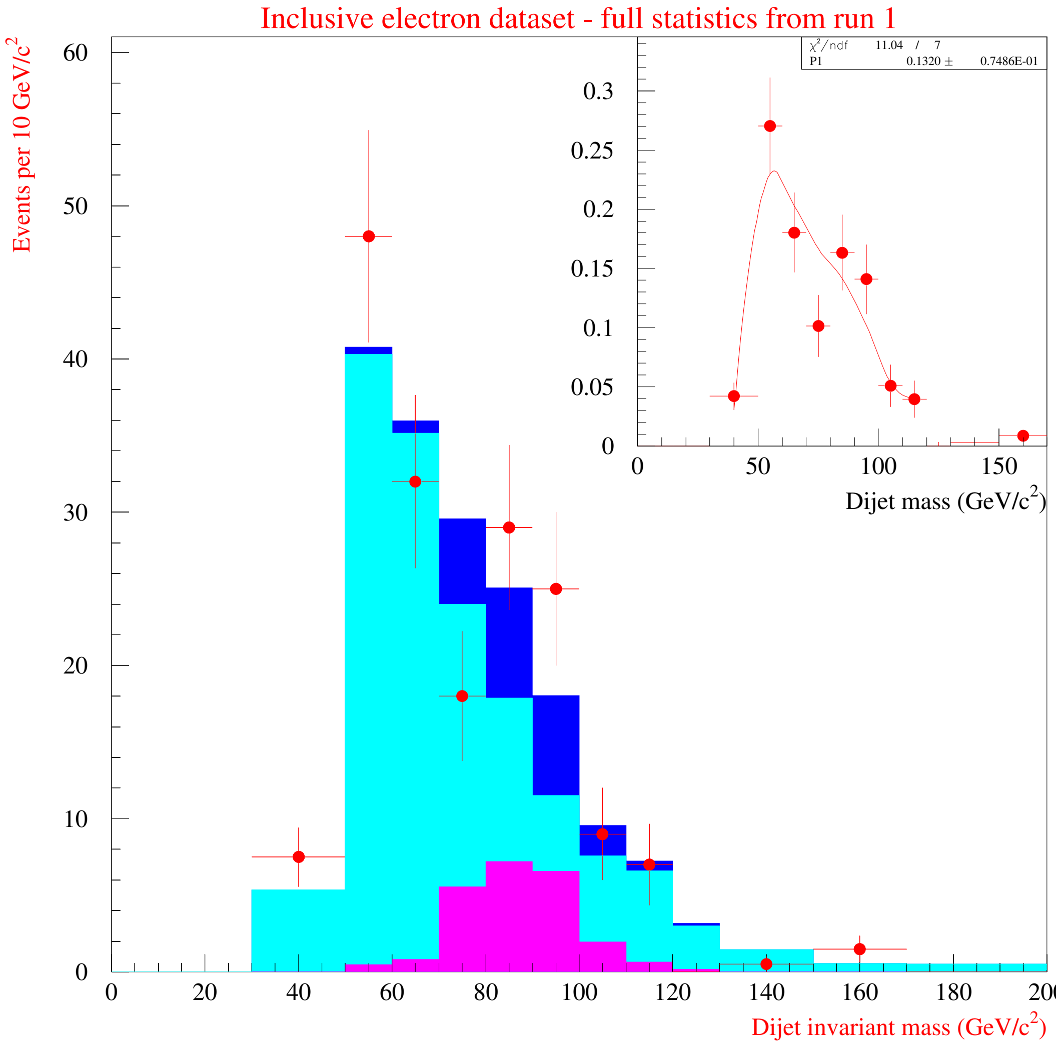

Figure A.11: *Two component fit to the double SECVTX data, selected with $N_{jet} = 2$, $\Sigma_3 E_T < 3$ GeV, $\Delta\Phi > 3$ (points with error bars). The fit region is $30 \div 120$ GeV/c^2. The inset shows the fit, the histograms give the results for the template fractions. The background template (lightest histogram) is not corrected for a signal contamination.*

The common problem of χ^2 fits to a low statistics spectrum as the one under study is the presence of bins with a number of entries too small for the Gaussian approximation to be valid: the fluctuations in these bins are higher than what is assumed by the χ^2 in the fit. We prevent small statistics bins from distorting our fits by using a variable bin size and by excluding from the fit the zone where the number of entries sinks under six per bin.

The results of the fits are shown in Fig. A.10 and Fig. A.11. The spectrum is well understood as the sum of background and signal, the χ^2 being 10.5 (11.0) for seven degrees of freedom with (without) the iterative correction described above; if one tries to fit the spectrum with the background template alone, the χ^2 obtained

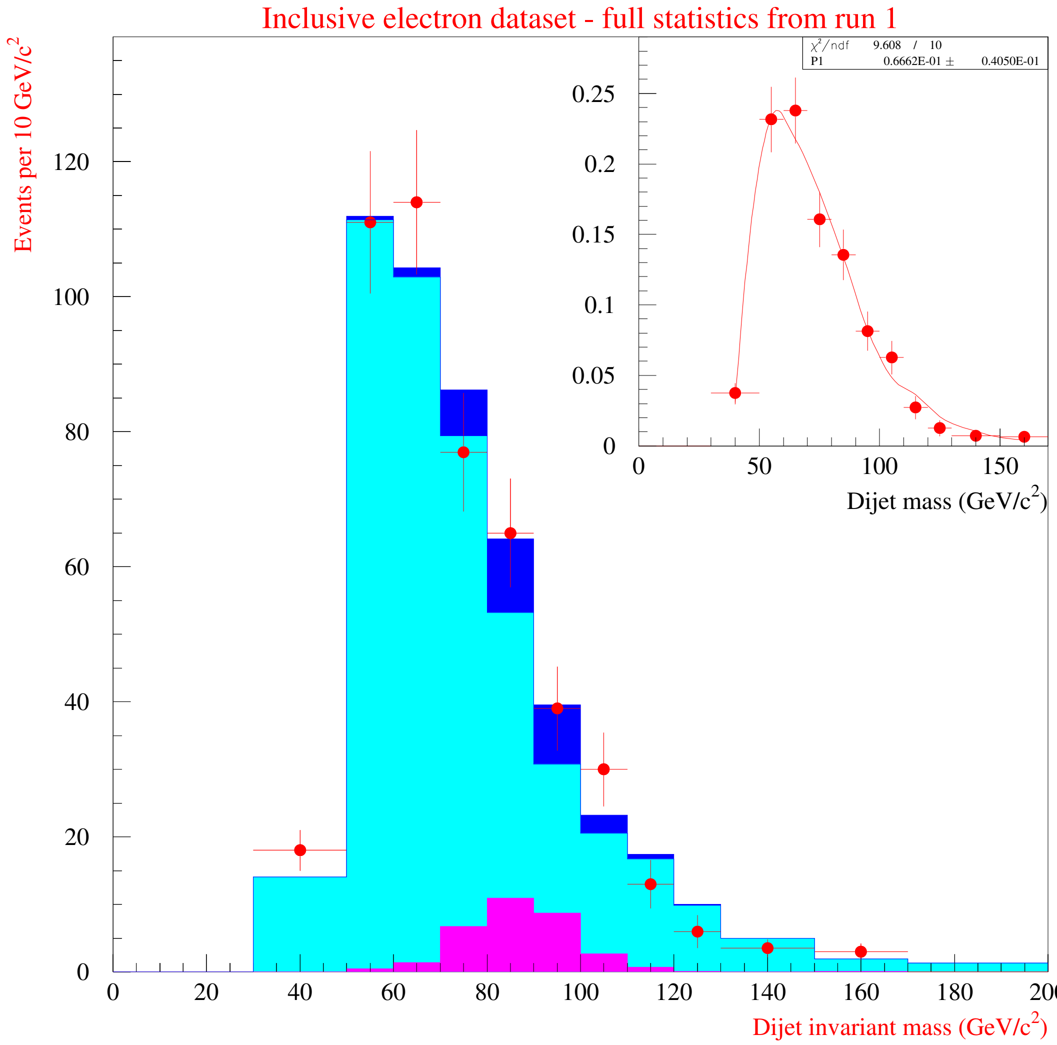

Figure A.12: *Two-component fit to the double SECVTX data, selected with* $N_{jet} = 2$, $\Sigma_3 E_T < 10$ *GeV*, $\Delta\Phi > 3$ *(points with error bars). The fit region is* $30 \div 170$ GeV/c^2*. The inset shows the fit, the histograms give the results for the template fractions. The background template (lightest histogram) is corrected for the fraction of* Z *signal accepted in the (+0) dataset, as described in the text.*

raises to 14.1/7.

