Model Independent Search for Deviations from the Standard Model at the Tevatron: Final States with Missing Energy

von

Philipp A. Biallass

Diplomarbeit in PHYSIK

vorgelegt der

Fakultät für Mathematik, Informatik und Naturwissenschaften
der Rheinisch-Westfälischen Technischen Hochschule Aachen

im Dezember 2004

angefertigt im

III. Physikalischen Institut A
Prof. Dr. Thomas Hebbeker

Contents

Introduction

Science attempts to describe and understand the complexity of nature. In the wide field of science, *Particle Physics* deals with the most fundamental processes in nature by investigating the forces and interactions between elementary particles, the components of all matter.
The interaction between single elementary particles dominates at high energies, for example conditions comparable to the first moments after the Big Bang. This is why Particle Physics is often also called High Energy Physics. The idea is to develop a theoretical model of the world at the quantum scale and use this knowledge to explain phenomena at the largest scales, too. In the past years it has become clear that only by investigating the fundamental interactions between subatomic particles the past, present and future of the universe can be understood. Nature's laws seem to inevitably connect particle physics to cosmology, which deals with questions like the formation of galaxies. This connection stresses the importance and fascination of research in this field of physics.

The Standard Model

The theoretical model to describe the basic interactions (electromagnetic, weak and strong force) is the *Standard Model* of fundamental particles and forces (SM). The mathematical foundation of the Standard Model is called Quantum Field Theory. Major contributions to this theoretical framework were made by S.L. Glashow, A. Salam and S. Weinberg. The Standard Model predicts the strength of a particular force, i.e. couplings, and the course of an interaction and the produced particles, i.e. cross sections. It is the task of Experimental Particle Physics to measure the predictions made by theorists in order to test the Standard Model. Tremendous efforts have been made in the past to test this model by experiments at large international facilities like *CERN* (Centre Européenne pour la Recherche Nucléaire). So far, no significant deviations have been found, and the success story of the Standard Model continues.
The surprising fact is that the Standard Model predicts its own downfall as it does not include the last fundamental force, gravity, and as it is restricted to energies below a certain level ($E < m_{\text{Planck}} = 10^{19}$ GeV). In addition to this, certain phenomena in the universe (like dark matter) cannot be explained by the Standard Model alone, and one SM particle, the Higgs, has not been found yet.
It has always been the dream of physics to find a "Theory of Everything" which addresses all these problems and mysteries. At the moment we are on the verge of a new era in physics as new facilities like the Large Hadron Collider (LHC) promise to breach the

current boundary to a new theory, and physics beyond the Standard Model seems to be detectable. This physics beyond the Standard Model is what is often called *New Physics*.

Analysis Summary

At the moment the *TEVATRON* collider at *Fermilab* (FNAL, Fermi National Accelerator Laboratory, Chicago/USA) is taking data at the highest energies currently available in the world. Until construction of the LHC is completed, it is the most likely place to enter the regime of New Physics.
The analysis presented in this diploma thesis attempts an alternative approach to a data analysis and wants to successfully perform a *Model Independent Search for New Physics* for the first time at DØ-Run II, one of the experiments at Fermilab. A complementary analysis, focusing on different final states, can be found in the diploma thesis by Oliver Kraff [1]. Model-independent in this context means that the physical data are compared to simulations (Monte Carlos) assuming the Standard Model. These comparisons are conducted using all particles detected and by investigating simple distributions like particle momentum. The aim is to scan a large part of the physical phase space and to find deviations that indicate New Physics without specifically testing a certain theoretical extension of the SM. As this is an alternative concept which has only been conducted a few times before, Model Independent Search provides great potential of discoveries as well as unexpected complications.

In this analysis a data set corresponding to an integrated luminosity $\mathcal{L} \approx 220\,\mathrm{pb}^{-1}$ of $p\bar{p}$-collisions is considered. These data were recorded by the DØ experiment between July 2002 and April 2004 with a center of mass energy of 1.96 TeV. The transverse momentum distributions of certain combinations of electrons, photons, muons, jets and Missing Transversal Energy (MET) are reviewed and compared to Monte Carlo simulations. A specific algorithm is presented here to systematically identify the region of greatest deviation in each distribution using Poisson statistics. Then another algorithm tests the statistical significance of this deviation.
As a result of this analysis, the Search Algorithm shows a satisfactory performance and the general agreement between data and Standard Model prediction is well. Nevertheless, some analysis channels show significant deviations. In the context of this Model Independent Search, evidence for the existence of New Physics cannot be claimed, as the discrepancies could as well be caused by detector effects not simulated properly in the Monte Carlo simulations. Further studies should examine these deviations more closely.

Chapter 1

Theoretical Foundations

1.1 Theoretical Framework of Particles and Interactions

1.1.1 Notations and Conventions

The paper presented here uses "God-given" units, where

$$\hbar = c = 1 \; . \tag{1.1}$$

In this system (natural units),

$$[\text{length}] = [\text{time}] = [\text{energy}]^{-1} = [\text{mass}]^{-1} \; . \tag{1.2}$$

The mass (m) of a particle is therefore equal to its rest energy (mc^2), and also to its inverse Compton wavelength ($mc/\hbar$). For example,

$$m_{\text{electron}} = 9.109 \times 10^{-28} \text{ g} = 0.511 \text{ MeV} = (3.862 \times 10^{-11} \text{ cm})^{-1} \; . \tag{1.3}$$

This text uses the common conventions for relativity, where the metric tensor $g_{\mu\nu} = g^{\mu\nu}$ runs with Greek indices over 0, 1, 2, 3 or t, x, y, z. γ^μ are the gamma matrices with $\mu = 0, ..., 3$. Repeated indices are summed in all cases. As a consequence, for a particle with mass m holds

$$p^2 = p^\mu p_\mu = E^2 - |\mathbf{p}|^2 = m^2 \; . \tag{1.4}$$

Throughout the text, particles and anti-particles are abbreviated by the name of the corresponding particle, and indices specifying the particle's charge are often omitted. Thus, the reaction $Z^0 \to \mu^+\mu^-$ is described by writing $Z \to \mu\mu$.

1.1.2 The Standard Model

This analysis seeks deviations from the Standard Model (SM), a theoretical framework describing nature which has been tested with enourmous accuracy and success in the past. One could question the value of an analysis which seeks to find alternatives to a model so demonstrably trustworthy. To understand the reasons for believing that a theory beyond the Standard Model exists, the principles and main features of the SM should first be introduced. As the mathematical formalism of the SM,*Quantum Field Theory* (QFT), is extremely complex, this Section is meant only to address a few of the basics and to outline the formalism. Details can be found in the literature, e.g. [2].

Quantum Field Theory is a relativistic field theory which can be regarded as an expansion of Quantum Mechanics. In contrast to every day life, matter is no longer characterized by particles, but for every particle a corresponding field is postulated. A field is a physical object which can emerge everywhere in space-time, e.g. a scalar field $\Phi(\vec{x}, t)$. With the help of the field concept, commonly known from "classical" Electro Dynamics, many problems of relativistic Quantum Mechanics can be solved.
After some simple steps, QFT uncovers numerous milestones: (1) Requiring a *continuous symmetry* (e.g. a global $U(1)$) of the quantized field inevitably leads to the existence of *anti-particles* with opposite quantum number (e.g. electrical charge); (2) the spectrum of states has positive energy $E \geq 0$; (3) causality is preserved. This causality is very important as it leads to the conclusion that QFT is "local", i.e. a measurement of the field at the space-time-point x does not affect a measurement at y.
Most concepts from Quantum Mechanics, like the construction of a Lagrangian $\mathcal{L}$, a Hamiltonian $\mathcal{H}$, and the principle of minimal virtue $\delta S = 0$, are retained. From this well known formalism, equations of motion for non-interacting particles can be derived, e.g. the Dirac-equation describing the spinor field Ψ (spin-$\frac{1}{2}$ fermion):

$$(i\partial_\mu \gamma^\mu - m)\Psi = 0 \; . \tag{1.5}$$

Until now only Lorentz-symmetry has been assumed, expressed by the invariance under Poincaré transformations. This philosophy of symmetries determines the structure and beauty of the Standard Model: One postulates fields (e.g. spinor Ψ or vector field A_μ) and symmetries, and one constructs the most global Lagrangian which is compatible with these symmetries. In addition to the Poincaré symmetry, *local symmetries* (e.g. $SU(2)$) are demanded, defined by:

$$\Psi_b \longrightarrow [e^{i\varepsilon^A T^A}]_{ab} \Psi_b \quad \text{with} \quad A = 1, .., D \; . \tag{1.6}$$

Here ε^A are real parameters and T^A are the generators of the transformation. D is the dimension of the corresponding symmetry group (e.g. $D = 1$ for $U(1)$) and a, b are the *inner* quantum numbers. These local symmetries are called inner symmetries, and their quantum numbers are discrete. The inner indices are traits of a particle which add to the spin and the mass. An example is the $SU(3)$ symmetry which defines the strong force and the Quantum Chromo Dynamics (QCD). Its eight generators correspond to the eight gluon fields G^A_μ, and the inner degrees of freedom are the three colors r, g, b.

At the point where local symmetries are assumed, the theory of free fields migrates to a theory of interacting fields. Expressions in the Lagrangian like $\overline{\Psi} i\partial_\mu \gamma^\mu \Psi$, where $\overline{\Psi} = \Psi^+ \gamma^0$, is not invariant under the local symmetry transformation. *Gauge fields* A_μ^A must be introduced, fields with a defined transformation behaviour which ensures the invariance of the Lagrangian. In order to ensure invariance of all terms of $\mathcal{L}$, the covariant derivative is defined:

$$\partial \longrightarrow \partial - igA_\mu^A T^A \; . \tag{1.7}$$

As a consequence, terms in $\mathcal{L}$ with combinations of matter fields like Ψ and gauge fields like A_μ^A are generated, representing the interaction between the fields with the coupling constant g. These vector fields correspond to spin 1 particles which mediate the forces. In the Standard Model, three fundamental forces are realized and can be described using the formalism of symmetry and gauge field:

- $SU(3)$ strong force $\longrightarrow$ eight gluon fields G_μ^A
- $SU(2) \times U(1)_Y$ electroweak force $\longrightarrow$ three W-boson fields W_μ^A and one hyper charge boson field B_μ

The vector bosons which can be observed in nature (W, Z, γ) are linear combinations of these electroweak fields, e.g. for the photon field A_μ which mediates the electromagnetic force with the couplings g and g':

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}} \left(g' W_\mu^3 + g B_\mu \right) \; . \tag{1.8}$$

On the other hand, the matter fields represent the fermions which interact with one another:

- Quarks: strong, weak and electromagnetic interaction
- Leptons: weak and electromagnetic interaction
- Neutrinos: weak interaction

	I	II	II	Charge
Leptons	e	μ	τ	-1
	ν_e	ν_μ	ν_τ	0
Quarks	u	c	t	$+2/3$
	d	s	b	-1/3

Table 1.1: *The three generations of matter*

Each of these matter fields is devided into three families, e.g. electron/muon/tau. In the SM there are only three families, but no theoretical explanation of this number is given. Only the weak force can mediate transformations from one family to another, as well as transitions inside a family (e.g. $e^- \leftrightarrow \nu_e$). All fermions of the Standard Model are summarized in Table 1.1. The final Lagrangian of the SM can be outlined by:

$$\mathcal{L}_{SM} = \mathcal{L}_{Kin} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \; . \tag{1.9}$$

The kinetic term $\mathcal{L}_{Kin}$ includes all interactions between the different fields: the free fields, interactions between the fermion fields and gauge boson fields, and combinations of gauge

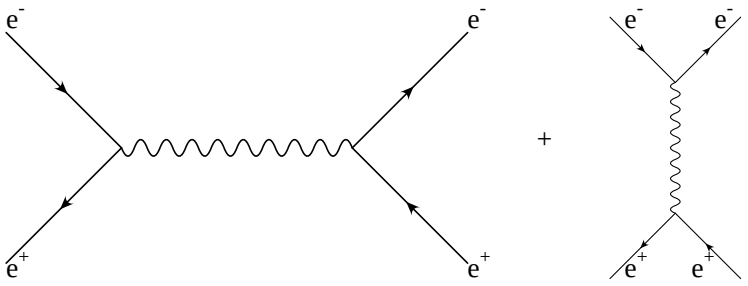

Figure 1.1: *The two Feynman diagrams contributing to Bhabba scattering* $e^+e^- \rightarrow e^+e^-$

boson fields only (e.g. possible self-interactions). $\mathcal{L}_{Higgs}$ accounts for a complex scalar field ϕ which fullfills all required symmetries. This Higgs field will play an important role in the formalism of *electroweak-symmetry-breaking* discussed at the end of this section. So far $\mathcal{L}_{SM}$ does not include any terms of the form $m^2\overline{\Psi}\Psi$ as they violate gauge-invariance. As a consequence, all gauge bosons and all fermions are without mass. With the help of symmetry-breaking the mediators of the weak force, $W^{\pm}$ and Z^0, get a proper mass term. Terms in $\mathcal{L}_{Yukawa}$ can be combined to produce all fermion masses.

Of course, by simply postulating a Lagrangian the interactions cannot be described properly. The goal of QFT is to calculate cross sections of particle interactions, i.e. to describe the dynamics of the theory. This is done by applying "Feynman rules" to a certain scatter process $ij \rightarrow kl$. The interaction is illustrated by Feynman graphs, see "Bhabba scattering" ($e^+e^- \rightarrow e^+e^-$) as an example in Figure 1.1. These graphs are calculation rules for the specific process, they are determined by the combination of interacting fields in $\mathcal{L}_{SM}$. The outer lines represent the initial- and final state particles; the inner lines represent the Feynman propagators; and the vertices combine inner and outer lines. The number of lines which end at a vertex is defined by the structure of the Lagrangian, i.e. which fields interact. Each vertex has a certain factor which includes the coupling constant of the specific force.

As Figure 1.1 shows, more than one graph can contribute to a single process, as only initial and final state are well defined. All possible inner lines are only subjected to the structure of $\mathcal{L}_{SM}$. As a consequence, *loop corrections* must be implemented. Loops are additional inner lines which can cause the cross sections to become divergent. This non-physical behaviour can only be solved by the formalism of *renormalisation*.
Up to now a local and lorentz-invariant theory with $\mathcal{L}(\phi_0; M_0, \lambda_0)$ has been described, where ϕ_0 is an arbitrary field, M_0 its mass parameter and λ_0 the coupling constant. If a "cut-off" Λ is defined which restricts the momenta of loops, than relations between the physical parameters (M, λ) and the "naked" parameters (index 0) can be obtained:

$$M = M(M_0, \lambda_0) \quad \text{and} \quad \lambda = \lambda(M_0, \lambda_0)\ . \tag{1.10}$$

These relations are divergent in Λ. The cut-off parameter has the meaning of an upper energy limit. The theory, i.e. the Standard Model, is only valid up to this energy regime. If $\mathcal{L}(\phi_0; M_0, \lambda_0)$ is a reasonable theory, then predictions made by the theory should not depend on the renormalisation scale Λ, i.e. they should be finite for $\Lambda \rightarrow \infty$. The divergence

of $M = M(M_0, \lambda_0)$ is no problem as M_0 and λ_0 are hypothetical parameters which cannot be measured by any physical apparatus.
Again, a complex formalism of renormalisation exists which cannot be explained in detail at this point. The essence is that besides the postulation of certain symmetries of a Lagrangian, a valid theory has to be renormalisable. This means that with the help of a *rescaling of parameters and fields*, all cross sections have to become finite. This requirement strongly restricts the multitude of possible terms contributing to $\mathcal{L}$. With the help of these complex techniques, it can be prooved that the Standard Model is a renormalisable theory. The strategy of constructing the Standard Model (symmetries and renormalistion) also holds for most extensions of the SM, e.g. Supersymmetry or Grand Unified Theory, and thus represents the core of QFT.

One might think that the scheme of loop corrections and renormalisation is only an abstract issue which does not affect physical observables. But it does affect them strongly. The rescaling of parameters and fields leads to a *scale dependence* of the couplings, e.g. for QCD:

$$\lambda(\mu_2) = \frac{\lambda(\mu_1)}{1 - \beta_0 \lambda(\mu_1) \ln(\frac{\mu_2^2}{\mu_1^2})} \quad \text{with} \quad \beta_0 < 0 \; . \tag{1.11}$$

Here, β_0 is a constant parameter and μ_2 is the energy scale where the coupling is tested. The other scale parameter μ_1 defines a fixed energy scale, and it is related to the Landau pole μ^* via:

$$\mu^* = \mu_1 \exp \frac{1}{2\beta_0 \lambda(\mu_1)} \; . \tag{1.12}$$

The Landau pole marks the energy region where the interaction becomes non-perturbartive. Equation 1.11 contains two important physical statements. The first is that $\lambda(\mu) \to 0$ for $\mu \to \infty$. This fact is called "asymptotic freedom" of QCD and means, that for very large energy scales quarks can be regarded as free particles which no longer interact strongly. Thus with the help of electrons and a large transfer momentum q^2 the electromagnetic substructure of hadrons can be probed (see experiments at Hera in Hamburg).
The second statement regards the μ^*. By measuring λ at a certain scale, μ^* can be calculated using the above expression. The result is that μ^* amounts to several hundred MeV, the typical scale of hadron masses. For $\mu \lesssim 1$ GeV the coupling becomes so large that free quarks cannot be observed. This is called the "confinement of quarks and gluons".

In Figure 1.2 the scale dependence of the forces is shown. In addition to the different slopes, one can see that there is a point where the electromagnetic and the weak force merge into a single force, the *electroweak* force. This represents another important issue in QFT: the "spontaneous breaking of a symmetry". The Higgs-mechanism, which was mentioned earlier and stated as the origin of mass, is closely linked to this unification of forces.
The Higgs-mechanism starts with the task of defining the ground state of the Standard Model. The Higgs field ϕ in $\mathcal{L}_{Higgs}$ is the only candidate for such a ground state. Even though the spectrum of the possible ground states is degenerated, nature realizes a specific state ϕ_0. Whereas the Lagrangian is not affected by this particular choice, the state itself

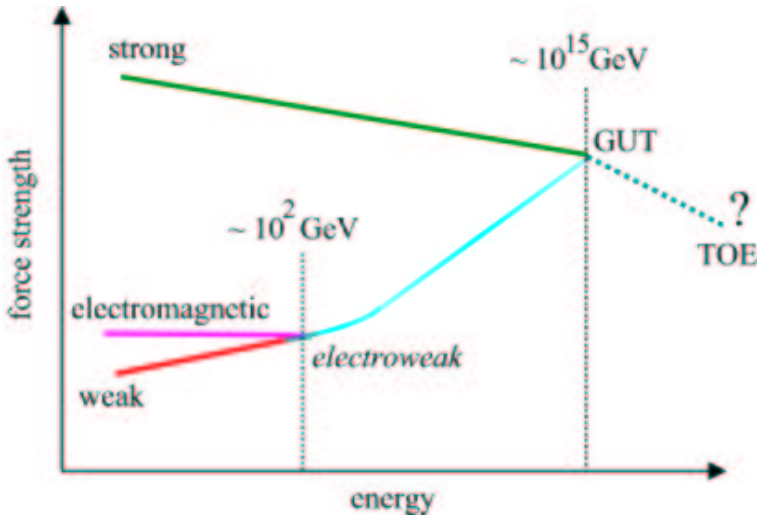

Figure 1.2: *Scale dependence and unification of forces [3].*

is no longer invariant under the electroweak symmetry transformation $SU(2) \times U(1)_Y$, so the symmetry is broken spontaneously. The Higgs field can be written as:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} . \tag{1.13}$$

Here v represents the constant value of the minimum and $H(x)$ is the dynamical Higgs field. If this expression is inserted into the Lagrangian $\mathcal{L}_{SM}$, the parts with couplings between the Higgs field and the gauge boson and fermion fields generate mass terms, e.g. terms in $\mathcal{L}_{Yukawa}$ like $\sim v\overline{\Psi}\Psi$. In this way the experimentally observed massive gauge bosons and massive fermions can be derived within QFT and the Standard Model.
The ground state of the Higgs field introduces a scale for the theory and determines the mass of the Higgs particle. If we assume $v \approx 250$ GeV, then at scales much higher than the scale of the ground state, v can be neglected and thus all mass terms diminish again. Without these mass terms W and Z become massless, and the unified electroweak force emerges.

The description of the Higgs-mechanism concludes this overview of the Standard Model and its founding principles. As the energy scale of the electroweak symmetry breaking is just around the TeV scale where current experiments are running, there is anticipation that very interesting physics can be observed in this unexplored regime. The Higgs particle itself should have a mass of this magnitude, and extensions of the Standard Model could be verified (e.g. SUSY). Using the techniques described above results in the construction of even higher symmetries, corresponding to Lagrangians with renormalisable fields and thus new theories. These theories are designed to solve problems of the SM, such as the unification of all forces or the origin of the 18 "nature given" parameters of the Standard Model. The symmetries of these new theories are expected to be broken at approximately the TeV scale, and the discovery of New Physics might be possible.

1.2 Hadron Collider Physics

At a hadron collider, e.g. the Tevatron, protons collide with anti-protons. Both are not pointlike particles, but they consist of several partons. The simple picture of three quarks (uud) building the proton has to be replaced by the complex formalism of the proton's *structure function*:

$$F_2 = \sum_i q_i^2 x[f_i(x,Q^2) + \bar{f}_i(x,Q^2)] + \sum_j q_j^2 x[g_j(x,Q^2)] \; . \tag{1.14}$$

Here the sum includes all quark flavours $i = u, d, ...$; q_i is the specific charge of the quark and Q^2 is the momentum transfer of the interaction. In addition to the fermion part of F_2, bosonic partons with a "color charge" q_j also contribute, accounting for the gluons inside the nucleon. This structure function defines the composition of the proton:
The variable x is called "Bjorken x" and redistributes the four-momentum of the incoming proton P^μ. Each parton gets a fraction resulting in the four-momentum k^μ, i.e. $k^\mu = xP^\mu$. In deep inelastic scattering, only two partons interact. As each of the partons has a certain fraction x of the total momentum P^μ, the entire center of mass energy of the accelerator is not available for the reaction. This is the reason why the laboratory frame (detector) is not the rest system, and most events are boosted along the z-axis.
$f_i(x,Q^2)$ is the parton density function (pdf), i.e. the density of quarks with flavour i which have a relative momentum between x and x+dx. Equivalently, $g_j(x,Q^2)$ represents the gluon density. What Equation 1.14 means is that a proton consists of a combination of all quarks, anti-quarks and gluons. In general, a distinction is drawn between "valence quarks" (uud for proton) and "sea quarks", which are both part of a hadron. The existence of sea quarks, e.g. a virtual strange quark in a neutron, can be explained by the presence of gluons keeping the hadron together. Besides the exchange of gluons between two valence quarks, the radiation of gluons converting into a quark anti-quark pair is also possible. As a consequence, anti-quarks and heavy quarks can be created, even though their density function for large x is naturally much smaller than, for example, the one of a u-quark, as they are the consequence of radiation processes.
In addition to the momentum given to the sea quarks, experimental data show that the gluons take about half of the initial momentum P^μ. Figure 1.3 summarizes the different shares of the components of the proton (valence quark density, sea quark density and gluon density). One can see that for small x, the gluons and the sea quarks dominate.

In Equation 1.14, the structure function and all parton densities are not only a function of x, but they also depend on the momentum transfer Q^2. This "scaling violation" results in a structure function F_2 which is shifted towards smaller x-values as $|Q^2|$ rises. If the momentum transfer rises, gluon radiation is enhanced, and the proton is thus dominated by gluons and sea quarks. At a TeV collider, valence quarks are no longer the main contributor to the total cross section of inelastic scattering.

These pdf's are all determined by experiments (e.g. Hera) as theroretical QCD calculations are very complicated. Monte Carlo simulations need these parton density functions as an

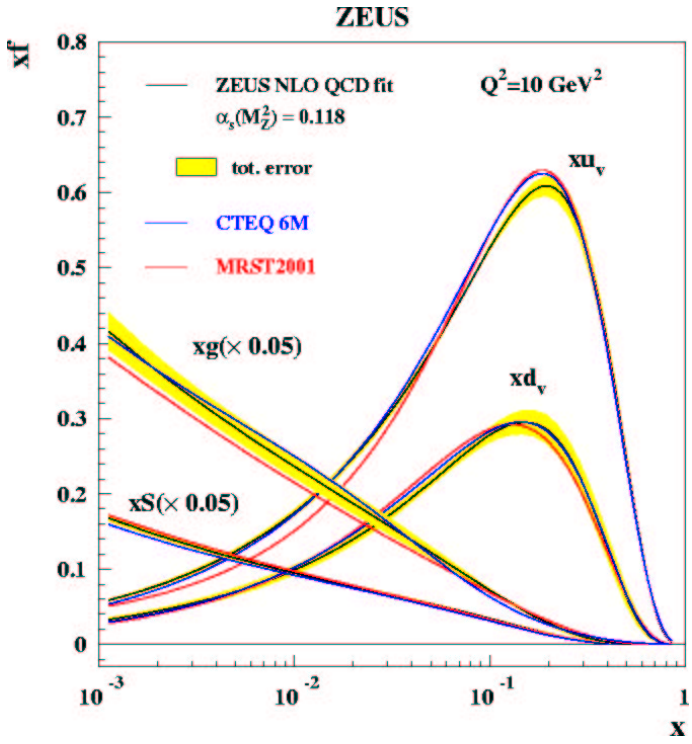

Figure 1.3: *Parton density functions of the proton times momentum fraction (xf), ZEUS data compared to the parametrisations CTEQ and MRST. Valence quarks (u_v, d_v), sea quarks (S) and gluon density (g) are plotted as a function of the momentum fraction x [4].*

input, so different parametrizations of the experimental data are determined. Examples for different parametrizations can be found in Figure 1.3. The cross section of a specific process depends on the assumption of a certain pdf, so an inherent systematic uncertainty is attached to the choice of a pdf (see Section 8.2). Parton density function and differential cross section of a process $p + \bar{p} \rightarrow F$ are connected by:

$$\frac{\mathrm{d}\sigma_F(\sqrt{s}, Q^2)}{\mathrm{d}Q^2} = \sum_{i,j} \int \mathrm{d}x_i \mathrm{d}x_j \, f_i(x_i, Q^2) f_j(x_j, Q^2) \frac{\mathrm{d}\sigma_F^{ij}(x_i, x_j, Q^2)}{\mathrm{d}Q^2} \, . \tag{1.15}$$

Here, F represents an arbitrary final state and $\sqrt{s}$ is the center of mass energy. One has to include all possible processes $parton_i + parton_j \rightarrow F$. All these contributions are incorporated in the differential cross section of the individual process $\sigma_F^{ij}(x_i, x_j, Q^2)$, then these are weighted according to the specific parton density functions $f_{i,j}$.

After the principles of cross section determination at a hadron collider have been introduced, the total cross sections of different process like $b\bar{b}$-production or W-production have to be discussed. Figure 1.4 shows the total cross sections of various Standard Model processes as a function of the center of mass energy. Arrows mark the energy of different facilities like Tevatron or LHC (Large Hadron Collider). Considering Tevatron as an example, one can see that QCD-processes like $b\bar{b}$-production with $\sigma \approx 20000$ nb are orders of magnitude larger than the processes important for many analyses, e.g. $W \rightarrow e\nu$ with $\sigma \approx 3$ nb. Most of the events created at a hadron collider are QCD-only events without

any leptons in the final state. The triggers of an experiment must be able to filter these events as the amount of stored data would exceed resources available. As QCD processes dominate, a physics analysis must deal with a large background of all-jet events, mimicking signatures important for the specific analysis (see Section 6.2).

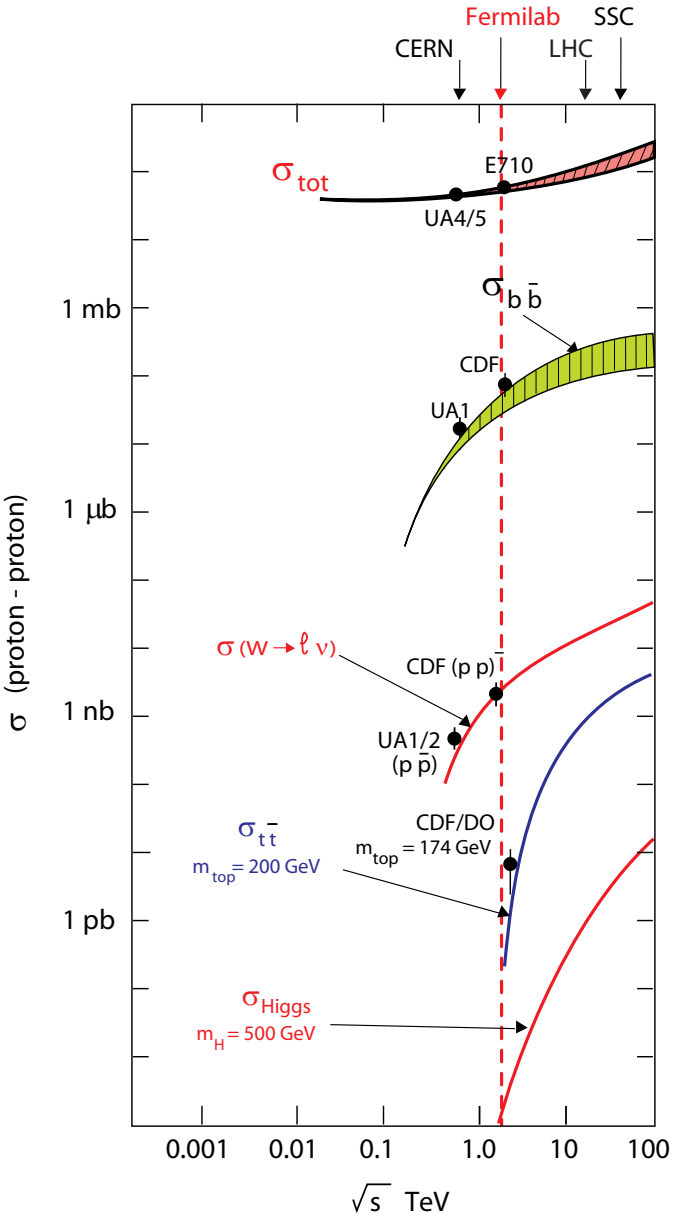

Figure 1.4: *Different total cross sections as a function of the center of mass energy $\sqrt{s}$ of a hadron collider [5].*

1.3 Important Quantities of High Energy Physics

This section introduces certain fundamental quantities which characterize the physics at hadron colliders.

1.3.1 The Coordinate System and Kinematic Angles

First of all a coordinate system has to be chosen to describe the kinematics of the initial and final state particles. Therefore the nominal interaction point, i.e. the geometric center of the detector, is located at the origin, and the z-axis runs parallel to the beam. The two other axes x and y span the transverse plane which is very important for the measurement of the tranverse momentum p_T. In order to obtain complete momentum information, the polar angle θ must be measured.
The direction of particles is defined by two variables: The azimuthal angle φ and the pseudorapidity η. The pseudorapidity is defined only in the high energy limit where rest masses can be neglected. Here, a connection between η and the angle θ in the rest system of the detector can be utilized:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \ . \tag{1.16}$$

For particles escaping perpendicular to the beam pipe ($\theta = 90°$) the pseudorapidity is zero. Towards $+z$ the pseudorapidity is positive and rising infinitely as θ becomes smaller. The same behaviour can be found towards $-z$ with the difference that the pseudorapidity reaches negative values.
The choice of η is convenient as differences in the pseudorapidity $\Delta\eta$ are Lorentz-invariants. As partons with different momenta along z interact with each other, each event is subjected to a Lorentz-boost, so invariant quantities are needed. Differences in the azimuthal angle $\Delta\varphi$ are also invariant, so a variable ΔR defining the spatial distance of two detector objects can be constructed:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} \ . \tag{1.17}$$

1.3.2 The Kinematic Variables of p_T and MET

At a hadron collider particles with a substructure (protons and anti-protons) are accelerated. As a consequence all events are boosted in the $\pm z$-direction. The tranverse component of the momentum, p_T, is not affected by this boost, so events are characterized best by the transverse momenta of the particles observed by the detector:

$$p_T = \sqrt{(p_x)^2 + (p_y)^2} = p \cdot \sin\theta \ . \tag{1.18}$$

As the center of mass energy is limited, objects with very large transverse momentum have only little momentum along the $z - axis$. They are thus likely to have a small pseudo-

rapidity η and can be measured within the well instrumented central region of the detector.