The number of Z events expected in the whole spectrum is $N_{exp} = 16 \pm 3$ events, while the fit gives $N_{fit} = 28.9 \pm 13.1$ events (23.4 ± 13.4 without correction). Not unexpectedly, the number of fitted signal events is in good agreement with the results obtained by the counting experiment (23.4 ± 8.3 events), given the fraction of signal events (about 25%) expected to fall outside the search bin ($75 \div 100$ GeV/c^2) defined there. The discrepancy with the expected signal is also confirmed by these fits.

We can try and loosen up some of our selection cuts to see if we remain sensitive to the small amount of signal in the dataset. The sensitiveness is actually quickly

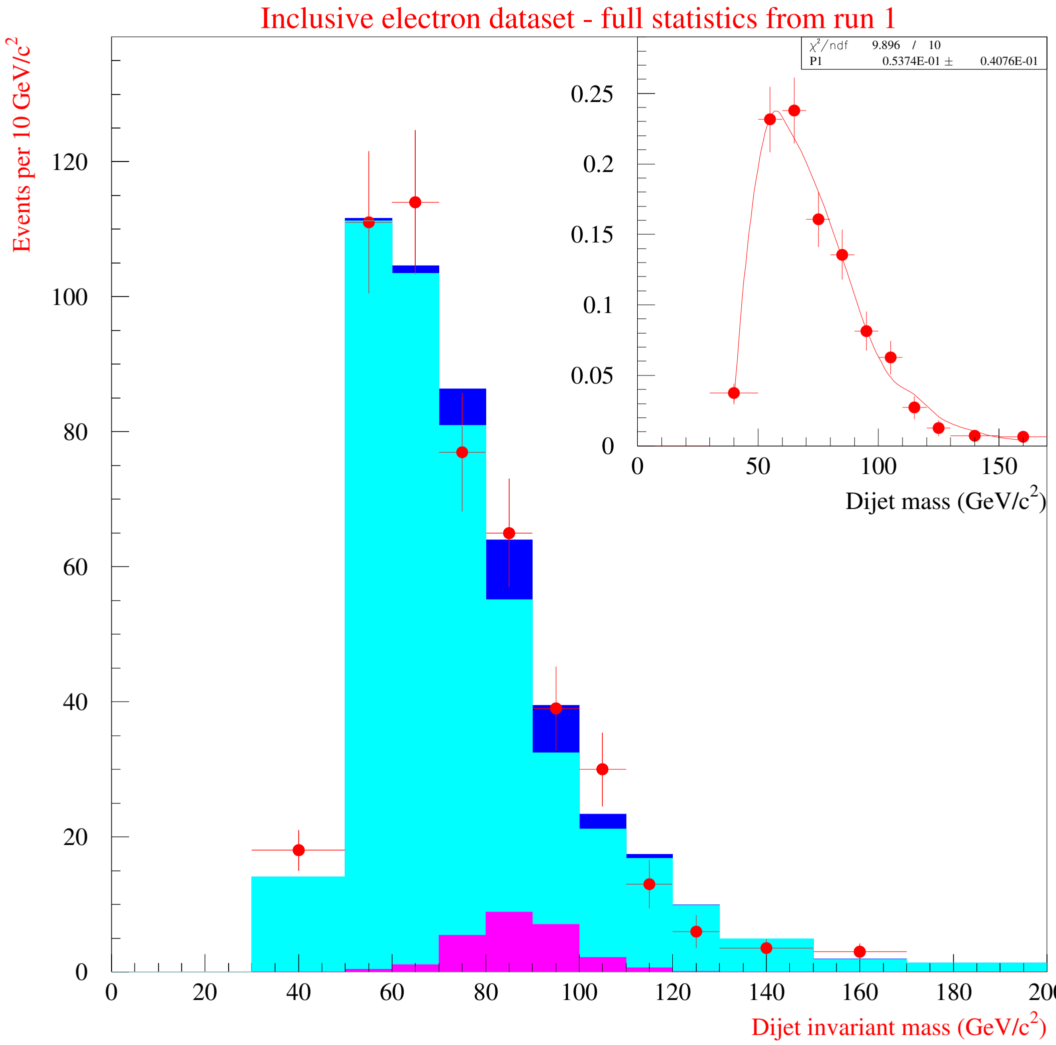

Figure A.13: *Two-component fit to the double SECVTX data, selected with $N_{jet} = 2$, $\Sigma_3 E_T < 10\ GeV$, $\Delta\Phi > 3$ (points with error bars). The fit region is $30 \div 170\ GeV/c^2$. The inset shows the fit, the histograms give the results for the template fractions. The background template (lightest histogram) is not corrected for a signal contamination.*

jeopardized as the S/N drops; anyways, we do get consistent results, albeit within very large error bars that do not allow to draw any further conclusion. See for instance the plots in Fig. A.12 and Fig. A.13, where the $\Sigma_3 E_T$ cut has been loosened to 10 GeV. The χ^2 is 9.6/10 (9.9/10) for the corrected (uncorrected) fit, and the signal obtained is $N_{fit} = 31.9 \pm 19.5$ ($N_{fit} = 25.8 \pm 19.6$), to be compared with the Monte Carlo prediction[18] $N_{exp} = 28 \pm 6$. While these numbers cannot really confirm the nature of the excess we see in the tighter selection, they appear to be more in line with the expected number of events.

[18]To estimate this number the efficiency derived in Sec. A.6 is used.

A.9 Conclusions

We have searched 8 million inclusive electron events for a $Z \to b\bar{b}$ signal. The depressingly low S/N ratio at trigger level (less than 10^{-3}) implies that a really tight selection is required to spot a signal. We have attempted to do so by using double SECVTX tagging plus some kinematic criteria suited to the discrimination of an electroweak production from the QCD background; that way we select 187 events, 62 of whose have a reconstructed dijet mass between 75 and 100 GeV/c^2, where the Z decay should yield about 11 events. We find some hint of the signal in our final dataset, quantifiable in an excess of 23.4 ± 8.3 events having a suggestive shape in the mass distribution.

Appendix B

In Favor of the Muon Dataset

B.1 Introduction

In App. A we reviewed the results of the search for $Z \to b\bar{b}$ decays in the inclusive electron dataset. The same selection tools described in the analysis of the muon dataset were used to increase the signal/noise ratio, but a stricter cut on the $\Sigma_3 E_T$ was necessary in order to spot a signal in the dijet mass distribution. Other things being equal, one would naively guess that the inclusive electron sample is more promising for a Z search than the inclusive muon sample, due to the higher acceptance our detector grants to tracked electrons than to CMUP muons[1]. On the other hand, Monte Carlo simulations predict that the muon dataset yields a signal two times as big—a fact confirmed by the size of the signal found in our analysis of the two datasets (compare the results of Ch. 6 and App. A). Some discussion of this apparent anomaly is therefore necessary.

B.2 Analysis of the Discrepancy

At the SECVTX level (where the events have already been required to contain a tight lepton and at least one SECVTX tagged jet), the two datasets are approximately equal in size: $134,548$ electron events and $105,782$ muon events. Yet the expected signal/noise ratio is already much larger for the muon dataset, where 724 ± 119 $Z \to b\bar{b}$ events are predicted, as compared with the 327 ± 58 of the electron dataset. This inconsistency is due to a couple of differences in the characteristics of electron and muon events. First of all, one must note that for a jet containing a $P_T > 8\ GeV/c$ muon the 10 GeV threshold of the SECVTX algorithm is much harder to pass than for a jet containing an electron of the same P_T, due of course to the low energy deposited in the calorimeters by the muon: this makes the SECVTX electron dataset much richer of low mass events[2], and the number of events at the SECVTX level

[1]See Sec. 2.4.4. The rapidity acceptance of the CMU-CMP chambers is $|\eta| < 0.6$, while electrons can be tracked up to $|\eta| \simeq 1.1$.

[2]Of course that difference is evident only after one corrects the muon jet momentum to account for the muon P_T, as we do according to the corrections described in Ch. 5.

in the two samples much less meaningful for a direct comparison. But by far the most important factor is the dramatic dependence of the efficiency of the electron identification cuts on the electron jet E_T. In fact, the tight cut on the hadronic over e.m. energy ratio of the electron cluster ($HAD\_EM_{3\times3} < 0.04$, a standard selection cut in b-quark analyses at CDF) makes it much harder for an electron to be identified in a high E_T jet than in a low E_T one, as can be seen in Fig. B.1.

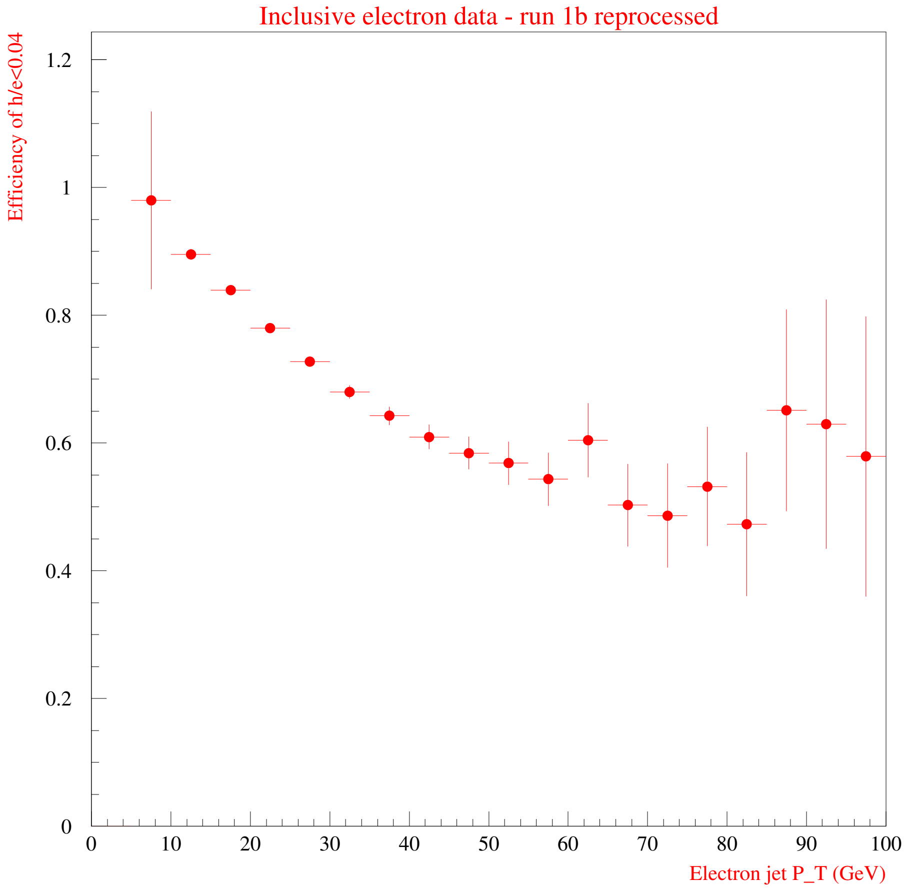

Figure B.1: *Survival probability for electrons passing the trigger requirements (that already include a cut $HAD\_EM_{3\times3} < 0.125$) to pass the $HAD\_EM_{3\times3} < 0.04$ requirement, as a function of the electron jet E_T .*

One can then take a pragmatic viewpoint, and obtain from the data themselves the ratio of muon over electron events *in the mass range of interest for our analysis.* The mass spectra for the two datasets at four different selection levels are shown in Fig. B.2. By considering only events with a dijet invariant mass in the range $80 \div 100\ GeV/c^2$, one deconvolutes the effect of the electron identification efficiency, and sidesteps the low-E_T inefficiency of the 10 GeV E_T threshold posed by SECVTX on muon jets. The results are shown in Table B.1.

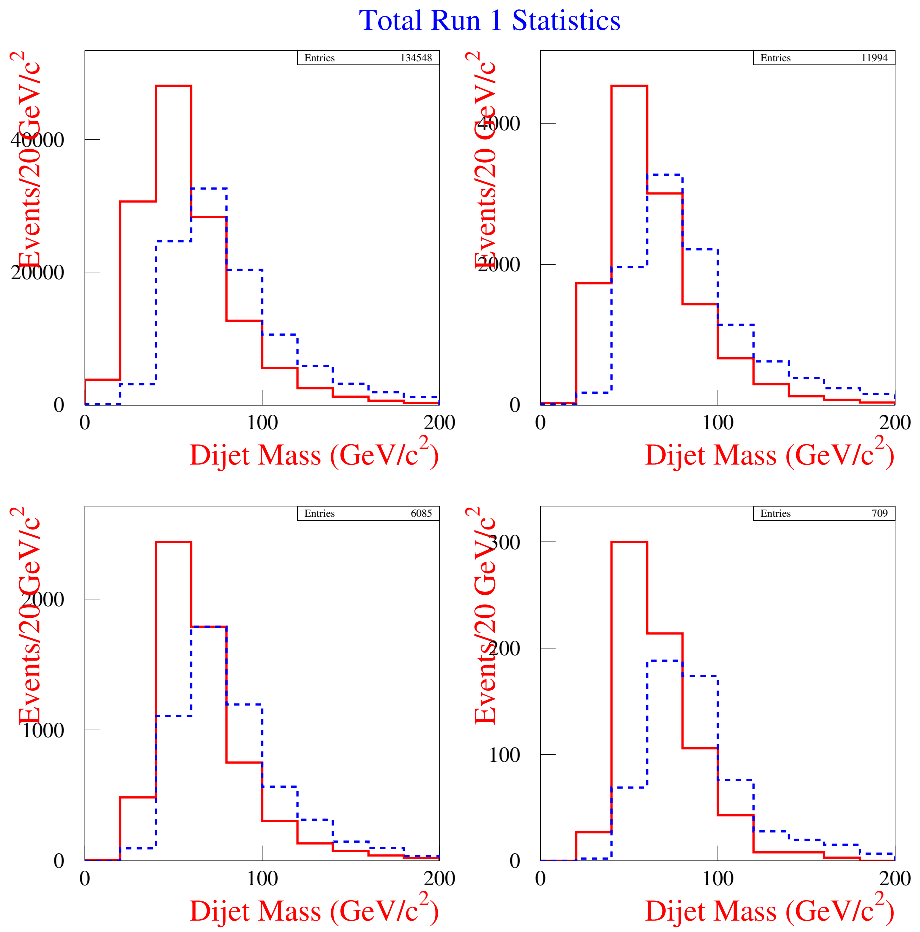

Figure B.2: *Mass spectra of electron (full line) and muon (hatched line) events at four stages of our selection. Top left: SECVTX level; top right: after the additional the kinematic requirements; bottom left: the double tag samples; bottom right: the final selection.*

The ratio (1.66 ± 0.20) is indeed in favor of muon events, and is in fair agreement with what the Monte Carlo predicts it to be (1.90 ± 0.08) for $Z \to b\bar{b}$ decays[3] Of course, the number thus derived can be affected by systematics due to non considered differences between the QCD $b\bar{b}$ production process (used for the estimate) and the real $Z \to b\bar{b}$ decay: but it is a good first-order guess.

In addition to what has just been estimated, there is another effect to factor in: the lepton originating from a $Z \to b\bar{b}$ decay is likely to be central, the more so

[3] We obtain this value from the ratio of events with two SECVTX tags in the muon and electron Monte Carlo datasets (we assume of course that the application of the kinematic cuts has the same effect on the two datasets). After normalizing to the run 1 luminosity and applying all corrections, the numbers are respectively 202 ± 40 and 103 ± 21, as we already know from Ch. 3 and App. A.