For all physical objects detected, p_x and p_y can be determined. As the colliding proton beam and anti-proton beam circle along z, no transverse momentum is present prior to the collision. The law of momentum conservation demands that also after the interaction, no overall p_T is allowed to remain. Momentum balance implies:

$$\left(\sum_{particles} p_{x,y}\right)^{before\ interaction} = \left(\sum_{particles} p_{x,y}\right)^{after\ interaction} = 0\ . \tag{1.19}$$

Extending the momentum balance also to the longitudinal z-component of the momentum is not possible as the detector is not hermetic. After the hard scattering, some fragments of the proton and the anti-proton can remain undetected while by escaping along the beam pipe.

In addition to particles flying in the very forward region of the detector, particles not interacting with the detector material are also not part of the energy balance in p_x and p_y. Examples for this would be neutrinos or non-SM particles which do not interaction with the detector material. With the help of momentum conservation, information about the transverse component of the Missing Transverse Energy (MET) can be found:

$$\left(\sum_{particles} p_{x,y}\right)^{\text{after interaction}} + MET_{x,y} = 0 \tag{1.20}$$

$$\Longrightarrow MET = \sqrt{(MET_x)^2 + (MET_y)^2}\ . \tag{1.21}$$

1.3.3 Luminosity and Cross Section

The most important properties of a particle accelerator are its center of mass energy $\sqrt{s}$ and its instantaneous luminosity l. The reason for this is a simple connection between the event rate $R = \frac{\mathrm{d}N}{\mathrm{d}t}$, the cross section of a certain interaction σ and the instanteneous luminosity l:

$$R = \sigma \cdot l\ . \tag{1.22}$$

Event rates are what a detector measures. As the cross section of a specific process is a function of the available energy $\sqrt{s}$, event rates are both affected by the amount of particle acceleration and the luminosity.

In principle, luminosity characterizes the intensity of the two beams heading for each other and their mutual penetration. In order to enhance the probability of a hard process, many particles should meet at an area as tiny as possible. Suppose a proton beam and an anti-proton beam divided into B seperate bunches with a Gaussian bunch profile, each

containing n_p particles and $n_{\overline{p}}$ particles, respectively. The profile of the beam cycling with a frequency f can be described by its widths σ_x and σ_y; luminosity is given by:

$$l = f \cdot \frac{B\ n_p\ n_{\overline{p}}}{4\pi\ \sigma_x\ \sigma_y}\ . \tag{1.23}$$

The beam optics of the accelerator make sure that the profile of the beam is minimized within the detector using focusing magnets, thus leading to high luminosity and high event rates.
As data are recorded during a certain amount of time $[t_1, t_2]$, expression 1.22 has to be integrated. The quantity $\mathcal{L} = \int_{t_1}^{t_2} l\, \mathrm{d}t$ is called integrated luminosity (unit [1 μb^{-1}=10^{30} cm^{-2}]) and is a measure of the amount of events seen by the experiment.

The cross section σ of a certain Standard Model process, e.g. W-production, characterizes how probable this interaction is if the two initial particles collide. It has the unit of an area (1 pb=10^{-36} cm^2) which suggests the association of two balls with a certain profile colliding. If this surface is large, then a deflection is probable. As elementary particles interact via the fundamental forces, the cross section is a measure of the strengths of these forces and the probability of a certain interaction. If this probability is large and the luminosity of the beam relatively high, then large event rates can be expected of this specific process.

Chapter 2

Tevatron and the DØ-Detector

2.1 The Tevatron Accelerator

This analysis is part of the international DØ-collaboration, one of the biggest experiments in particle physics in the world. Together with the second big experiment (CDF), both detectors are located at *FERMILAB* (FNAL, Fermi National Accelerator Laboratory), named after the Nobel Prize winner Enrico Fermi, who lived and worked in Chicago. Fermilab is located at an area of 27.5 km^2 in Batavia near Chicago, Illinois. Figure 2.1 shows an aerial view of the whole laboratory terrain. One can easily see the two big rings, the Main Injector (lower ring) and the Tevatron (upper ring). The Tevatron is a proton-antiproton collider at the world's highest center of mass energy of $\sqrt{s} = 1.96$ TeV. Besides this big machine and its two detectors DØ and CDF, various other experiments like MINOS and MiniBooNE are located at Fermilab, and the rest of the site is a natural reserve and even accommodates a herd of American bison.

Fermilab was founded on November 21, 1967, under the Director R. Wilson. During its existence excellent research has been performed which resulted in several groundbreaking discoveries. Three major components of the Standard Model were discovered at Fermilab: The bottom quark (May-June 1977), and the top quark (February 1995) in Collider Run I. In July 2000, Fermilab experimenters announced the first direct observation of the tau neutrino, the last part of the three generations of matter to be observed. With the inauguration of Collider Run II of the Tevatron in March 2001 along with the DØ-detector Run II upgrade, discoveries like the Higgs, as the final slot of the Standard Model, and the detection of New Physics were aimed.

The Machine Operation

The Tevatron is a collider which has been upgraded several times in order to reach new frontiers in energy and luminosity. The old Tevatron tunnel was recycled for Run I and II, and the whole complex of the Main Injector had to be built in order to reach the design instantaneous luminosity of $l = 5 \cdot 10^{31}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ for Run II. With a radius of 1 km

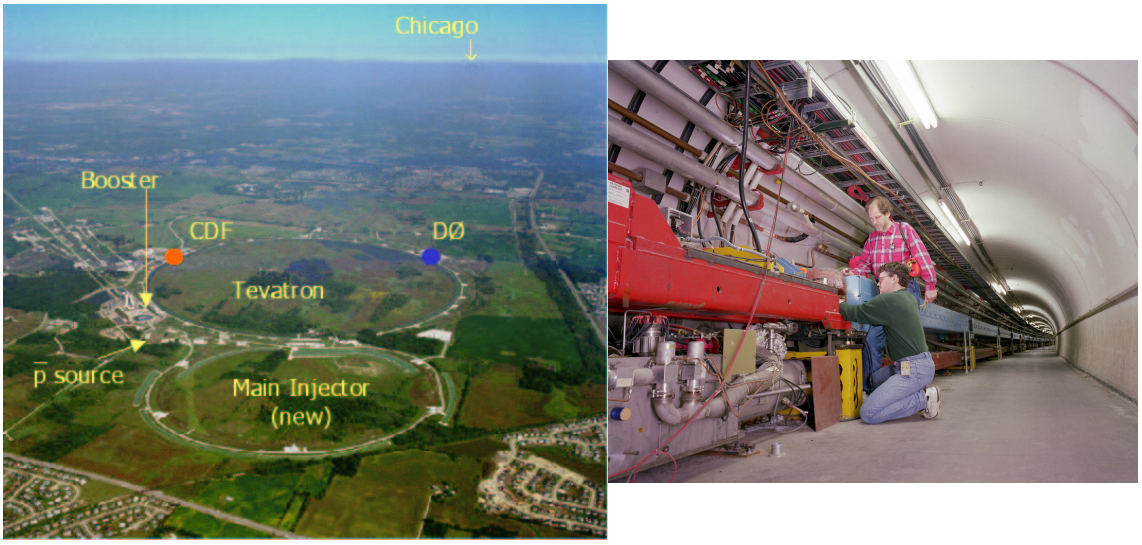

Figure 2.1: *Aerial view of Fermilab (left) and inside the tunnel of the Tevatron (right) [6].*

the Tevatron reaches a center of mass energy of $\sqrt{s} = 1.96$ TeV. Of course this immense energy is only achieved by using a chain of accelerators. The following text summarizes the Tevatron structure, which is displayed in Fig. 2.2, details can be found in [7]/[8].

There are two separate tasks which are assigned to the Tevatron machine: (1) cumulating enough protons and anti-protons to form two separate beams running in a joint beam pipe, and (2) accelerating these particles to the desired energy and focusing them inside the detectors in order to collide. The beam formation process starts with H-atoms, which get an additional electron to build negatively charged ions. With a Cockcroft-Walton-accelerator, basically providing a very huge voltage, the H^- ions reach an energy of 750 keV and are injected in the linear accelerator (LINAC), where drift tubes enhance the energy up to 400 MeV. Then the ions lose their electrons by traversing carbon foil, and a proton beam enters the first of three "synchrotons", the Booster. Charged particles in a homogeneous magnetic field follow the law $p = qrB$ (p is the particle momentum, q its charge, r the radius of the ring and B the machine's magnetic field). Thus in order to increase the particle's momentum the magnetic field has to be increased as the radius is a fixed parameter here. This is what a synchrotron does: While "cavities" continuously accelerate the particles with the help of electromagnetic waves, the magnetic field is raised simultaneously. At the end 8 GeV protons leave the Booster and enter the Main Injector, the first big ring with a radius of 0.5 km. Here the proton beam reaches an energy of 120 GeV, but no anti-proton beam has yet been created.
Therefore the protons leave the Main Injector and hit the Anti-Proton Source, a nickel target. A magnet sorts out anti-protons and rejects all other secondary particles which dominate the reaction as 10^5 protons are needed to create only one anti-proton. The anti-protons vary considerably in energy, so a "debuncher" creates anti-protons of 8 GeV mean energy using the technique of stochastic cooling. These anti-protons are collected in the Accumulator. Then a new proton beam is filled in the Main Injector and accelerated to an energy of 150 GeV. After this beam is injected into the Tevatron, the accumulated anti-protons enter the Main Injector, gain energy and also are filled into the Tevatron.

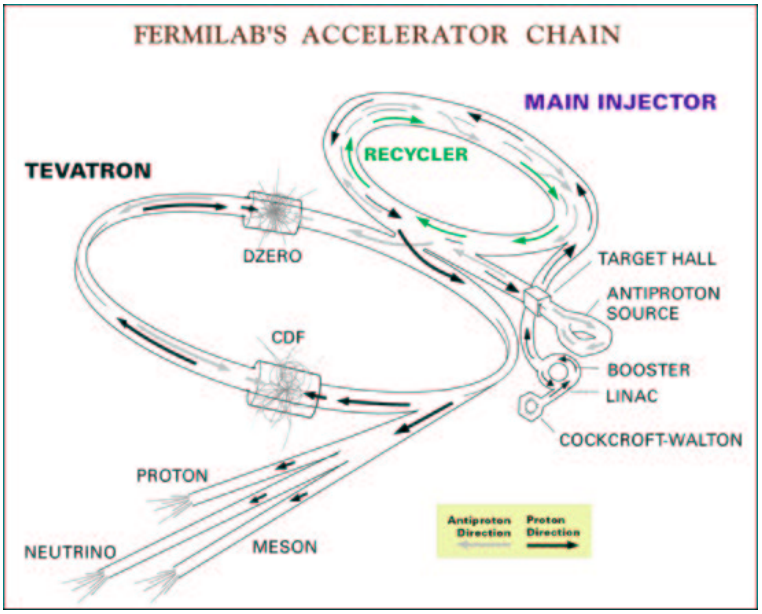

Figure 2.2: *The chain of acceleration [9].*

Finally both the proton and the anti-proton beam cycle the Tevatron ring, the protons running clockwise and the anti-protons counterclockwise. Within this last synchrotron the designed beam energy of 0.98 TeV is achieved and a Store begins, i.e. the beams revolve up to one day and actual collisions are induced. A magnetic field of 4.2 T is needed to keep the particles in their trajectory, a technological challenge. The Tevatron was a pioneer using more than 1000 superconducting dipole magnets for this purpose. In Figure 2.1 one can see the inside of the Tevatron tunnel with its sequence of magnets enclosing the beam pipe. To guarantee superconduction the magnets are held at a temperature of 4.2 K cooled by liquid Helium. The need for superconduction leads to the risk of "quenches", magnets with small seeds of heat which end the state of superconduction. With any sudden drop in the magnetic field the beam is lost, which can and has several times resulted in destroyed magnets or damage done to the detectors. At two points along the Tevatron where the experiments CDF and DØ are located magnets focus the proton and anti-proton beams, minimizing the diameter of the beam (emittance) to increase luminosity. At these two crossing points the detectors have the task of measuring the final states of these collisions, identifying the particles, and quantifying their kinematic properties. In addition, a 120 GeV proton beam can be extracted from the Tevatron to the Switchyard to operate fixed target experiments.

The rate of the collisions is determined by the substructure of the beam. Protons and anti-protons are gathered in "bunches", one of them containing $\mathcal{O}(10^{11})$ protons or $\mathcal{O}(10^{10})$ anti-protons, respectively. A group of 12 bunches builds a "train" with a temporal separation of 396 ns between the bunches. In total three trains cycle inside the Tevatron at a separation of 7 µs. This leads to a mean rate of $1,700,000$ bunch crossings per second. The detectors have to cope with this immense flood of data in order to identify interesting physics, i.e. events with hard scatter processes.

During a store the detectors take data, so it is desired to maximize the initial luminosity of a store and the overall duration of the store. The record for instantaneous luminosity

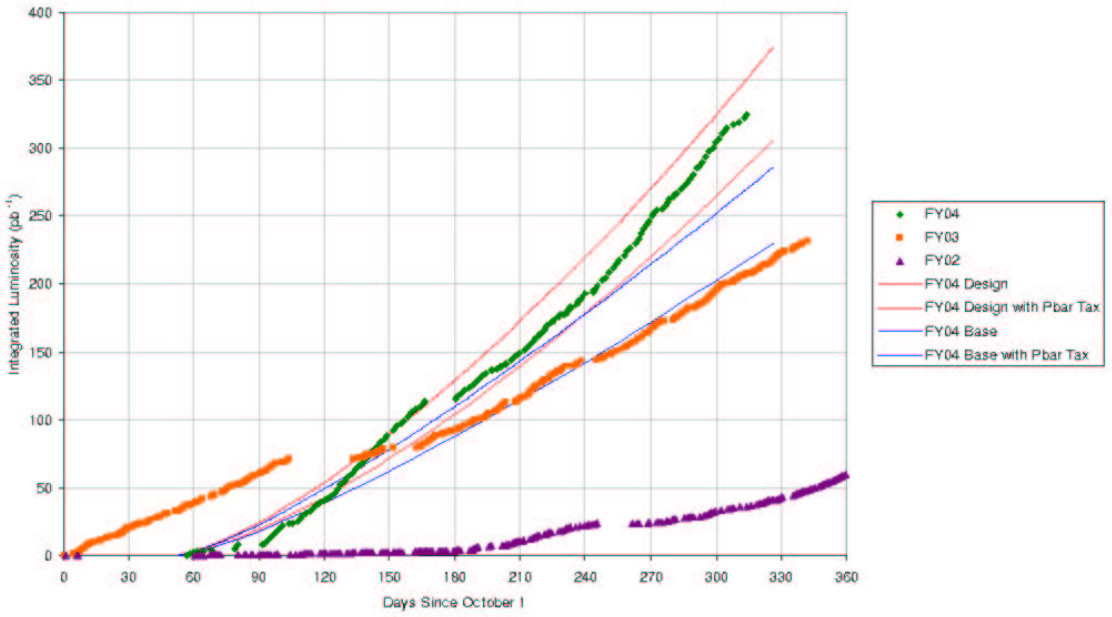

Figure 2.3: *Luminosity statistics [8].*

was at the beginning of store 3657 was $l = 1.02 \cdot 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, and the same store on $16-17$ June 2004 delivered the highest integrated luminosity of $\mathcal{L} = 4463\,\mathrm{nb}^{-1}$. Just like an hour can be devided into minutes and seconds, a store has its subsets called Run and Luminosity Block Number (LBN). During a store the initial luminosity declines exponentially because of the collisions and inevitable beam losses. At certain points of this decline the rates at which the experiments take data are enhanced by lowering the prescales of triggers (see Section 2.3). Each time these scale factors are modified, a new Run begins. Finally, a level of $\sim 1/4$ of the initial luminosity is reached, and the store is ended by directing both beams to a fixed absorber outside the ring (a store can last up to 24 hours). As mentioned earlier the accumulation of anti-protons is very inefficient, even if a new gathering of anti-protons starts at the beginning of each store. A new mode of operation is planned where anti-protons are continuously gathered by the Recycler, a separate ring of the Main Injector, to allow optimal operation of the Accumulator. On June 9, 2004, the first "mixed-pbar shot" was successful, combining 8 GeV anti-protons from the Recycler with 8 GeV anti-protons from the Accumulator to be injected into the Main Injector. This new integration of the Recycler promises to increase the Tevatron luminosity even further as shorter Stores with higher closing luminosities become possible.
In the past years the machine performance has increased steadily and regained reliability which many people doubted regarding the problems at the beginning of Collider Run II. Tevatron has just ended a long phase of shutdown until Nov 2004 where several parts of the machine (electron cooling for Recycler) and its detectors were revised. As of now the delivered integrated luminosity has exceeded its design values, see Figure 2.3 where data in fiscal year 2004 are above Design with Pbar Tax (a certain amount of anti-protons is used for studies at the Recycler and is thus lost for the luminosity of the Tevatron). The future looks promising that several fb^{-1} of data will reveal New Physics.

2.2 The DØ-Detector

Like the Tevatron the detector had to be upgraded for Run II to meet the requirements of the new high luminosity environment. Therefore the trigger capabilities of all subsystems were enhanced, a new radiation resistant and more precise tracking system was installed, and the old muon system was supplemented while keeping the old strength of full coverage in calorimetry and muon detection. The result is the modern particle detector displayed in Figure 2.4 and Figure 2.5 with the impressive measures of 20 m length and 13 m hight. The DØ-Detector can be subdivided into three big subsystems: (a) The inner part is the Inner Tracking System surrounded by the 2 T superconducting solenoid; (b) the Calorimeter cryostat; and (c) the Muon System with its embedded 1.8 T toroid magnet. All these parts and their function in measuring particles will be introduced in the following paragraphs. Details can be found in [10] and [11].

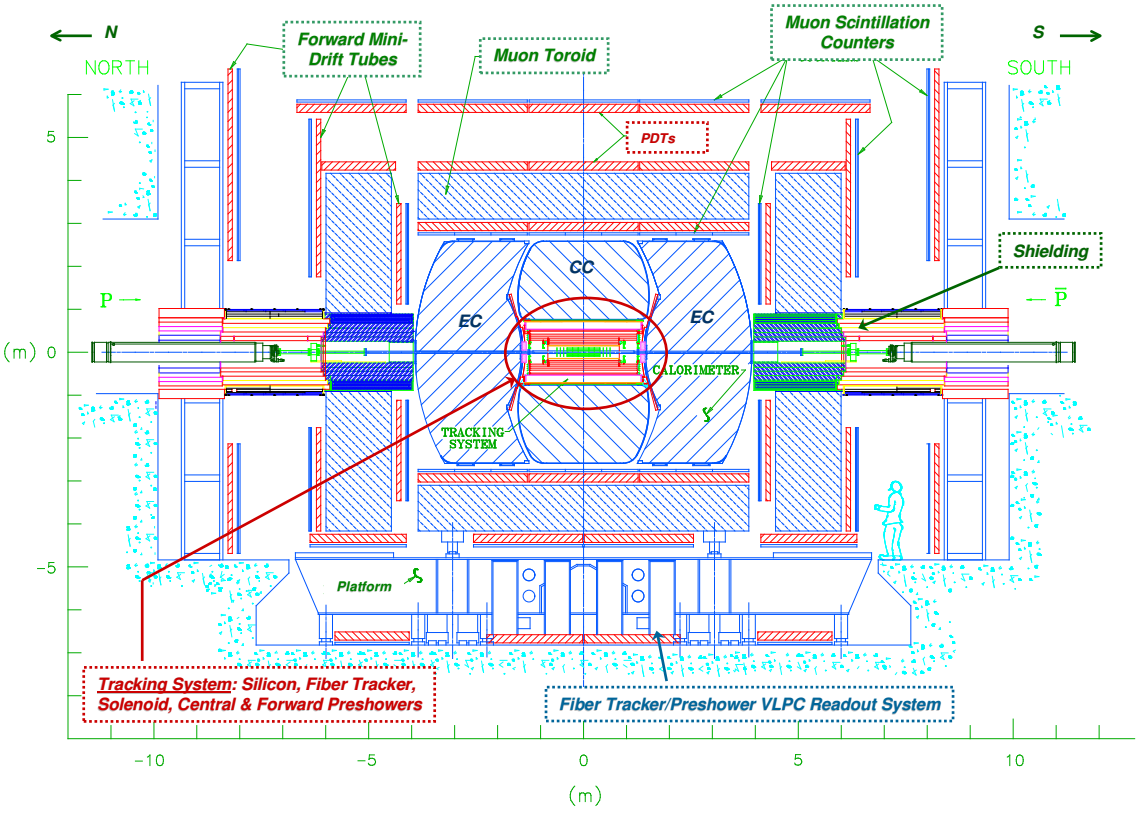

Figure 2.4: *Schematic profile of the DØ detector with all its subsystems [6].*

2.2.1 The Inner Tracking System

In a tracker the momenta of charged particles are determined by their curvature in a magnetic field. This extremely uniform field is generated by the 2.8 m long solenoid magnet enclosing the tracking detectors. This superconducting magnet has to be cooled so that solenoid and cryostat represent 1.1 radiation lengths of dead material. On the other hand, the tracker provides a precision charged particle momentum measurement of $\Delta(p_T)/p_T^2 \approx 0.002\ \mathrm{GeV}^{-1}$, a value much better than the resolution of the muon system.

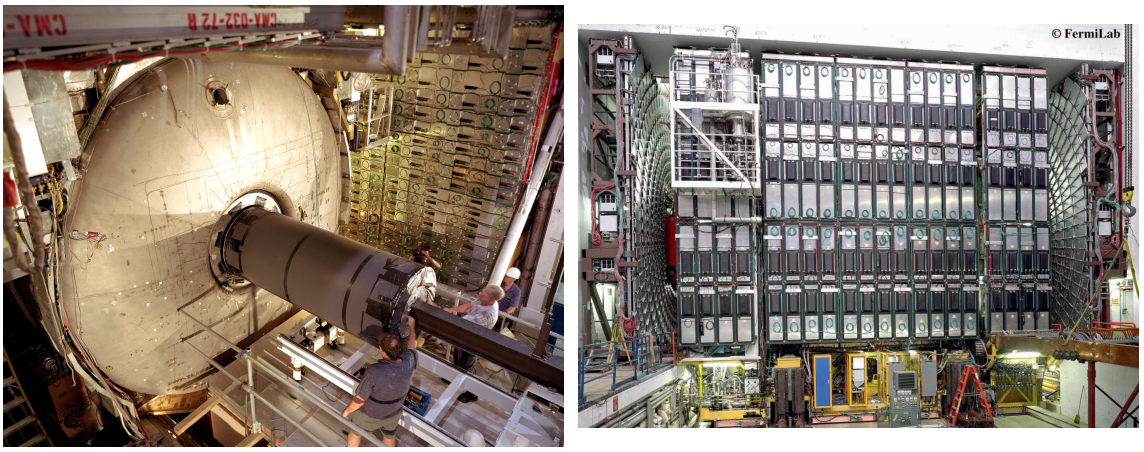

Figure 2.5: *On the left picture the open detector is shown during the installation of the Inner Tracker. On the right an exterior view can be seen with the scintillation counters of the Muon System building the outer shell [6].*

As a consequence this analysis uses only muon momentum information measured with the tracker.

The tracker itself consists of two layers: The Silicon Vertex Detector *SMT* being the closest to the nominal vertex and the Scintillating Fiber Tracker *CFT*, both displayed in Figure 2.6. The SMT is the high resolution part of the tracking system, enabling secondary vertex measurement, which is important for B-physics and identifying b-jets. Since the collider interaction point is extended in z with $\sigma_z \approx 25$ cm, a combination of barrel detectors measuring the coordinates $r - \varphi$ and disk detectors measuring $r - z$ and $r - \varphi$ was chosen. Because of these disk detectors, even particles with large pseudorapidities can have SMT- hits, thus a specific η-coverage is dependent on the amount of required SMT-hits. The barrel detectors consist of 4 layers of 50 and 60.5 µm silicon strips, most of the layers consisting of double-sided detectors for two-coordinate information. Each layer is divided into six segments, 12 cm per segment. All six segments are sealed off by one of the 12 double sided "F" disks and the six remaining disks are located at each side next to the barrel. Four large diameter single-sided "H" disks enclose the SMT at $|z| = 110$ cm and $|z| = 120$ cm, covering the very forward η-region.

To enhance tracking efficiency, the CFT, covering the range $|\eta| < 1.62$ (with in principle 16 measurement points available), is the second component of the tracker. This yields more hits available for the track reconstruction. The CFT consists of $74,000$ scintillating fibers mounted on eight concentric cylinders at radii from 19.5 to 51.5 cm surrounding the silicon vertex detector. The cylinders support a doublet layer of fibers and four having a doublet at stereo angles for information on z-coordinates. These fibers consist of three materials increasing light trapping and mechanical robustness. The scintillated light is guided by 11 m long fiber waveguides to Visible Light Photon Counters, cryogenic photomultipliers with a quantum efficiency of $\sim 70\%$ capable of detecting single photons.

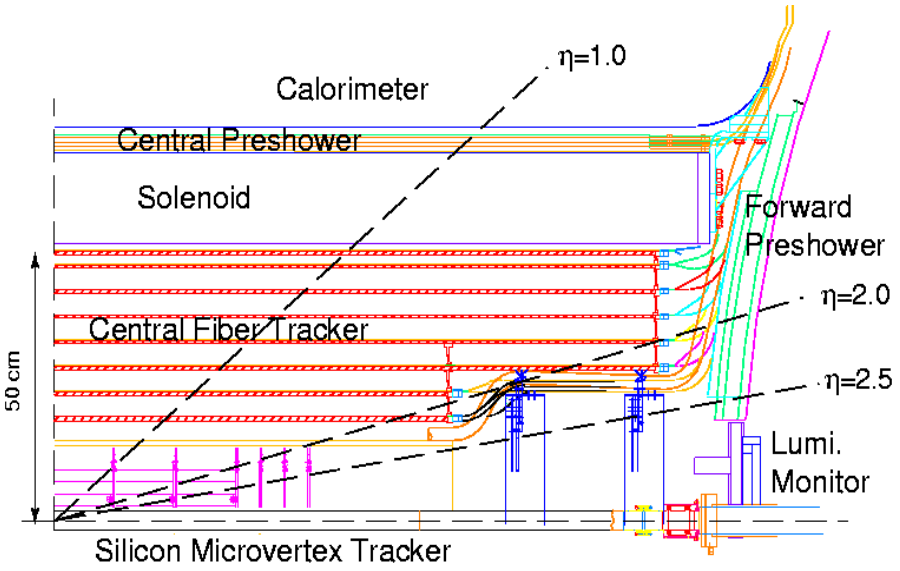

Figure 2.6: *Schematic view of one quadrant of the Inner Tracking System.*

The muon-resolution of the tracker can be parameterized as (see [12]):

$$\frac{\sigma(\frac{1}{p_T})}{\frac{1}{p_T}} = \sqrt{a^2 \cdot p_T^2 + b^2} \tag{2.1}$$

$$\text{and} \quad \sigma(p_T) = \sigma(\frac{1}{p_T}) \cdot p_T^2 \, . \tag{2.2}$$

This means that the error of the measurement increases dramatically as p_T becomes larger. Parameter $a = 0.002\,\text{GeV}^{-1}$ is caused by the limited coordinate resolution; parameter $b = 0.03$ accounts for the effect of multiple scattering. In order to minimize this error and enable a viable measurement of high p_T muons, a combination of SMT and CFT is needed. This conclusion is supported by Figure 2.7 where the relative error of the sagitta ($\sim 1/p_T$) measurement as a function of p_T is plotted ($W \to \mu\nu$ Monte Carlo events). One can easily identify two straight lines with two different slopes, the slope corresponding to the parameter a. The one with the bigger slope corresponds to tracks reconstructed by CFT-only hits, the other one to tracks including ≥ 3 SMT-hits. This illustrates that the SMT-detector has to be used to achieve adequate momentum resolution.

2.2.2 The Calorimeter

No modifications to the core of the original Run I uranium liquid-argon calorimeter itself were made for Run II, but the readout electronics were replaced, the Central/Forward Preshower Detector was designed and the readout phototubes of the Intercryostat Detector (ICD) were moved outside the magnetic field.
The Central Preshower Detector is placed in the gap between the solenoid coil and the central calorimeter cryostat and covers a region of $|\eta| < 1.2$. It is designed to aid electron identification by making precision position measurements, support triggering and to correct electromagnetic energy for the inevitable showering effects of the solenoid. Together

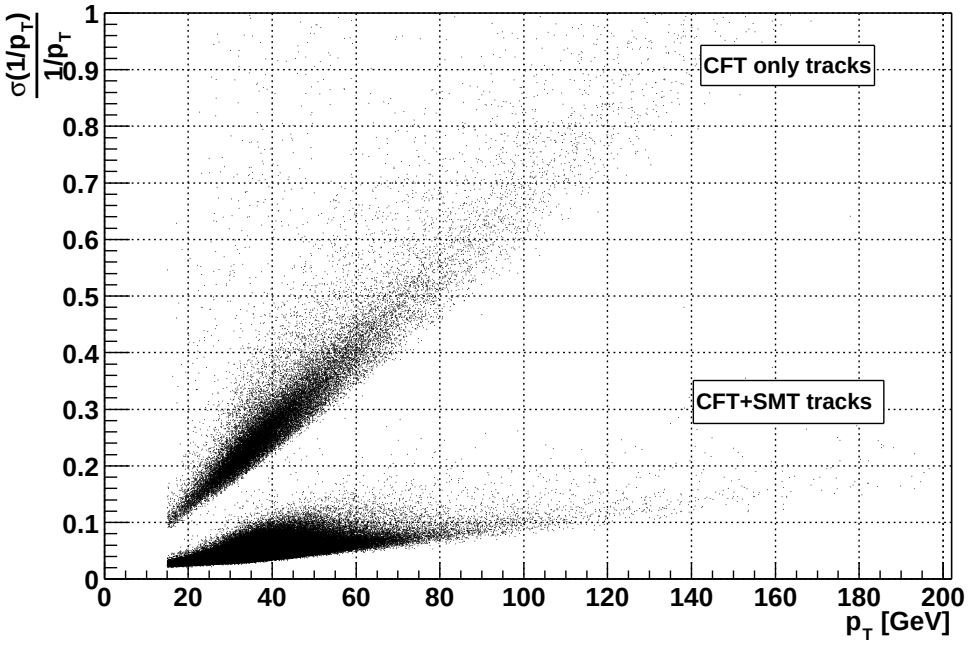

Figure 2.7: *p_T resolution of the Inner Tracking System, a $W \rightarrow \mu\nu$ Monte Carlo simulation is shown.*

with an additional lead absorber, the material before these 3 layers of scintillating strips totals two radiation lengths.
The Forward Preshower Detector at $1.4 < |\eta| < 2.5$ is constructed similarly, with additional lead absorbers accounting for the absence of the solenoid. Both detectors can be seen in Figure 2.6 outside the inner tracker. The data analysed in this paper do not make use of both of these Preshower Detectors as they are still under commissioning and the software is not completed yet.
The next layer of the detector is the massive cryostat accommodating the calorimeter. In order to enable access to the Inner Tracker, the calorimeter is divided into the central calorimeter (CC) covering $|\eta| < 1$ and a pair of end-cap calorimeters (ECN and ECS) leading to an overall coverage of $|\eta| < 4$. The boundary between CC and EC is instrumented with the Intercryostat detectors, two scintillation counter arrays, in order to correct for energy deposited in the uninstrumented walls.
Calorimetry must provide the energy measurement and identification of electrons, photons and jets. This is realized by three distinct types of modules, all using the principle of metal absorber plates and signal boards embedded in the active medium liquid argon: An electromagnetic section (EM) consisting of four layers with thin uranium absorber plates, three layers of fine-hadronic section (FH) with thicker uranium plates, and finally a coarse-hadronic section (CH) with thick copper or steel plates. Electromagnetic objects are completely absorbed within the EM, in contrast to hadronic jets, which begin showering in the EM-section but are not stopped until they reach the outer coarse plates which protect the outer Muon System against jet remnants. In addition to this, EM and hadronic objects can be distinguished by their shower profiles measured by combining different calorimeter cells. The measurement of the shower shape also leads to the design of "pseudo-projective"