Selection	All ele	Ele 80 ÷ 100	All μ	μ 80 ÷ 100	Ratio
SECVTX	134,548	12,685	105,782	19,876	1.57 ± 0.02
+ Kin.	11.994	1,431	10,424	2,136	1.49 ± 0.05
2 SVX	6,085	753	5,457	1,222	1.62 ± 0.08
+Kin.	709	106	588	176	1.66 ± 0.20

Table B.1: *This table shows the number of events accepted by the selection cuts in the electron and muon datasets, and their ratio in the dijet mass window* $80 \div 100$ GeV/c^2 . *One clearly sees that the data themselves predict an acceptance 1.66 times higher for Z events with a tight muon w.r.t. Z events with a tight electron. Muons pass the requirements described in Ch. 3; electrons pass the requirements described in App. A.*

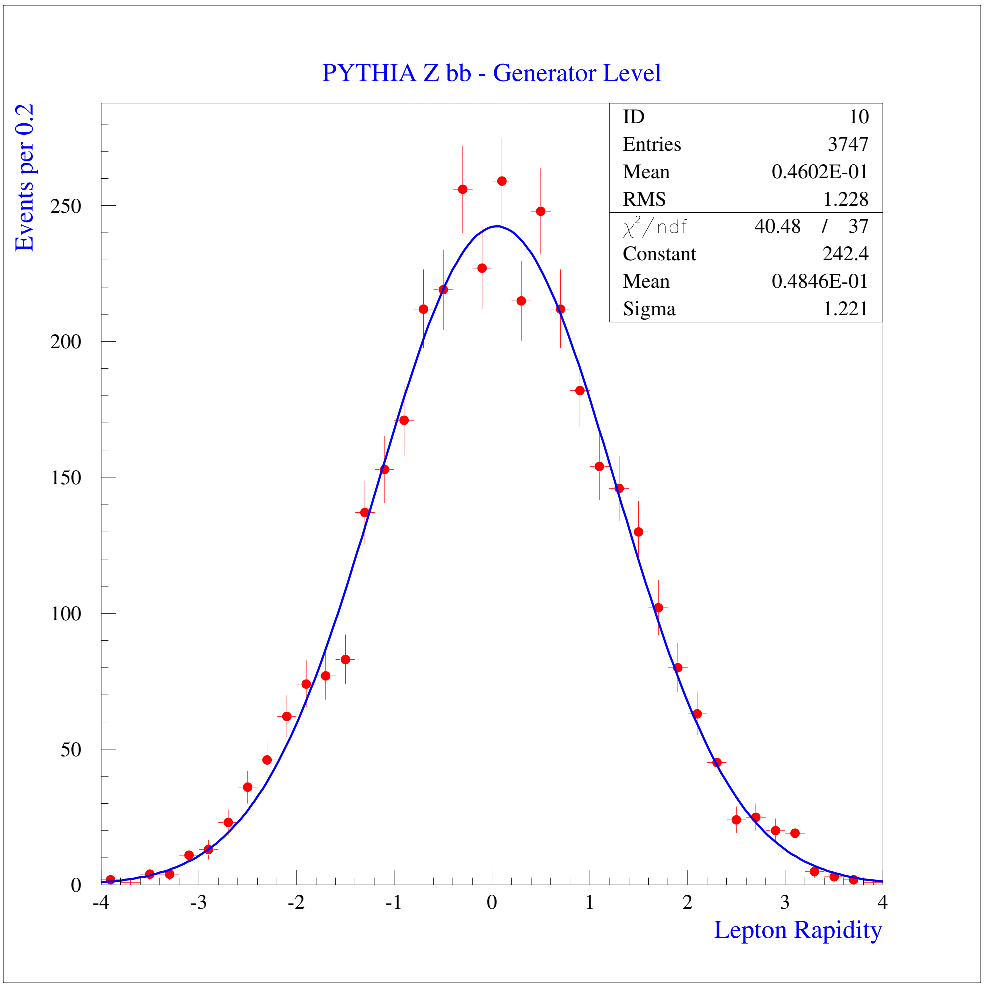

Figure B.3: *Rapidity distribution of leptons having* $P_T > 8$ GeV/c *from the decay of a b quark originated in a* $Z \to b\bar{b}$ *decay.*

if it has to pass a trigger requiring about 8 GeV/c of transverse momentum. The Monte Carlo indeed shows that the generated leptons with $P_T > 8$ GeV/c have a

Gaussian distribution in pseudorapidity with a sigma of about 1.2 (see Fig. B.3). CMUP muons, being more central than tight electrons ($|\eta| < 0.6$ versus $|\eta| < 1.1$), are therefore *favored* for the collection of Z events. In a Monte Carlo generation (PYTHIA, V5_7) of $60,000$ Z events with no requirements other than a cut at 8 GeV/c on the P_T of the produced lepton at generator level, we find that $1,424$ out of $3,747$ events fall in the CMUP acceptance, to be compared with the $2,352$ out of $3,747$ having pseudorapidity in the range $-1.1 \div 1.1$. We therefore expect the muon dataset to be favored by a factor $f = 1,424/2,352 \times 2.2/1.2 = 1.11 \pm 0.03$. The final estimate for the ratio of Z events in the two datasets must then be placed at $(1.66 \pm 0.20 \times 1.11 \pm 0.03 = 1.84 \pm 0.23$, in excellent agreement to the Monte Carlo prediction.

We conclude that there is no mistery concerning the larger signal we get in the muon dataset: our detector is able to identify very well isolated electrons coming from W or Z decay, but when they are produced in a high energy hadronic jet they are much harder to identify—something that does not apply to CMUP muons.

Appendix C

The Modeling of Multiple Interactions

C.1 Introduction

During run 1b the high luminosity reached by the Tevatron has pushed the average number of hard interactions per bunch crossing in CDF to values as high as four. The relative rarity of interactions capable of firing the trigger system allows us to consider only one interaction per collected event as a "hard scattering", while all the others, identified via the reconstruction of primary vertices using VTX informations, can be thought of as "satellite interactions". Most of the times, one does not care to study the effect of satellite interactions, since their impact is irrelevant on most of the features of the events one is interested in; one is then just concerned with keeping under control the effective belonging of particular physics objects (electrons, muons, missing E_T) to one and the same hard scattering, as far as our detector is able to discern. Sometimes, though, the "soft" physics of the so-called *minimum bias* interactions—accompanying the hard scattering that triggered the event—interferes with what one wishes to detect. That is the case, for instance, of searches for diffractive physics or double parton scattering, where a single additional hard interaction can hide the searched signal[1] or become an independent source of background. It is also the case of our search for Z bosons decaying to pairs of b quarks, where a relevant feature of the signal is its different radiation pattern from that of the background produced by QCD processes: to detect this difference one has to understand very well the energy flow in the calorimeters. One of the variables we use to discriminate Z decays to $b\bar{b}$ from QCD b-quark pair production is the sum of transverse energy in the calorimeter clusters, once the two leading jets are removed. This variable cannot be the ground for comparisons of Monte Carlo simulations with real data, with the idea of extracting the contribution of multiple interactions to the energy

[1]In the search for diffractive physics signatures, rapidity gaps in the energy flow measured by the calorimeter system are sought: this can be done only if just a single interaction occurred during the bunch crossing. As for the signal of double parton scattering, it has been extracted at CDF[65] by looking at uncorrelated dijet systems of very low energy, a feature shared with a double hard scattering from two separate $p\bar{p}$ pairs.

flow, because the Monte Carlo generators do not describe with sufficient accuracy another effect affecting the energy flow, namely the underlying event. One is thus forced to understand the contribution of multiple interactions from the real data. This is what we will try to do in the following.

In Sec. C.2 we introduce thw two datasets we wish to use for a comparison of the energy flux from satellite interactions. In Sec. C.3 we describe how we can select a subset of events where the triggering vertex—which we need to discard, to get rid of trigger biases—is easily identificable and removable. In Sec. C.4 we describe some additional tricks necessary for an unbiased comparison, and in Sec. C.5 we compare the energy flux for the two datasets. We conclude in Sec. C.6.