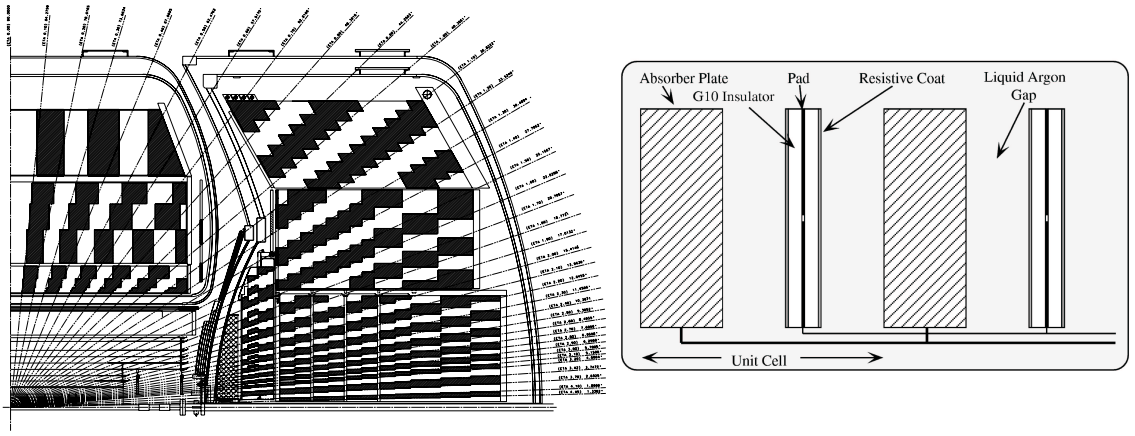

Figure 2.8: *Cross section of the liquid-argon calorimeter (left) [13] and zoomed view of a single cell (right) [14]*

readout towers, which means that the centers of cells of increasing shower depth lie on rays projecting out from the center of the interaction region (see Figure 2.8). These towers have typical transverse sizes of $\Delta\eta = 0.1$ and $\Delta\varphi = 0.1$, limiting the resolution in η and φ; only the third EM layer is segmented finer ($\Delta\eta = \Delta\varphi = 0.05$) to enable a better shower shape measurement and to enhance the matching between a track and a calorimeter entry. The complete EM calorimeter represents 20.6 radiation lengths of material. Fine and coarse hadronic calorimeters sum up to $6.4\lambda_A$, where λ_A is the absorption lengths. This results in the total radial dimension of the whole calorimeter of ≈ 1.60 m at $|\eta| = 0$.

Figure 2.8 shows a schematic view of the liquid argon gap and signal board unit cell. EM, FH and CH only vary in the thickness of the absorber plates and the gaps. A high voltage of $\approx 2.0 - 2.5$ kV leads to a fast extraction of the electrons created by an ionizing particle traversing the liquid argon.

The p_T-resolution of the calorimeter can be parametrized for jets and electromagnetic objects as (see [12]):

$$\frac{\sigma(p_T)}{p_T} = \sqrt{\frac{N^2}{{p_T}^2} + \frac{S^2}{p_T} + C^2}\,. \tag{2.3}$$

As N,C and S are constants (see Table 2.1 for values of electrons with $|\eta| < 1.1$ and jets with $|\eta| < 0.5$ from [12]), this formula means that the resolution of the calorimeter increases, i.e. the relative error drops, if the deposited energy gets larger. This is contrary to the tracker, where the relative error of the transversal momentum increases proportional to p_T.

object	C	S [$\sqrt{\text{GeV}}$]	N [GeV]
electrons	0.044	0.224	–
jets	0.0893	0.753	5.05

Table 2.1: *Values for the energy resolution parameters.*

2.2.3 The Muon System

The upgrade of the Muon System was driven by the higher event rates and backgrounds expected in Run II, resulting in a detector $|\eta|$ coverage of ~ 2 and new efficient triggers. Like the calorimeter, the Muon System is split into two parts: the Wide Angle Muon System (WAMUS) covering the central region $|\eta| \lesssim 1$ and the Forward Muon System (FAMUS) maximizing the acceptance with $1 \lesssim |\eta| \lesssim 2$. They both share the design of three layers (A,B,C), where A is closest to the interaction region, and the toroid magnet is located between the A and B layers (see Figure 2.4).
In the central region, layer A consists of 630 new scintillation counters and the old Run I four planes of proportional drift tube chambers (PDT). With a maximum drift time of 450 ns of the PDT's, the scintillation counters enable faster triggering and background rejection. As the Muon System not only provides muon identification but also a momentum measurement independent of the central tracking, a 1 m thick toroid magnet with a field of 1.9 T is located between the A and B layers. Muons with a momentum > 3.5 GeV at $\eta = 0$ and > 5 GeV at larger η emerge from the iron toroid. Because of the magnetic field they are deflected in z so that in combination with the following layers a momentum measurement is possible, but only with a resolution of $\Delta(p_T)/p_T^2 \approx 0.004\ \text{GeV}^{-1}$ which is only half as good as the tracker resolution. Like the tracker resolution, it can be parameterized by:

$$\frac{\sigma(\frac{1}{p_T})}{\frac{1}{p_T}} = \sqrt{A^2 \cdot p_T^2 + B^2}\ . \tag{2.4}$$

Again, parameter $A = 0.00437\ \text{GeV}^{-1}$ is the resolution term and parameter $B = 0.348$ accounts for the effect of multiple scattering (see [15]).
Layer B consists of three planes of drift chambers followed by three planes of PDT's in layer C. The outer skin of the detector is made up by the C-layer scintillation counters (see Figure 2.5) achieving a timing resolution of 2.5 ns. This is necessary because of the 396ns bunch spacing when another event can produce muons. All scintillators in the central region are rectangular, with the A-layer covering $\Delta\varphi \approx 4.5°$ and $\Delta z \approx 0.85$ m.
The bottom region of the detector is not covered completely with PDT's and scintillators due to interfering support structures. This is displayed in the $\eta - \varphi$ scatter plot (Figure 2.9, $W \to \mu\nu$ Monte Carlo simulation) where one can clearly identify the uninstrumented regions of the "bottom hole" (this bottom hole has recently been covered by scintillators, but the data analyzed here did not include this newest instrumentation). In addition to this, the fine white lines in this plot illustrate the gaps between the different muon PDT chambers. The straight lines in φ show the octet structure of the Muon System best.
In the forward region the PDT's suffered radiation damage in Run I and were replaced by mini-drift proportional tubes (MDT) with a maximum electron drift time of only 60 ns and a better coordinate resolution (PDT ≈ 1 mm, MDT ≈ 0.8 mm). These chambers are supplemented by 4608 scintillation counters separated into 3 layers (A,B,C), all of them segmented into $\Delta\varphi = 4.5°$ and $\Delta\eta = 0.1$.
In this analysis the muon system is used only for muon identification, requiring combinations of layer hits, and to reject background like cosmics or proton anti-proton remnants. The tracking features of the "local" muon system, in contrast to a "central" muon track,

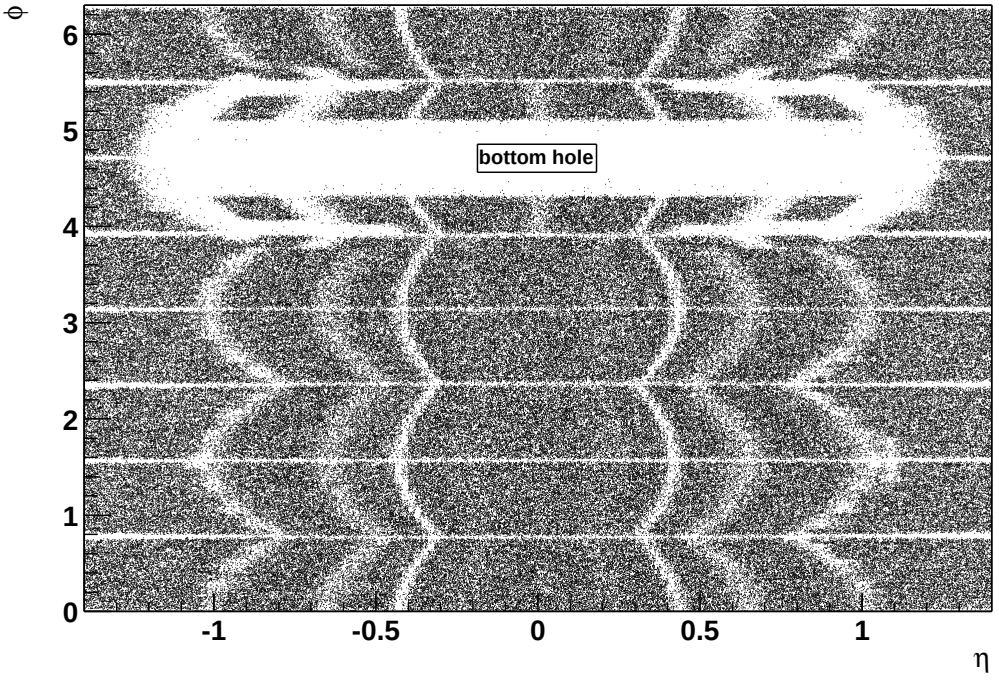

Figure 2.9: *$\eta - \varphi$ scatter plot of a $W \rightarrow \mu\nu$ Monte Carlo simulation, the uninstrumented regions of the Muon System can easily be identified.*

are employed only by matching the local track to the corresponding central track; no local p_T information is taken into consideration.
In order to further reject the background, a thick iron/lead casing surrounds the beam pipe extending from the calorimeter to the accelerator tunnel. This shielding reduces the occupancy of the detector, the probability of fake tracks and it slows down the detector aging.

2.3 Luminosity and Triggers

A high energy experiment is characterized by its center of mass energy and by its luminosity (see Section 1.3.3). To determine the instantaneous luminosity of the DØ-experiment the rate of inelastic $p\bar{p}$ collisions is measured, as:

$$l = \frac{1}{\sigma_{p\bar{p},eff}} \frac{\mathrm{d}N}{\mathrm{d}t}(p\bar{p}) \tag{2.5}$$

$$\text{with} \quad \sigma_{p\bar{p},eff} = \epsilon_l \times A \times \sigma_{p\bar{p}} \quad \text{luminosity constant [mb]} . \tag{2.6}$$

Three processes contribute to this measured rate $\frac{\mathrm{d}N}{\mathrm{d}t}(p\bar{p})$, as the inelastic cross section $\sigma_{p\bar{p}}$ can be separated into "single diffractive", "double diffractive" and "hard" scattering. In a single diffractive scattering, one of the initial particles breaks to pieces. In a double diffractive scattering proton and anti-proton fracture to hadron jets; nevertheless the interacting quarks remain unchanged and new particles are formed only during the process

of hadronization. In a hard scattering initial particles are transformed into new particles while the fragments of the (anti-)proton scatter at small angles along the beam pipe. All of these processes can be detected by two luminosity monitors, each with 24 plastic scintillator wedges mounted on the north and south calorimeter end-cap at $z = \pm 140$ cm. Since they cover the geometric acceptance of $2.7 < |\eta| < 4.4$, and are very efficient, these detectors can measure $\approx 76\%$ of the total inelastic cross section $\sigma_{p\bar{p}}$.
In order to compute the luminosity only one value is missing: The total inelastic cross section. Multiplied by $\epsilon_l \cdot A \approx 76\%$ accounting for efficiency and acceptance, this value is called *luminosity constant*. Since January 2004 a new luminosity constant has been calculated as $\sigma_{p\bar{p},eff} = 46 \pm 3$ mb. Within this constant DØ and CDF have agreed upon adopting the same value for the total inelastic cross section $\sigma_{p\bar{p}} = 60.7 \pm 2.4$ mb, which is the result of a reanalysis of the world's data (see [16]). As a consequence, the estimated error on the DØ luminosity decreases from 10% to 6.5%.
The luminosity monitors also act as a fundamental trigger for the experiment. If both monitors get hit by charged particles in coincidence, a gate is opened so that the detector records physical data. The monitors have a time-of-flight resolution of ≈ 200 ps. This enables a good distinction between real collisions and beam halo particles mimicking hard scatter processes, and the z-position of the primary vertex can be measured.

Over a time period of 60 s all luminosity related quantities are measured and a Luminosity Block Number (LBN) is formed. Each store can be divided into Runs, and each Run can be divided into LBN's, the fundamental luminosity unit. Integrating the luminosity of all LBN's in a Run provides the *delivered* luminosity of the specific Run. This number characterizes the performance of the Tevatron accelerator but it can not directly be used for physics analyses. The number important is the *recorded* luminosity. Recorded luminosity means that the collision events selected are stored on tape. The problem now is that saving all the information of one event takes a lot of time. The DØ-detector and its data acquisition are able to record data at a rate of 50 Hz. This is much less than the mean collision rate of 1.7 MHz. In order to select the events desired and to reject background, *triggers* are needed. A trigger is a tool that analyses parts of the detector data in order to decide if the event is worth of being recorded. Different particles require different triggers, as different parts of the physical data are important to make a decision. At the moment the trigger-list (version global_CMT-13.20 [17]) contains 376 different triggers, some of them specialized for electrons or jets and some for objects like di-muons and taus. All of them have in common that three steps of decision are required for the event selection: Level 1, Level 2 and Level 3.
The luminosity monitors provide an inelastic collision trigger (Level 0) and the next decision has to be made by L1, a pure hardware trigger. This trigger combines primitive information like minimum transverse energy (E_T) deposition in the calorimeter or tracks of a certain transverse momentum (p_T) in the muon chambers. The trigger therefore uses the electrical signals from the detector components and tests if a certain threshold is exceeded. Once an event passes the L1 trigger, the entire detector is read out. As the maximum L1 acceptance rate of $5 - 10$ kHz is too high for the input event bandwidth of L2 and L3, the rate has to be decreased artificially. This is done by the *prescale factor*. A Level 1 trigger with a prescale of 10 only passes on every 10th event to the next level. As a consequence, the

luminosity of an analysis strongly depends on the triggers and its prescales used for the analysis, and the recorded luminosity is smaller than the delivered luminosity. As during a store the machine luminosity declines, the L1 rates shrink also, and prescales can be redefined in order to use the full 50 Hz of data acquisition. Every time new prescales are set, a new Run begins.
After a prescale is fixed, Level 2 examines the event. This level is a hardware trigger which combines the information of different parts of the detector, for example matching calorimeter hits with tracks. Therefore this level uses processors to reorganize the detector and L1 information and it scores a rejection factor of 10 before events are passed on to Level 3.
The final decision is made by Level 3, a software trigger made up by a computer "farm" which performs a nearly complete reconstruction of the event. An event passing Level 3 of at least one of the various active triggers will be written to tape, resulting in the optimal Level 3 accept rate of 50 Hz.

Chapter 3

The Concept of Model Independent Search

3.1 Motivation

The Standard Model (SM) so far has done a great job describing the physics up to the energies available with current accelerators. Nevertheless, nobody would deny that the SM is *incomplete* in describing the nature of matter. There is great anticipation that New Physics will be detectable around the TeV-scale where the electro-weak symmetry is broken. At the moment Particle Physics is at a point where everybody is waiting for some New Physics like Supersymmetry (SUSY) to surface and experimentalists are waiting for the LHC to be completed. Theorists are under pressure as they are expected to provide solutions to all the questions SM cannot answer, such as: What is Dark Matter? What mechanism sets the values for the parameters of the SM? How can gravity be integrated? ...
There is a multitude of theories which implement extensions of the Standard Model; some of them seem to be promising, but most of them are not based on any experimental evidence, and one can easily get lost in their theoretical complexity.

The idea of Model Independent Search (MIS) is to *be based only on the Standard Model* and not to test the data for any particular extension of the SM. MIS is independent in the sense that it is not biased to look only for a specific deviation from the SM. The history of physics has shown that chance is often a scientist's best friend. The danger is that if one looks too closely for deviations, one looses the sense of the whole picture. Therefore Model Independent Search tries to cover a phase space as broadly as possible in order to cover regions where a "Conventional Analysis" fails to look.
This does not mean that MIS is the superior strategy. But the combination of analyses with dedicated selection cuts and "quasi" model independent searches appears to be the ideal tactic to enhance the understanding of the detector and pinpoint any deviations from the SM. The following comparison of approaches is meant to introduce both the tactics and to expose the differences:

Conventional analysis: The goal is to test a certain theoretical model, for example SUSY. The optimal strategy depends on the hypothetical particle or the exact properties of the model. Often the expected cross sections are low compared to SM-processes, so selection cuts are optimized in order to suppress background while keeping as much signal as possible. Therefore specific physics assumptions of the tested model have to be made in order to properly describe any new particle and its decays.
The choice of a certain model facilitates decisions regarding selections cuts, but it also biases as it depends on the correctness of the model.

Model Independent Search: The philosophy is to look at every detected event without prejudice. Selection cuts are chosen to identify physical objects (electron, muon...) and optimize the quality of the measurement. With simple distributions like $\sum p_T$, data are compared to SM-predictions in a large energy range, using as many different particle topologies as possible (e.g. final states with electrons+jets, electrons+photons, ...). In this way it is possible to detect deviations in parts of the phase space neglected by common extensions of the Standard Model.

3.2 Concept

Complete model independence is something almost impossible to accomplish. In order to quantify deviations one has to assume some kind of model. The analysis presented here also depends on a model, the Standard Model, as it compares data to Monte Carlo simulations assuming the Standard Model (MC). In addition to this, this analysis also performs some selection cuts in order to define proper physics objects.
Tevatron is a hadron-collider, so QCD reactions dominate. The cross section for $b\bar{b}$ production is $\sigma \approx 20$ µb (see Figure 1.4). Compared to this the W-production cross section of $\sigma \approx 25$ nb seems almost negligible.
The analysis presented in this paper focuses on final states with leptons as they seem most promising for New Physics. As a consequence this Model Independent Search does not include events with pure QCD interactions which create only hadron jets. In general MIS should include QCD, but it would go too far at this point as additional problems arise and the amount of data would multiply.
This analysis demands at least *one lepton per event* because a clean signature is needed. The τ-lepton is not explicitly selected; only its decay products, electrons and muons, are considered. The following physical objects are defined:

$$Electrons \mathrel{\widehat{=}} e \;||\; muons \mathrel{\widehat{=}} \mu \;||\; jets \mathrel{\widehat{=}} j \;||\; photons \mathrel{\widehat{=}} \gamma \;||\; MET \mathrel{\widehat{=}} \text{“}\nu\text{”}$$

All of them can be identified and distinguished by the DØ-detector, but only particles meeting certain selection cuts are regarded as physical objects (see Section 4.3).
The diploma thesis of Oliver Kraff [1] selects all possible combinations of electrons, muons and jets. This represents a broad scan of the physical objects recorded with the detector. Events without Missing Transverse Energy (MET) are also selected and scanned for de-

viations from SM.
The diploma thesis at hand focuses on a subset of this sample by requiring a minimum value of MET and by also taking account of photons. This refinement makes sense as almost all theories beyond SM require non-SM particles. By requireing MET, two decay scenarios of these particles are possible:

1. Non-SM particles with SM final states, possibly by violating Baryon number conservation. One of these SM final states is a neutrino as MET is required.

2. Stable non-SM particles remain and leave the detector without being detected, possible candidates for Dark Matter. This leads to a contribution to the MET.

It is important to stress that these two scenarios do not compromise the model independence of the analysis. They are just interpretations of the object MET and serve to illustrate the importance of looking at MET in particular.
For this analysis at least one muon or one electron is required to select an event, in addition to the MET. Two big data samples are chosen and analyzed separately: One with $1e + X$ called EM1TRK-skim and the other with $1\mu + X$ called 1MUloose-skim (see Section 4.1). Both samples complement one another and provide consistency checks. In order to systematically define all possible combinations of particles, exclusive and inclusive event classes are defined:

Exclusive event class: The event contains a certain combination of physical objects; no additional object passing the selection cuts is present.
Example: $1e\,0\mu\,2j\,0\gamma\,1\nu$

Inclusive event class: A certain minimum combination of physical objects is required; further objects are admitted but not considered in the computation of $\sum p_T$.
Example: $1e\,1\gamma\,1\nu + X$

Comparison with SM-Prediction

Model Independent Search attempts to minimize any bias and scan the phase space as broadly as possible. In this context it seems natural that elementary quantities are used to identify deviations from SM. Complex distributions like transversal mass or the ratio of two measured quantities often assume certain expectations of the result. For example, if one plots the transversal mass in a certain energy region, one expects to observe the peak of a decayed particle. These complex distributions should also be investigated for deviations, but as a first trial more general distributions and quantities should be chosen. In this analysis three variables sensitive to New Physics are compared to MC simulations.

- total number of events in every event class (i.e. total cross section)
- $\sum p_T$ of all particles contributing to the event class (i.e. differential cross section)
  Example: $\sum p_T = p_T(e) + p_T(\gamma)$
- MET distribution of every event class (i.e. differential cross section)

The first variable is the integral of the $\sum p_T$-distribution. It serves as an orientation to quantify the general agreement between data and MC. In the course of the analysis this comparison also helps to enhance the understanding of the detector and of the various selection cuts.
The second variable can be used to compare data to MC in a detailed way and in different energy ranges. Resonances would become apparent either in sharp peaks (e.g. a non-SM particle with a long lifetime decaying into two SM particles) or in an excess of data spread over a certain region (e.g. a resonance like $t\bar{t}$ decaying into more than two SM-particles).
As the third variable, for every event class the MET-distribution is examined and compared to MC-prediction. MET is not included into the sum of p_T as this quantity is associated with considerable uncertainty (see Section 5.3). The measurement of p_T is much better understood, so MET is separated in order to increase the accuracy of the measurement.

Pros and Cons

+ As indicated in the Motivation (Section 3.1) one of the advantages of Model Independent Search is the lack of prejudice dictated by theory. Looking at small parts of the physical phase space is also very important, but it is essential to not lose the sense of the "big picture". History has shown several times that new physics can appear in unexpected regions which the present theories do not cover. MIS does not guarantee sensitivity to all regions, but the opportunity is enhanced to detect some deviations from the SM. MIS is an alternative approach to data analysis, and as at the moment a sense of direction on where theory should go in the future is missing, every approach is worth trying.
  Besides the potential of *discoveries*, MIS provides valid information to facilitate a *broader understanding* of the detector, of the Monte Carlo event generator (e.g. simulation of jet multiplicities) and of the detector simulation (e.g. assumed resolutions). A detector not well understood, and MC not describing the data properly, will result in the downfall of any experiment and its analyses. MIS can be seen as a kind of calibration tool to tune the detector properly and detect discrepancies readily.

− The downside of MIS which looks at all events detected is that one depends on the quality of MC describing *all* data. A multitude of SM-Monte Carlos must be used to describe all combinations of physical objects. This is a significant disadvantage as some SM processes are unimportant for conventional analyses, so the official DØ-Monte Carlos lack statistics, or not much work is spent on finetuning the MC.
  The issue of QCD-background needs to be raised at this point (see also Section 6.2). At the moment where this analysis was performed, DØ did not provide QCD-Monte Carlos with sufficient statistics as their accumulation was still in the early stages. One might question why QCD is a problem for an analysis requesting at least one lepton, but QCD can also produce "fake" electrons, for example by the decay of π^0 into gammas producing an electromagnetic shower with a fake track or by the leptonic decay of a charged pion. As MIS does not only require complex particle topologies, a lot of the event classes will be "contaminated" with QCD events. Producing Monte Carlo samples of these "fake-lepton QCD events" with satisfying statistics is a techni-

cal problem as many QCD processes must be simulated and an enormous number of events must be generated to develop a proper background sample (see Section 6.2). As a result, the analysis presented here extracts the QCD-background from the data. Another disadvantage of MIS is that one does not expect a certain specified signal, so the selection cuts cannot be optimized. This restricts the sensitivity of MIS to relatively large deviations from SM. SUSY-signals with small cross sections cannot be observed. Here it is important to note that MIS is only an alternative form of analysis. Many other analyses attend to these small signals; it is not the task of MIS to cover all needs.

Related to the absence of a signal is the fact that MIS cannot set any limits. Only the level of consistency between observed data and the MC models can be quantified such that regions of significant deviation can be identified. Also, the interpretation of these deviations is not trivial, as it is difficult to relate an observed deviation to a specific extension of the SM. The idea of MIS is to identify these regions and then start to refine the analysis by sacrificing model independence. Finding definite explanations for a deviations is often a separate step and a seperate analysis.

$\pm$ Finally, one should mention that it is a chance as well as an obstacle that a Model Independent Analysis has received rather limited attention so far. There are only a few articles and papers providing information and guidelines on MIS (see [18], [19] and [20]). It is an interesting idea where the benefits are difficult to estimate, but the lack of broad experience with MIS presents many possible problems.

Chapter 4

The Data Sample and Object Identification

4.1 The Data Sample

In this analysis a data set recorded by the DØ experiment between July 2002 and April 2004 and reconstructed with the p14 version of the DØ reconstruction software "D0reco" is considered. As explained in the previous chapter, not all collisions are recorded; only events of physical interest are selected by the triggers. After the decision has been made that a certain event is worth saving, the reconstruction starts. This is done by the FNAL processor farm system, several hundred CPUs which deal with the 50Hz rate of data streaming from the detector. As a single event sizes $\approx$ 250kByte, a peak data transfer of 12.5MByte per second is accomplished, while the reconstruction of one event takes about 10 seconds [21].
The results of this reconstruction are condensed by a factor of 10 and written on tape in the Thumbnail (.tmb) format. These data contain all relevant physical information, ranging from basic values like energy or rapidity to sophisticated object identifications like trackmatch or isolation. They can be accessed any time using the SAM system (Sequential data Access via Meta-data).

In order to reduce the amount of storage, the data are divided into separate *skims*. Each skim represents a loose pre-selection of data and is chosen to meet the needs of a certain analysis. This analysis uses two skims provided by the Common Sample Group [22] as two data samples are analyzed separately (see Section 3.2). These skims are part of the "pass1 tmb-fixing", an offline re-reconstruction of all data to correct problems and imperfections. A calorimeter-noise-reduction tool called T42 is not used in this analysis as this is implemented in pass2, which is not available yet.

The skim conditions implemented have to be looser than the selection cuts applied later. As the Model Independent Search tries to cover a phase space as wide as possible, very loose skims are used:

1MUloose-skim: at least one scintillator hit and at least two wire hits are required (loose muon), $p_T > 8$ GeV, no trigger requirement

EM1TRK-skim: electromagnetic object in calorimeter (e or γ), 5 GeV track within $\Delta\varphi = \pm 0.1$ of calorimeter object, $p_T > 8$ GeV, no trigger requirement

The thumbnails of these skims are processed with the *Top Analyze*-package [23] to apply the certified object-ID criteria and the standardized corrections (e.g. smearing). The output of this package is a further reduced root-tuple which can be used by the final analysis in combination with the ROOT-framework [24].

Finally, the luminosity of the data sample has to be determined, and the data quality has to be ensured. This is done by excluding Runs graded as "bad" in the Offline Run Quality Database [25]. As this analysis deals with calorimeter, tracker and muon system objects, many different quality requirements are made: MuonQuality=bad, CalQuality=bad, CFTQuality=bad and SMTQuality=bad are excluded. This is combined with the JETMET-Bad-Luminosityblock-List [26] and the Ring-of-Fire-Bad-Luminosityblock-List [27]. Bad Runs are the result of hardware problems found during data acquisition or in offline analyses (Global Monitoring), JETMET-bad-luminosityblocks are due to calorimeter noise and the ring-of-fire is a misunderstood effect of the calorimeter observable where a ring of calorimeter cells mimics a signal. All these events are excluded from the computation of the luminosity done by the program "runrange_luminosity" [28]. This utility calculates the luminosity for a set of triggers and run ranges within a data sample and returns a bad-luminosityblock-list for the analysis. As the reconstructed luminosity depends on the triggers and its prescales, only triggers with the same L1-definition, i.e. prescales, are combined in the analysis. The luminosity of the whole data set is:

1MUloose-skim: 218 ± 14 pb^{-1} reconstructed luminosity, corresponding to 237818 good luminosityblocks

EM1TRK-skim: 229 ± 15 pb^{-1} reconstructed luminosity, corresponding to 237966 good luminosityblocks

The relative error of the luminosity of 6.5% is not a relevant systematic error in this analysis as the MC samplesare not scaled to data using luminosity and efficiencies, but, by taking the W-peak as a normalisation tool (see Section 6.1). Nevertheless the luminosity provides important information about the statistics available and also is neccessary to determine and understand cut-efficiencies.

4.2 The Set of Triggers

A Model Independent Search tries to analyze all kind of particle combinations. Therefore, triggers requiring very few restrictions are used in this analysis. A combination of inclu-

sive single-electron-triggers and inclusive single-muon-triggers is considered, in each case all using either the same L1-definition or prescale one, so that a luminosity calculation is possible. By combining similar triggers, an event which for example does not fire a specific single muon trigger even though trigger requirements are met, can still be detected by another single muon trigger. These inclusive triggers need high p_T-thresholds in order to cope with the high rates and keep prescales low. As the data sample considered reflects two years of data taking, triggers have been changed in different trigger-lists [17]. The following combinations of triggers are used in this analysis, detailed trigger definitions can be found in Appendix A.