C.2 The Data Samples

We wish to model in our $Z \to b\bar{b}$ simulations the amount of energy flowing from the satellite vertices present in these datasets; we will use the inclusive electron dataset for the purpose of the present study.

The dataset we consider for this study is a selection of the large data sample coming from the COMBINED_ELEB_CEM trigger (run 1b). The initial sample amounts to about 5.6 million events containing an electron with E_T greater than 8 GeV, landed in the central calorimeter, and passing tight identification requirements (see App. A). We select events where one electron passes tight cuts (see Table A.3 on page 156 for the cuts used) and one jet is tagged by the SECVTX algorithm (see Table 3.3 on page 49 for the options used); these two requirements leave us with about 115,000 events. In what follows we will refer to this sample as the "SECVTX" sample.

We use minimum bias data collected during run 1b to try and model the pileup contributions in the SECVTX sample. The minimum bias trigger, requiring only a BBC coincidence, was heavily prescaled (see Sec. 2.5). It is therefore difficult to gather enough such events collected at luminosities greater than $15 \times 10^{30}\ cm^{-2}s^{-1}$.

We extract the distribution of instantaneous luminosity for the $115,000$ SECVTX events, binning it in 25 bins from 0 to $25 \times 10^{30}\ cm^{-2}s^{-1}$. This distribution has been used to select a sample of minimum bias events (in the following referred to as the MINBIAS sample) such that its cumulative luminosity distribution was exactly the same as that of the SECVTX dataset (see Fig. C.1). Because of the poor statistics available in the highest luminosity bins, we have decided to substitute the ten bins from 15×10^{30} to 25×10^{30} into a single bin at $16 \times 10^{30}\ cm^{-2}s^{-1}$. Due to the small fraction of events falling in the last bins, the approximation is surely a safe one for our purposes.

C.3 Associating Tracks and Jets to Vertices

We need to associate observable quantities in our detector to the interaction that originated them if we wish to compare the latter. In the following we describe an unequivocal definition of track P_T flow and energy flow for the primary vertices.

C.3.1 Matching Tracks to Vertices

To make comparisons in the activity of satellite vertices we must characterize it as well as we can. The information provided by the VTX (number of hits and segments in the drift chamber associated to each vertex) is really insufficient for that purpose, since these numbers are only indirectly linked to the real energy flux from the vertices. A better idea is to associate charged tracks seen in the CTC to the vertices, to enable a quantitative estimate of the P_T flow from each of them—albeit only in the central detector.

We take all the charged tracks reconstructed in the CTC, having P_T greater than 400 MeV/c, and all the primary vertices reconstructed in the event; we then associate each of the former to one of the latter by a "best match" algorithm: we compute the number of standard deviations from the hypothesis that a track comes from a given vertex, for each of the vertices in the event, using the z intercept error of the track and the z error of the reconstructed vertex. The track is then associated to the vertex with lowest significance, provided that their z distance is less than 5 cm. We thus obtain a list of tracks for each vertex, and can then characterize the latter by computing the number of tracks and the sum of the charged P_T in each.

C.3.2 Matching Jets to Vertices

The best instrument we can devise to understand the energy flow from vertices is the association of jets in the calorimeter with the vertex they originated from. However, the track momentum measurement in CDF extends to only ± 1.1 in pseudorapidity. This fact makes it impossible to correctly associate each cluster seen in the calorimeter to a primary vertex from which the energy has flown.

What one can still do, in the case of jets landed in the central calorimeter (as is most frequently the case with the first two jets in our SECVTX sample, where the b-tag requirement and the presence of a tight electron restrict the jet pseudorapidity distribution), is computing the sum of the P_T of those tracks landed in the calorimeter inside the $\eta\phi$ circles defining of the two leading jets, $\Sigma_{12}P_T$, for each of the vertices they belong to. That allows us to choose as triggering vertex the one that was really responsible for the appearance of the two jets, instead than just choosing the one with highest number of VTX segments—as is normally done—, or the one with highest ΣP_T flow *tout court*, or the one the electron track belongs to; that will also allow us to be free from possible biases due to assuming the most active vertex as the one that caused the event to be collected by the trigger system. Figure C.2 shows how the "safe" events can be chosen: if a vertex has a z value between ± 30 cm from the detector center, its tracks pointing to central jets will always be recorded by the CTC system, and a $\Sigma_{12}P_T$ value above 10 GeV/c assures us that we are considering the true vertex that originated the two leading jets of the event. The falloff of the z distribution at $\pm 30cm$ in the distribution of Fig. C.2 reflects the SVX acceptance. We will therefore restrict our analysis to events where at least one primary vertex fits to our definition of a true "triggering vertex".

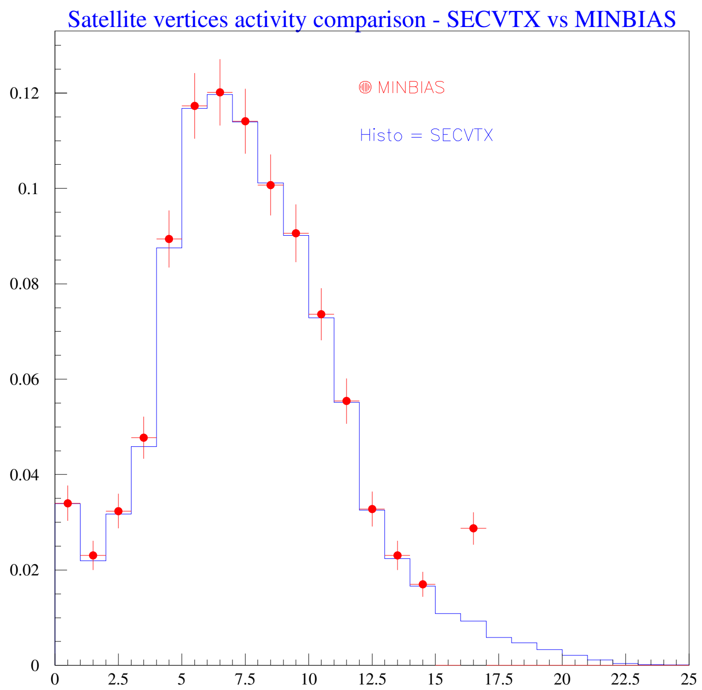

Figure C.1: *Instantaneous luminosity distribution for the SECVTX dataset (histogram) and for the MINBIAS dataset (dots).*

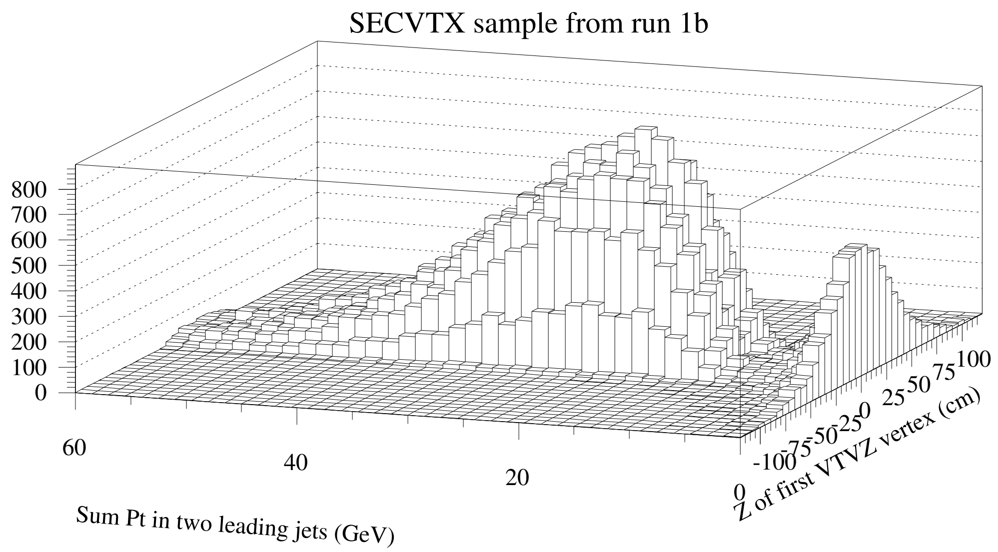

Figure C.2: *Sum of track P_T in the two leading jet cones versus the z of the leading VTX vertex. One can notice how the first VTX vertex is not always the one from which the two jets originate. One can also notice how the true vertex has always more than 10 GeV/c of charged P_T and is tightly restricted between ± 30 cm from the center of the detector.*

C.4 Unbiasing the Data

The data thus gathered can now be compared to the SECVTX dataset using variables that quantify the activity from the primary vertices in the event, once the triggering one has been discarded. One wishes to determine if that activity is comparable in the two datasets or if there remains some discrepancy that may prevent us from using the minimum bias as a modeling of what happens around the triggering vertex in the SECVTX dataset (or in other datasets). Variables useful to this purpose are the number of tracks from each vertex, the sum of their P_T, the number of VTX segments and hits associated to each vertex, and the number of vertices itself.