Electron triggers for lists up to version global CMT-11.04, corresponding to runnumber < 178722, prescales ranging from $1-50$: All these triggers require electromagnetic objects at a large coverage $|\eta| \lesssim 3$, the E_T thresholds are ranging from $12-30$ GeV.
`EM_HI` or `EM_HI_SH` or `EM_HI_EMFR8` or `EM_HI_TR` or `EM_HI_SH_TR`

Electron-triggers for lists starting with version global CMT-12.10, corresponding to 178721 <runnumber< 192159: These triggers have different L1-definitions but all share the same prescale 1. The Level 3 requirements can be summarized by EM objects with $E_T > 20-30$ GeV at $|\eta| < 3.6$.
`E1_SHT20` or `E2_SHT20` or `E3_SHT20` or `E1_SH30` or `E2_SH30` or `E3_SH30`

Muon-triggers for lists up to version global CMT-10.03, corresponding to runnumber< 173102, prescales ranging from $1-361$: This trigger requires a certain combination of scintillator and wire chamber hits. At $|\eta| \lesssim 1.6$ low muon momentum thresholds are used ($p_T > 10$ GeV).
`MU_W_L2M5_TRK10`

Muon-triggers for lists starting with version global CMT-10.30, corresponding to 173352 <runnumber< 192159, prescales ranging from $1-2$: Again, a certain combination of scintillator and wire chamber hits is required, $p_T > 10$ GeV and $|\eta| \lesssim 1.6$ remain unchanged from the "old" trigger.
`MUW_W_L2M3_TRK10`

4.3 Object Identification and Selection Cuts

Model Independent Search attempts to describe various combinations of particles. Therefore, simple selection cuts must be made to ensure the overall quality of the measurement. New Physics is expected to appear in the high-p_T region as this is the field observable only with the high center of mass energies available now. This analysis focuses on high-p_T objects as describing the whole momentum range (high-p_T and low-p_T) leads to additional problems, e.g. more QCD-background or worse calorimeter resolution for small E_T. Most selected objects are required to have a relatively large minimal p_T; small η values and

additional quality cuts are made to ensure the reliability of the measured event. In this way possible deviations from the Standard Model can be filterd out.
All the following cuts were tested and studied using Monte Carlo simulations and a test data sample of only $\approx 20\,\mathrm{pb}^{-1}$. The whole infrastructure of the analysis (including the Search Algorithm) was developed with this test sample. In this way the complete data set remained untouched as a "black box" and the risk of bias is minimized.

4.3.1 Electron Candidates

In the central calorimeter (CC), electromagnetic (EM) clusters are defined as a set of towers in a cone of radius $R = \sqrt{\Delta\eta^2 + \Delta\varphi^2} = 0.4$ around an initial tower selected on the basis of its energy content. Recall that this cone includes both the electromagnetic and the hadronic calorimeter, but the energy deposit in the electromagnetic calorimeter must dominate ($> 90\%$). These EM clusters represent the reconstructed EM candidates, and further cuts decide if this cluster really contains an electromagnetic shower.
Electrons are selected combining the quality cuts recommended by the EM-ID Group [29], supplemented by a few additional cuts:

- cluster definition $|ID| = 11$ (associated track)
- electromagnetic fraction $emf > 0.9$
- electromagnetic shower shape: H-Matrix(7)< 20
- isolation $iso < 0.15$
- trackmatch
- track origins from the primary vertex: $|z_{track} - z_{vertex}| < 2$ cm
- leading electron with $p_T > 30$ GeV
- additional electron with $p_T > 15$ GeV
- central calorimeter: detector-$|\eta| < 1.1$

The cluster definition $|ID| = 11$ (electron/positron candidate) is a refinement of the simple EM-cluster definition performed by the reconstruction. The electromagnetic fraction of the total energy deposit is checked, and track information is brought in. The electron's angle information is taken from the associated track.

If an electromagnetic object traverses the detector, most energy will be deposited in the EM layers of the calorimeter. Among all reconstructed clusters, genuine EM showers are expected to have a large EM fraction:

$$emf = \frac{E_{EM}}{E_{tot}} > 0.9\,, \tag{4.1}$$

where E_{EM} corresponds to the cluster energy in the EM section of the calorimeter, and E_{tot} is the total energy in the $R = 0.4$ cone.

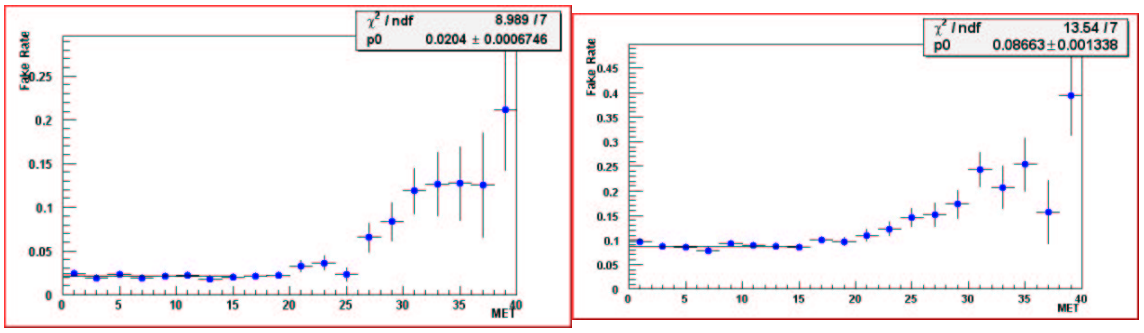

Figure 4.1: *Fake probability of track matching cut vs MET for (a) Central Calorimeter (b) Central Calorimeter with spatial requirement only (no E_T/p_T), taken from [29]*

Besides its energy deposition, electromagnetic showers can be distinguished from hadronic showers by comparing the longitudinal and lateral shower shape. Hadronic showers are much broader and travel a long way in the calorimeter until they are completely absorbed by the detector material. In order to quantify this, seven correlated observables are used for shower shape analyses (see [30]): The four EM energy fractions in each layer, the total EM energy, vertex z-position, and transverse shower width in φ. The covariance matrix is calculated for each tower in η using Monte Carlo electrons. With H, the inverse of this covariance matrix, a χ^2 is determined which is a measure of how similar the shower is to an electron shower.

This analysis focuses on isolated leptons in order to suppress QCD-background. Electrons and muons are also the decay products of, for example, B-mesons in a hadron jet, or jets can be misidentified as electrons. Leptons from QCD-events tend to be near the energy deposit of the hadron jet, so the "cluster isolation" is an important factor:

$$f_{iso} = \frac{E_{tot}(R < 0.4) - E_{EM}(R < 0.2)}{E_{EM}(R < 0.2)} < 0.15 \; . \tag{4.2}$$

Here a second conus of $R = 0.2$ combining clusters is used. This means that only a small fraction of additional energy is allowed outside the $R = 0.2$-cone of the electromagnetic calorimeter possessing most of the energy from the primary EM shower. A nearby hadron jet would lead to a much greater energy deposit in the larger cone, and the EM candidate would not pass the isolation cut.

For electrons and positrons, an associated track candidate is required in order to separate them from photons or QCD contamination. In the central region the trackmatch is defined using the χ^2 [31]:

$$\chi^2 = \left(\frac{\delta\varphi}{\sigma_\varphi}\right)^2 + \left(\frac{\delta z}{\sigma_z}\right)^2 + \left(\frac{\frac{E_T}{p_T} - 1}{\sigma_{E_T/p_T}}\right)^2 \; . \tag{4.3}$$

In the above expression, $\delta\varphi$ and δz are the differences between the extrapolated track position and the EM cluster position at the third layer of the calorimeter (finest segmen-

tation). These differences illustrate the spatial matching between the track candidate and the EM cluster, as the z-coordinate is equivalent to an η-information. E_T/p_T is the transverse energy of the EM cluster as measured by the calorimeter divided by the transverse momentum of the track. The σ variables are the "root-mean-squares" of the experimental distributions of each quantity. A trackmatch is found true if the probability of the χ^2 exceeds 0.01.
The E_T/p_T quantity is a very powerful discriminator as only well measured tracks will meet this criterion. On the one hand, the tracker resolution declines as p_T rises, so objects with a large transverse momentum (high-p_T) are measured more precisely by the calorimeter (see Section 2.2). For high-p_T objects this results in a drop in efficiency for the trackmatch cut. On the other hand, the probability of a false trackmatch decreases using E_T/p_T from 8% to only 2% (see Figure 4.1). Fake trackmatches are often jets misidentified as EM candidates with an associated track. In order to increase the purity of the sample, a trackmatch including E_T/p_T is used in this analysis.

As Model Independent Search investigates different combinations of particles, it is important to ensure that all these particles result from the same primary vertex. With the current luminosity delivered by the Tevatron, each bunch crossing can produces multiple interactions. Most of them are QCD events because the cross sections are much larger than for electroweak boson production or similar processes. Nevertheless it is possible, even though very improbable, that in a single event two W-bosons are produced with two different vertices mimicking, for example, a $1e\,1\mu$ event. More likely is the combination of a true W-boson decaying into electron and neutrino and a "fake" muon arising from a second QCD hard scatter.
In order to anticipate similar scenarios, an assessment of the difference between the z-coordinate of the electron candidate and the z-coordinate of the vertex found by the 2-pass Primary Vertex finder [32] is made. This vertex finder determines the most probable primary vertex by combining track information with a probabilistic method that assigns a probability that the reconstructed vertex comes from a "minimum bias" interaction (additional scatter processes next to hard interaction). Figure 4.2 illustrates the distribution of this difference obtained from a $W \rightarrow e\nu$ Monte Carlo simulation. MC also simulates multiple interactions by underlying minimum-bias events. Most of the electron candidates originate from the primary vertex. The electrons outside the 2 cm range result from multiple hard interactions or mis-reconstructed vertices. If the vertex reconstruction algorithm fails, the primary vertex is set to the nominal interaction point at $z = 0$ cm. This results in a difference in z if the true origin of the electron is somewhere else.
It is important to understand that this z-coordinate cut is not comparable to a dca cut (distance-of-closest-approach) as 2 cm is a very loose requirement. The dca an be illustrated using a sphere with the primary vertex as its origin. If the track is a tangent to this sphere, then the radius of the sphere is the dca. Typical dca-cuts range in the < 0.01 cm region to reject background from secondary vertex decays. As this Model Independent Search is designed to remain sensitive to possible secondary decays of heavy non-SM particles, no dca cut is applied.

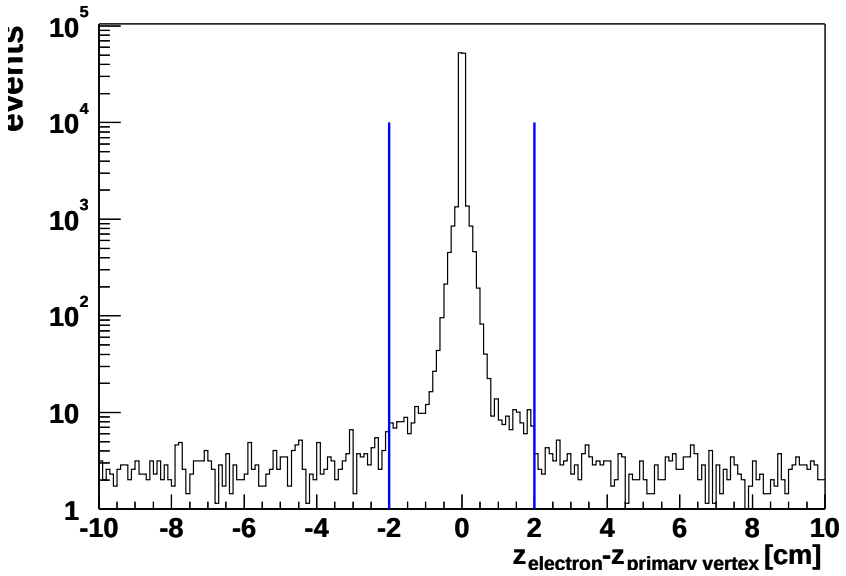

Figure 4.2: *z-distance between primary vertex and leading electron of simulated $W \rightarrow e\nu$ MC.*

The last cuts on transverse momentum and pseudorapidity define the kinematic and geometrical acceptances. The relatively high threshold of $p_T > 30$ GeV for the leading electron is primarily due to trigger requirements. In this region the single electron triggers used are fully efficient [33]. Additional electrons or electrons triggered by a muon-trigger (electrons in the 1MUloose-skim) only need $p_T > 15$ GeV, a value larger than most calorimeter noise. The geometrical acceptance is defined by detector-$|\eta| < 1.1$. Detector-η is computed with reference to the detector point of origin at $z = 0$, in contrast to physical-η which refers to the primary vertex found. EM objects are reconstructed using the calorimeter with its projective towers pointing to the detector point of origin.
Following the official EMid [29], the calorimeter is divided into Central Calorimeter (CC with det$-|\eta| < 1.1$) and Endcap Calorimeter (EC with $1.5 <$det$-|\eta| < 2.5$). These components have different efficiencies for reconstructing electrons and different resolutions. In order to create a pure sample of comparable and similar electromagnetic objects, only the CC region is regarded in this analysis.
The EM calorimeter has cracks in φ arising from the 32 separate modules. EM candidates depositing significant energy inside these cracks, called "non-fiducial" electrons, are measured insufficiently. As MCsimulates these electrons properly, the entire calorimeter is taken advantage of to enhance statistics.

4.3.2 Photon Candidates

Photons are EM candidates without an associated track. Therefore all electron cuts except the trackmatch are performed:

- veto on trackmatch
- photon with $p_T > 15$ GeV

As Photons are not used to fire any trigger, only the $p_T > 15$ GeV threshold is required. The isolation requirement is maintained to reduce QCD-contamination: π^0 are major components of a hadron jet; their decay into a pair of gammas produces EM-candidates, most of them unisolated. This analysis focuses on isolated photons from non-SM particle decays like $X' \to \gamma + X$. With the help of isolation they can be separate properly from the QCD-background and from low-p_T initial or final state radiation (see Section 5.1).

4.3.3 Muon Candidates

Muons are reconstructed using information from two independent detector systems: the muon detector with its scintillators and drift tubes, and the central tracker. A "local" track in the Muon System is the basis of muon identification; an associated "central" track provides precise p_T measurement. The analysis presented here considers only high quality muons with a local track and a correlated central track in order to combine the superior track resolution of the tracker with the clear object identification of the Muon System (only muons traverse the calorimeter). The potentially problematic background of mis-reconstructed muons and cosmics is therefore minimized. Muons are selected combining the certified muon object definitions by the Muon-ID Group [34] with additional quality cuts, special terms will be explained further below:

- $nseg = 3$ muon with "medium" quality
- standard timing cut against cosmics
- central track with $N_{SMT} \geq 3$
- $\chi^2_{track}/dof < 3.5$ for central track fit
- $\chi^2_{global} < 160$ for global fit between track and muon system
- track origins from assumed primary vertex: $|z_{track} - z_{vertex}| < 2$ cm
- isolation criteria
- muon with $p_T > 15$ GeV
- detector-$|\eta| < 1.4$

A $nseg = 3$ muon is required to have a central track matched to muon segments. The local track must include hits before and after the toroid (A and BC layer). These hits are fitted to form a local track. The local track is probably better designated a "track candidate" as a specific quality of the track fit is not required in $nseg = 3$. This track is then extrapolated to the tracker to match with a central track. If the fit of the local track does not converge, the matching is done starting from the tracker and extrapolating to the muon system. One of these overall fits must converge, but the local track itself can be poorly reconstructed. In principle both tracks can now be combined into a "global track" to give the 4-vector of the muon, its charge, η, φ and the z-coordinate. Unfortunately this global track is not well understood, and kinematic variables measured only by the central track are more reliable. As a consequence, the local muon system is used only to provide object identification by

matching the central track to the hits in the muon system.

The quality "medium" in combination with a certain *nseg*-value determines the number of wire and scintillator hits: At least two A layer wire hits, at least one A layer scintillator hit, at least two BC layer wire hits, and at least one BC scintillator hit. With hits before and after the toroid a local track can be defined more easily, and the requirement of numerous wire and scintillator hits minimizes noise effects.

The standard "loose" cut against cosmics is applied based on timing information from scintillator hits. To reject cosmic ray muons, the transit time in scintillator layers A, t_A, and in scintillator layers BC, t_{BC}, must be consistent with the time-of-flight for muons originating from a beam collision:
$|t_A| < 10\text{ns}$ and $-15\text{ns} < t_{BC} < 10\text{ns}$
As a C-layer scintillation counter achieves a timing resolution of 2.5 ns, this cut ensures that the muons studied emerge from the hard scatter process and are not of cosmic origin. In the final sample (i.e. all cuts applied) of selected muons 100% of the muons fulfill this timing cut.