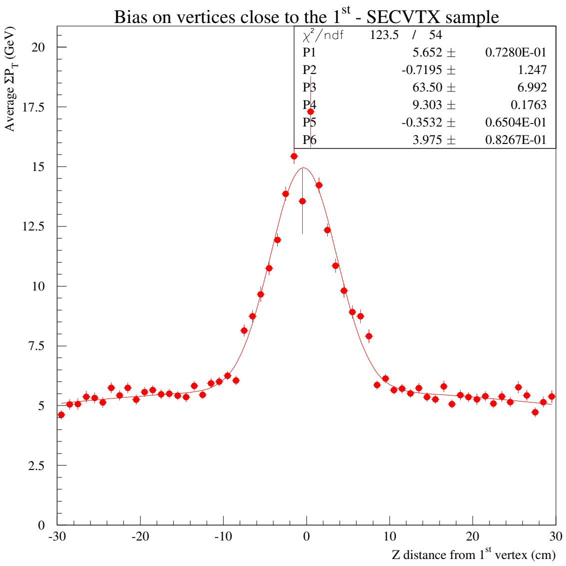

Figure C.3: *Average ΣP_T from the satellite vertices as a function of their z distance from the first vertex. One clearly notices how satellite vertices are likely to steal tracks from the triggering hard interaction.*

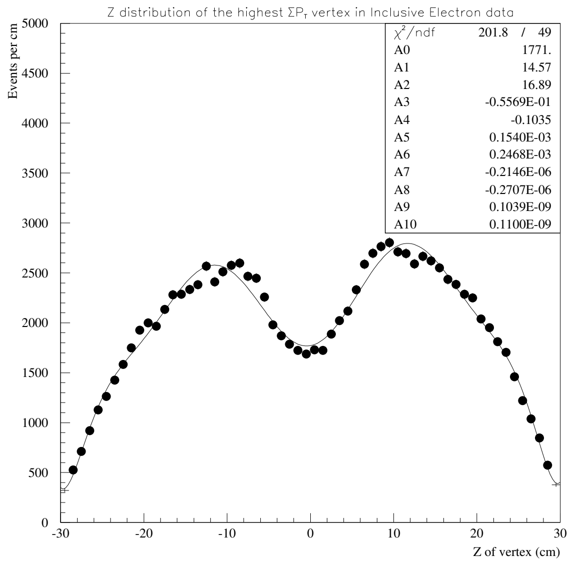

Figure C.4: *Z distribution of the primary vertex in SECVTX events.*
The fit parameters are used to model the ghost vertex in MINBIAS data (see text).

Of course, we must sort out the "triggering vertex" in the SECVTX sample: it had to contain a $E_T > 8\ GeV$ electron to begin with. But, even after that, the satellite vertices in the SECVTX sample cannot be compared to the vertices in the MINBIAS sample in a straightforward fashion. The presence of the triggering vertex in the SECVTX sample has to be taken into account, since even after its removal it is still able to create a bias on the other vertices; the reason for this is twofold:

1. some of the charged tracks originated from the hard scattering that contained the triggering electron may be mistakenly assigned to a different vertex, close in z (most of the times this is a fake vertex), thus increasing number of tracks and ΣP_T of these;

2. the removal of the triggering vertex in the SECVTX sample subtracts phase space to the satellite vertices in the z coordinate[2], thereby creating a disfavored condition compared to the MINBIAS sample.

These two effects have to be corrected for before any comparison can take place. To get a clear picture of the first problem, one may compute the average value of ΣP_T for the satellite vertices as a function of their z distance from the first vertex in the event. The result of such an exercise is shown in Fig. C.3: one very clearly notices how a satellite vertex reconstructed at less than 10 cm from the first vertex is likely to "eat up" tracks that do not belong to it. So we must discard satellite vertices that are closer than 10 cm to the first vertex.

The second effect can then be taken into account by playing on the MINBIAS data the very same trick we have just discussed: we must discard those vertices landed closer than 10 cm to the triggering vertex. But we do not exactly have a triggering vertex there: worse than that, we have it, but we do not have any idea of what that is, since the hits in the BBC counters (the only requirement of the minimum bias trigger) yield of course not enough information to tell apart two vertices. Still, what we can do is to extract the z distribution of the first vertex from the SECVTX sample (see Fig. C.4), and pretend that each MINBIAS event has a "triggering vertex" somewhere in the z axis, distributed exactly like the triggering vertices in the SECVTX sample. We will call that a "ghost vertex". Using a random number generator, we choose for each MINBIAS event a z value for a ghost triggering vertex according to the fitted polinomial distribution shown in Fig. C.4, and then remove any satellite vertices that lie closer than 10 cm from it.

C.5 The Comparison

After removing the triggering vertex and those satellite vertices too close to it, one must reorder in a proper way the vertices that are left. Since we are interested in their energetic activity, we order them by the ΣP_T they show. Once that is done, we can pick the first vertex (what was formerly the second, most of the times) in the SECVTX sample and the first vertex in the MINBIAS sample, and compare those variables that signal their activity; and so on for the other vertices.

A bit of care is still necessary, though. We must take into account the way we collected the MINBIAS sample with our triggers. The BBC coincidence forces these events to contain at least a couple of tracks, which is not really a big effect but may have an impact in the low end of the distributions we are going to compare. More to the point is the consideration that—to have a hit in the BBC counters—an interaction must fulfill spatial constraints that are different from other datasets (the SECVTX sample has been collected with a trigger that, in most part of its life, did not require any BBC coincidence). As a matter of fact, the z distribution of the satellite vertices in the MINBIAS sample is flatter than that of the SECVTX

[2]In fact, a second scattering close in z to the triggering one may fail to be reconstructed if its tracks can fit to the hypothesis of coming from the triggering vertex.

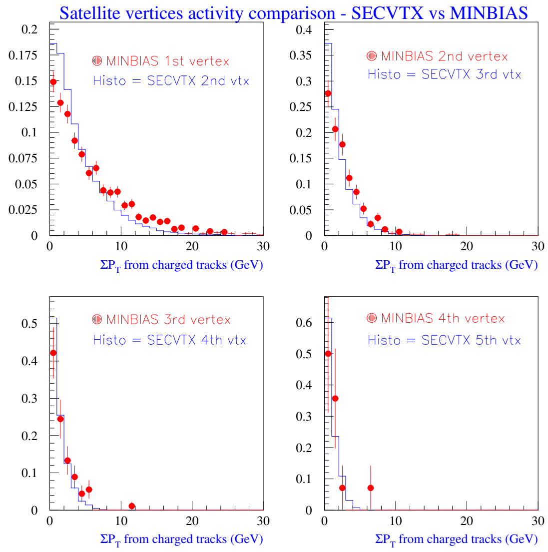

Figure C.5: *Sum of charged P_T from the satellite vertices in MINBIAS data (dots) and SECVTX data (histograms). The vertices are ordered by ΣP_T.*

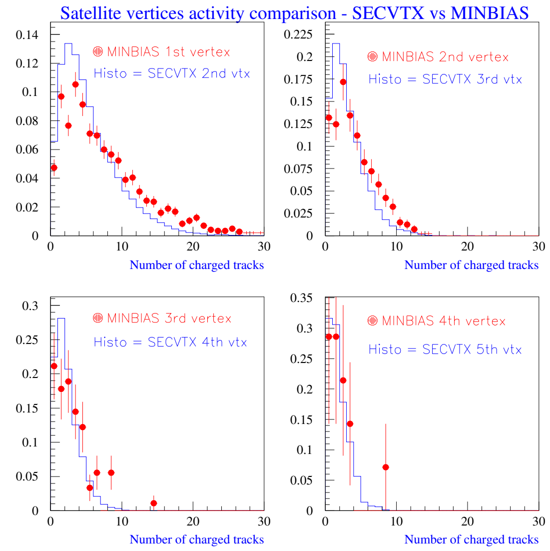

Figure C.6: *Number of charged tracks in the satellite vertices in MINBIAS and SECVTX data. The vertices are ordered by ΣP_T.*

sample; we interpret this fact as a result of the BBC requirement and sort it out by restricting our comparisons to vertices that have a z distance from the center of CDF amounting to less than 60 cm.

After these last fine tunings, we are able to overlay the normalized distributions for the two datasets. Figure C.5, Fig. C.6, Fig. C.7, and Fig. C.8 show the distributions of ΣP_T, number of charged tracks, number of segments and hits for the vertices of the two samples, in order by rank of the vertices. These comparisons show that indeed the activity in the SECVTX sample seems well modeled by a minimum bias sample. One also notices that, as expected, the MINBIAS data shows a deficit at small values of the variables. We interpret that with the residual effect of the BBC requirement; a second effect that may be partly responsible for the discrepancies in these distributions is the higher CTC occupancy in the SECVTX events, where dozens of charged tracks are reconstructed, although ignored in our analysis (they come most of the times from the triggering vertex): one expects a lower track-finding efficiency in these events, when compared to minimum bias events.

Another useful test of the equivalence of satellite interactions in the two datasets consists in counting the number of vertices that have ΣP_T over a certain threshold, using it to compare the two datasets. Figure C.9 shows that indeed the number of vertices, as well as their energy flow, is in good agreement in the two datasets.

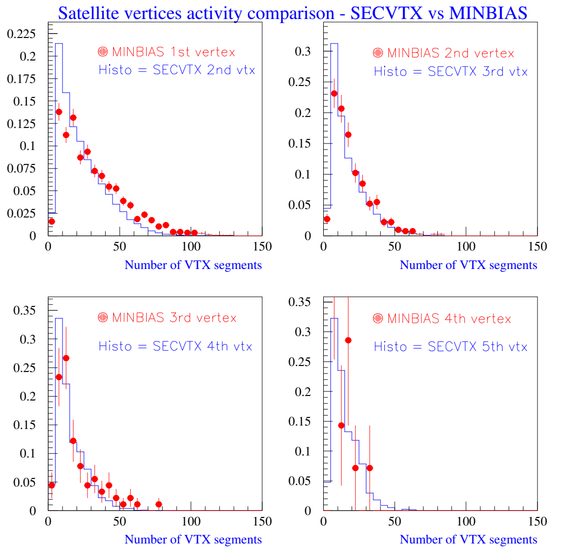

Figure C.7: *Number of VTX segments belonging to the satellite vertices in MINBIAS and SECVTX data. The vertices are ordered by* ΣP_T.

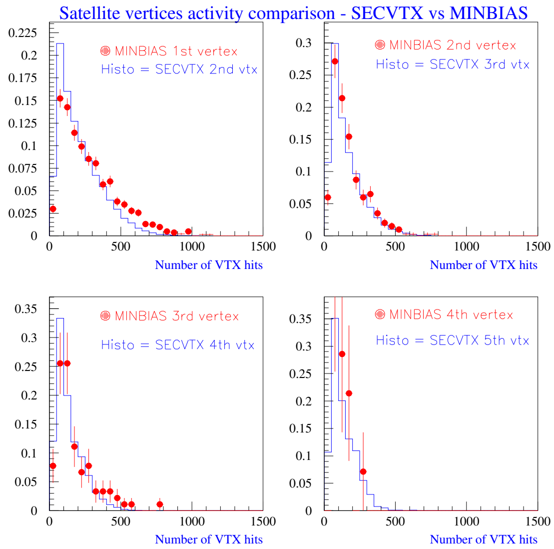

Figure C.8: *Number of VTX hits associated to the satellite vertices in MINBIAS and SECVTX data. The vertices are ordered by* ΣP_T.

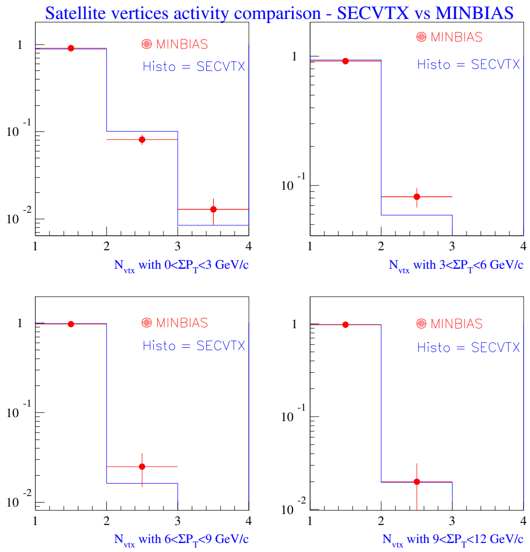

Figure C.9: *Number of primary vertices in different bins of* ΣP_T *for the two datasets. The triggering vertex in the SECVTX sample has been ignored.*

C.6 Conclusions

The checks we discussed above show that by simply accounting for the instantaneous luminosity of a given data sample one can use data collected with a minimum bias trigger to model the correct amount of satellite interactions in other triggers. We have indeed used this approach to include in our Monte Carlo simulations for Z production the effect of multiple interactions, that affects some of the variables we use to discriminate direct QCD $b\bar{b}$ production from the $Z \to b\bar{b}$ decay.

Bibliography

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

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

[3] A. Salam, in *Elementary Particle Theory: Relativistic Groups and Analicity (Nobel Symposium No. 8)*, ed. N. Svartholm (Almqvist and Wiksell, Sweden 1968) 367.

[4] P.W. Higgs, *Phys. Lett.* 12 (1964) 132; *Phys. Lett.* 13 (1964) 508.

[5] G.'t Hooft, *Nucl. Phys.* B33 (1971) 167.

[6] G. Arnison *et al.*, *Phys. Lett.* 122B (1983) 103.

[7] G. Arnison *et al.*, *Phys. Lett.* 126B (1983) 398.

[8] G. Banner *et al.*, *Phys. Lett.* 122B (1983) 476.

[9] P. Bagnaia *et al.*, *Phys. Lett.* 129B (1983) 130.

[10] M. Kobayashi and M. Maskawa, *Progr. Theor. Phys.* 49 (1973) 652.

[11] W. de Boer, *XVIII Physics in Collisions*, Frascati, 17-19 June 1998.

[12] R. Drucker, *"Electroweak Results from NuTeV"*, SLAC Summer Institute, Stanford (CA) 1998.

[13] J. Erker and P. Langacker, *"Bounds on the Standard Model Higgs Boson"*, HEP-PH/9801422.

[14] See http://www-cdf.fnal.gov/physics/ewk/ewk.html.

[15] J. Alitti *et al.*, *Phys. Lett.* 186B (1987) 452.

[16] J. Alitti *et al.*, *Z. Phys.* C49 (1991) 17.

[17] M. Mangano, private communication.

[18] F. Abe *et al.*, *Phys. Rev. Lett.* 80 (1998) 5720.

[19] *The Collider Detector at Fermilab*, a compilation of articles reprinted from Nuclear Instruments and Methods in Physics Research-A, North-Holland (1988).

[20] F. Abe *et al.*, *Phys. Rev.* D50 (1994) 5550.

[21] D. Amidei *et al.*, *Nucl. Instr. Meth.* A269 (1988) 51.

[22] F. Abe *et al.*, *Phys. Rev. Lett.* 77 (1996) 438.

[23] F. Abe *et al.*, *Phys. Rev.* D45 (1992) 1448.

[24] J.E. Huth *et al.*, Proc. of the 1990 Summer Study on High Energy Physics, *Research directions for the decade*, Snowmass, Colorado 1990, ed. E.L. Berger (World Scientific, Singapore 1992) 134.

[25] F. Abe *et al., Phys. Rev.* D47 (1993) 4857.

[26] D. Brown, *"A Study of Dijet Pt Matching with 1987 Data"*, CDF internal note CDF/ANAL/JET/CDFR/835 (1988).

[27] B. Flaugher, *"Calibration Stability and the Jet Energy Scale: Run 0 and Run 1a Comparisons"*, CDF internal note CDF/DOC/JET/CDFR/2902 (1994).

[28] T. Hessing, S. Behrends, P. Tipton, and B. Winer, *"Monte Carlo Fragmentation Tuning for Jet Energy and Resolution Analysis"*, CDF internal note CDF/ANAL/JET/GROUP/1131 (1990).

[29] M. Shapiro *et al.*, *"A User's Guide to QFL"*, CDF internal note CDF/ANAL/MONTECARLO/PUBLIC/1810 (1992).

[30] F. Abe *et al.*, *Phys. Rev. Lett.* 80 (1998) 2767.

[31] F. Abe *et al.*, *Phys. Rev. Lett.* 79 (1997) 3585.

[32] F. Abe *et al.*, *Phys. Rev. Lett.* 79 (1997) 1992.

[33] F. Abe *et al.*, *Phys. Rev.* D52 (1995) 4784.

[34] D. Amidei *et al.*, *"Study of the Jet Probability Btag Algorithm JPBTAG"*, CDF internal note CDF/PHYS/TOP/CDFR/2091 (1993).

[35] S. Dell'Agnello *et al.*, *"A Primary Vertex Finding Package"*, CDF internal note CDF/DOC/SECVTX/PUBLIC/1789 (1992).

[36] C. Caso *et al.*, "Review of Particle Physics", *Eur. Phys. Journ.* C3 (1998) 1.

[37] C. Miao and W. Yao, *"SECVTX Tagging Study with Track-Finding Degradation"*, CDF internal note CDF/ANAL/TRACKING/CDFR/3542 (1996).

[38] G. Unal, *"Tracking Efficiency and SVX B-tag Efficiency"*, CDF internal note CDF/ANAL/TOP/CDFR/3563 (1996).

[39] A. Warburton, *"Run 1A and 1B Low-p_T Single- and Double-Track CTC Pattern Recognition Efficiencies"*, CDF internal note CDF/ANAL/BOTTOM/CDFR/4139 (1997).

[40] F. Abe *et al.*, *Phys. Rev. Lett.* 80 (1998) 2779.

[41] P. Giromini *et al.*, *"Measurement of the Data to Simulation Scale Factor for the Tagging Efficiency of SECVTX and JPBTAG"*, CDF internal note CDF/ANAL/TOP/CDFR/4257 (1997).

[42] P. Giromini *et al.*, *"Rates of Jets with Heavy Flavor in the Data and in Herwig Simulations"*, CDF internal note CDF/ANAL/TOP/CDFR/4293 (1997).

[43] H. Bengtsson and T. Sjostrand, *Comput. Phys. Commun.* 46 (1987) 43.

[44] G. Marchesini *et al.*, *Comput. Phys. Commun.* 67 (1992) 465.

[45] R. Wilkinson, W. Wester, and R. Hollebeek, *"Search for the Hadronic Decays $W/Z \to jj$ Using Inclusive Dijets in Run 1c"*, CDF/ANAL/JET/CDFR/4191 (1997).

[46] D. Costanzo and S. Lami, *"Comments to the Search for the Hadronic Decays $W/Z \to$ jet jet in the run 1c data"*, CDF internal note CDF/ANAL/JET/CDFR/4471 (1998).

[47] F. Abe *et al.*, *Phys. Rev.* D50 (1994) 2966; F. Abe *et al.*, *Phys. Rev. Lett.* 73 (1994) 225.

[48] D. Hennessy, *"Getting Your Luminosity"*, CDF internal note CDF/DOC/CDF/CDFR/4317 (1997).

[49] C. Grosso-Pilcher, *"Luminosity for Run 1b: How Well Do We Know It ?"*, CDF internal note CDF/PHYS/CDF/CDFR/4350 (1997).

[50] CERN Program Library C301 (1992).

[51] F. Abe *et al.*, *Phys. Rev.* D49 (1994) 1.

[52] F. Abe *et al.*, *Phys. Rev. Lett.* 76 (1996) 3070; F. Abe *et al.*, *Phys. Rev.* D52 (1995) 2624.

[53] S. Abachi *et al.*, *Phys. Rev. Lett.* 75 (1995) 1456.

[54] W. Sakumoto, private communication.

[55] P. Avery, K. Read, and G. Trahern, Cornell Internal Note CSN-212 (1985).

[56] J.D. Lewis, *"Measurement of Run 1a Level 2 CMUP Trigger Efficiencies"*, CDF internal note CDF/ANAL/TRIGGER/CDFR/3130 (1995).

[57] M. Chertok *et al.*, *"Run 1b Level 2 Muon Trigger Efficiency Study"*, CDF internal note CDF/ANAL/TRIGGER/CDFR/4017 (1997).

[58] J. Olsen, *"Integrated Luminosity for the Dynamically Prescaled CMUP and FMU Triggers in Run 1b"*, CDF internal note CDF/DOC/BOTTOM/CDFR/4409 (1997).

[59] M. Paulini *et al.*, *"About the Sample Composition of SECVTX Tagged Single Lepton Data"*, CDF internal note CDF/ANAL/BOTTOM/CDFR/3757 (1996).

[60] M. Mangano, private communication; M. Seymour, private communication; D. Soper, private communication.

[61] N. Varclas, *"A Summary of Recent Color Coherence Results"*, Proc. XVIII Physics in Collision, Frascati, June 1998.

[62] F. Abe *et al.*, *Phys. Rev.* D50 (1994) 5562.

[63] B. Abbott *et al.*, Phys. Lett B414 (1997) 419.

[64] D. Hennessy, *"Luminosity at CDF"*, CDF internal note CDF/DOC/CDF/CDFR/4721 (1998).

[65] F. Abe *et al.*, *Phys. Rev.* D47 (1993) 4857; *Phys. Rev.* D56 (1997) 3811; *Phys. Rev. Lett.* 79 (1997) 584.

[66] R. Blair *et al.*, *"The CDF II Detector: Technical Design Report"*, FERMILAB-PUB-96/390-E (1996).

[67] A. Aldarese *et al.*, *"SVT: An Online Silicon Vertex Tracker for the CDF Upgrade"*, Proc. 7^{th} Pisa Meeting on Advanced Detectors, La Biodola, May 1997.

[68] R. Culbertson and M. Shochet, *"$Z \to b\bar{b}$ in Run 2"*, CDF internal note CDF/ANAL/TOP/CDFR/4158 (1997).

[69] C. Bigongiari and S. Leone, *"Proposal for a $Z \to b\bar{b}$ Trigger Using SVT"*, CDF internal note CDF/DOC/TRIGGER/CDFR/4824 (1998).

[70] T. Dorigo, *"A Trigger Breakdown of the Inclusive Central Electron Datasets"*, CDF internal note CDF/PHYS/BOTTOM/CDFR/4079 (1997).

[71] K. Byrum and A.B. Wicklund, *"The Level 2 Trigger Efficiency for 9 GeV Electrons in Run 1a"*, CDF internal note CDF/MEMO/ELECTRON/GROUP/3120 (1995).

[72] M. Paulini *et al.*, *"Time Dependent B0 Mixing in Inclusive Lepton Data Using Jet Charge and Soft Lepton Flavor Tags"*, CDF internal note CDF/ANAL/BOTTOM/CDFR/3810 (1996).

List of Figures

List of Tables