Muons originating from cosmic ray showers are an irreducible background for every analysis dealing with muons. Cosmic muons tend to be well measured in the Muon System and the tracker, so even after several cuts on the quality of the p_T measurement, and after timing cuts, some cosmic muons present at the time of the event can not be anticipated. Especially in the high-p_T region where muons from the vertex are mostly due to a bad track fit, the cosmic contamination becomes one of the major backgrounds. Next to New Physics and bad tracks, cosmic rays are the most probable source of muons with $p_T > 200$ GeV, which pass all selection criteria. A track which obviously does not originate from the vertex will not pass the track quality criteria discussed below. Therefore, cuts requiring a well measured track implicitly also perform a very loose cut on $dca \lesssim 1$ cm (distance-of-closest-approach) . As a consequence no additional anti-cosmic dca-cut is made to remain sensitive to possible secondary vertex decays of new particles.

As for the electron candidate, a cut on the difference between z-coordinate of the muon candidate and z-coordinate of the primary vertex is made. In this way all leptons belong to the same vertex, and badly reconstructed tracks or vertices are discarded.

As discussed earlier, the p_T-measurement is performed by the tracker. In order to ensure a precise measurement of the central track, the Silicon Vertex Detector is mandatory. The relative error of p_T increases with p_T, and tracks without SMT-hits have even worse resolution (see Figure 2.7). Poorly measured tracks have a much larger probability of producing a very high p_T measurement (> 200 GeV). By requiring at least three SMT-hits, a three-dimensional track reconstruction is possible. The resulting sagitta (p_T) measurement using the Silicon Vertex Detector has a better resolution, and the non-physical high-p_T-background is reduced. 100% of muons reconstructed with the SMT cut have more than

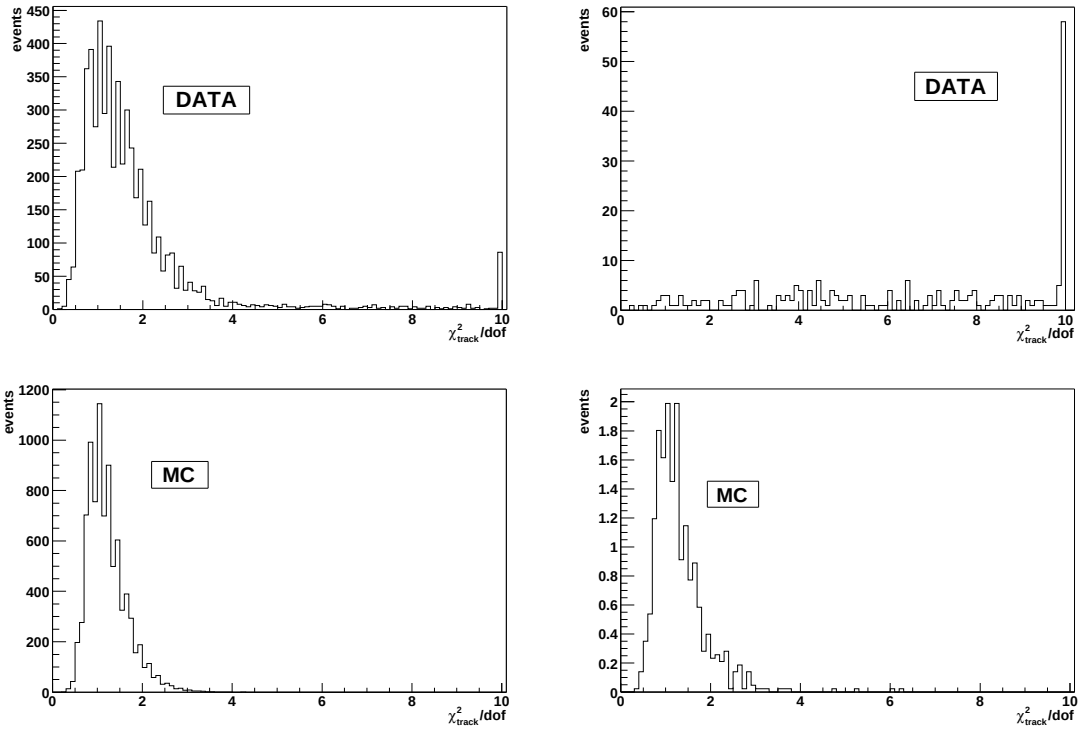

Figure 4.3: *These plots show the χ^2_{track} distribution of muons with all cuts applied except the χ^2_{track} cut. top: Data with $p_T(\mu) < 100$ GeV(left) and $p_T(\mu) > 100$ GeV(right) bottom: MC with $p_T(\mu) < 100$ GeV(left) and $p_T(\mu) > 100$ GeV(right)*

eight additional CFT hits, further improving the measurement, so no explicit CFT cut is performed.

The $\chi^2_{track}/dof < 3.5$ cut removes bad track fits which mimic isolated high-p_T muons. Some of these fakes are produced by in-flight decays of pions or kaons, resulting in a poorly reconstructed high-p_T track caused by a kink in the track at the point of decay and a low-p_T local muon (measured in the local muon system). For more information see [35]. This cut is very effective against this non-physical high-p_T tail, and the remaining high-p_T muons are expected to be genuine high-p_T muons. Figure 4.3 illustrates this, where a small data sample with all cuts applied except the χ^2_{track} cut is shown. The fraction of poor quality central tracks is enhanced for muons with $p_T > 100$ GeV due to poor track fits. The assumption of bad fits is supported by the fact that Monte Carlo events do not show this accumulation of large χ^2_{track}/dof values in the high-p_T tail.

The cut on the global track-χ^2 tightens the $nseg = 3$ requirement as it demands a certain quality of matching between the central track and hits in the muon system. If a track cannot be assigned to a local track with certainty, the quality of the muon measurement is called into question. Again, most unphysical high-p_T candidates are rejected by this cut, and the combination of SMT and both χ^2 cuts serves to distinguish between genuine high-p_T muons and ghost tracks.

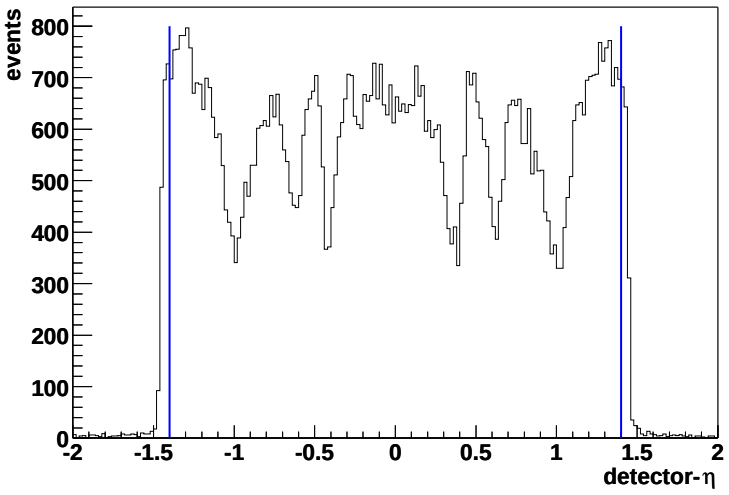

Figure 4.4: *Muon detector-η of data, trigger acceptance is reflected by sudden drop at $|\eta| \approx 1.5$*

This analysis emphasizes on isolated leptons to suppress QCD-background. Heavy quark jets produce non-isolated muons from semi-leptonic decays, which tend to have a low transverse momentum. The isolation criteria is defined by [12]:

- HaloCone$(0.1, 0.4) < 0.08\, p_T$, where HaloCone$(0.1, 0.4)$ is the sum of the E_T of calorimeter clusters in a hollow cone between $\Delta R = 0.1$ and $\Delta R = 0.4$ away from the muon. Cells in the electromagnetic and fine hadronic calorimeter are considered for this calculation, cells from the coarse hadronic calorimeter are excluded to reduce noise effects.
- TrkCone$(0.5) < 0.06\, p_T$, where TrkCone(0.5) is the sum of the p_T of all tracks within a cone of Radius $\Delta R = 0.5$ surrounding the muon. A possible track matched to the muon is excluded from this calculation.

In this way a measure of unclustered calorimeter energy surrounding the muon is combined with the information from the tracking detector to further isolate the muon candidate.

The p_T-requirement for muons is lower than that for electrons, as muons are identified more easily and cleanly within the muon system without any intruding calorimeter noise. The single muon triggers used are fully efficient when $p_T > 15$ GeV is used as a threshold [36]. In this way, a lower p_T-region of muons can be included in the analysis. One must recognize, however, the disadvantage of poorer high-p_T measurement as compared to electrons.
The geometric acceptance defined by detector-$|\eta| < 1.4$ is selected with respect to the trigger acceptance. All muon triggers used are wide-region triggers which are designed for and fully efficient in the central detector region. Outside this region, the efficiency drops, see Figure 4.4. Several other sinkings can be seen which are caused by the imperfect coverage of the Muon System (see also Figure 2.9). As Monte Carlo does not include a trigger simulation yet, this cut ensures that the trigger is efficient, and data can be compared directly to MC (assuming a constant trigger-efficiency which is flat in η).

4.3.4 Jet Candidates

If only the calorimeter information is utilized, jets at DØ are reconstructed using the *improved legacy cone* algorithm [37]. Here, only a bief description of this complex mechanism is made: A cone algorithm in general forms jets by associating towers whose centers lie within a circle of specific radius $R = \sqrt{\eta^2 + \varphi^2}$ in the $\eta \times \varphi$ space. This cone is the result of an iterative process beginning with seed towers which represent the sum of all cells sharing the same pseudo-rapidity and azimuthal angle. If this sum produces a positive energy (no noise), all 4-vector variables pointing from the reconstructed vertex within a cone around the seed tower are totaled and a new energy-weighted center is determined. This new point in $\eta \times \varphi$ is the center for a new trial cone, and the whole process is repeated until a stable solution is found where the balance point of energy deposition is aligned with the geometric axis of the cone.
For this analysis, a cone of $R = 0.5$ is chosen, and further quality selection cuts following the certified JetID [38] are applied to each clustered jet to remove false jets:

- electromagnetic fraction $0.05 < emf < 0.95$
- $\Delta R(EMobject, jet) > 0.5$
- coarse hadronic fraction $chf < 0.4$
- remove hot cells by $HotF < 10.$
- remove hot towers by $n90 > 1.$
- $L1SET$ anti-noise-cut
- jets with $E_T > 30$ GeV
- detector-$|\eta| < 1.5$

If a hadronic jet traverses the detector, the first location of massive instrumented material is the electromagnetic calorimeter. Like EM objects, jets start to shower in these cells of the calorimeter, so considerable energy will be deposited already here. In particular, jets with a large fraction of π^0's will result in a large EM fraction, as many photons will be produced.
To remove isolated EM particles, an upper limit of $emf < 0.95$ on the electromagnetic fraction of jets is defined (EM candidates are required to have $emf > 0.9$). As electrons and photons possess several other properties, e.g. characteristic shower shapes, jets can be distinguished from electrons or photons. In order to further minimize the risk of confusion, jet candidates that are close ($\Delta R < 0.5$) to electrons or photons with $p_T > 15$ GeV are discarded.

To remove jet candidates which deposit their energy primarily in the coarse hadronic section of the calorimeter, an upper limit of 0.4 on the fraction $chf = \frac{E_{ch}}{E_{tot}}$ is applied. In this way, jet candidates which cluster around noise in the coarse hadronic section are removed.

$HotF < 10$ removes false jets caused by hot calorimeter cells. A hot cell is an isolated object faking a high E_T signal that can be combined with surrounding general noise to form a jet candidate. The ratio of the highest to the next-to-highest transverse energy cell must be lower than 10 to prevent this effect.

Another anti-noise cut removes jets clustered from a single hot tower by requiring that 90% of the jet energy $n90$ is deposited in more than just one calorimeter tower. As towers have typically sizes of $\eta \times \varphi = 0.1 \times 0.1$, and since the jet cone is much larger, genuine jets deposit their energy in a combination of neighbouring towers.

The final noise reduction cut utilizes the energy measurement of the L1 calorimeter trigger. The comparison of the energy in the Level 1 compared to the precision readout is a good discriminant against noise problems caused by the readout electronics. $L1SET$ is defined as the scalar sum of the L1 trigger towers E_T inside the same cone of the tested jet, and in this context the detector can be divided into the regions CC with $|\eta| < 0.8$, EC with $|\eta| > 1.5$ and the ICD region with $0.8 < |\eta| < 1.5$. Jets must fulfill:

$$\frac{L1SET}{E_T(jet) \times (1 - chf)} \begin{cases} > 0.4 & \text{in CC, EC} \\ > 0.2 & \text{in ICD} \end{cases} . \tag{4.4}$$

As this analysis focuses on events with at least one isolated lepton, most jets contributed by Standard Model processes originate from the gluon radiation of the inital partons (see also Section 5.1 for definition of initial and final state radiation), or from heavy particle production with hadronic decay channels (e.g. $t\bar{t}$ production). As no jet trigger is inquired, a specific E_T threshold is arbitrary. In order to separate genuine jets from noise imitators, a relatively low E_T-requirement seems impractical. In addition to this, the Jet Energy Scale (see below) works only for jets with $E_T \gtrsim 10$ GeV. As a consequence, this analysis focuses on jets with a minimum transversal energy of 30 GeV.

The DØ-calorimeter has an excellent full pseudorapidity coverage of $|\eta| \lesssim 4$. A cut detector-$|\eta| < 1.5$ comparable to the selection criteria of muons and EM objects is chosen as the error of the Jet Energy Scale rises in the very forward region of the calorimeter (see [39]). Figure 4.5 shows the η-distribution of jets in a $W \to \mu\nu$ Monte Carlo events. These are typical examples of jets produced by gluon emission of the initial partons.

Jet Energy Scale (JES)

A detector measures a signal of a certain height if a particle transverses. This signal (ADC values) must be transformed into an energy value. If a jet showers in the calorimeter, only a certain amount of the total energy can be measured because inert dead absorbing material exists. The calorimeter, therefore, must be calibrated, which means that a relation between electronic signals from the calorimeter cells and energies at the particle level (e.g.

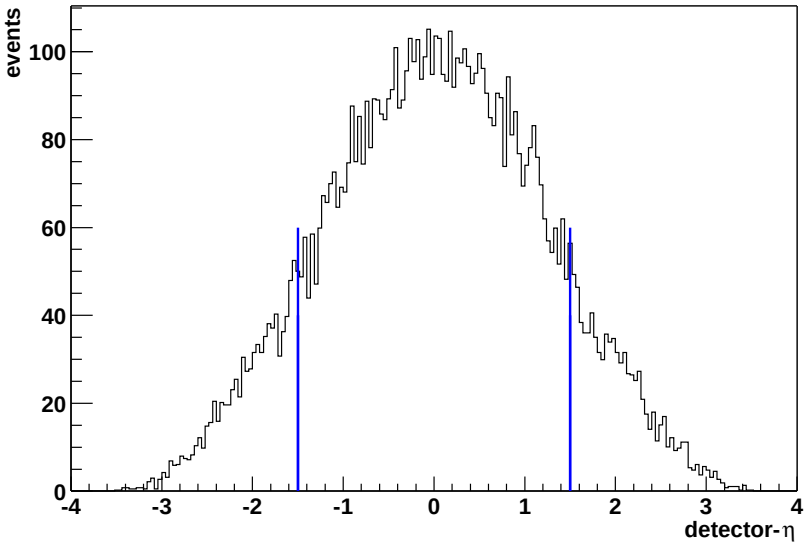

Figure 4.5: *Jet-η distribution of $W \rightarrow \mu\nu$ MC*

the "real" momentum value of a final state quark) must be determined.
This calibration is done in two steps:

- The electromagnetic calorimeter (EM scale) is calibrated using $Z \rightarrow ee$ or $J/\Psi \rightarrow ee$ events.
- For hadronic jets, transverse momentum conservation of $\gamma + jet$ events is used to determine the Jet Energy Scale as the EM scale of the photon has been measured in the first step.

In this way the measured hardware signal is translated into the GeV-scale by multiplying with a factor, the *Jet Energy Scale*. This energy value then reflects the true physical value of the jet.
The mechanism described above only outlines the principle of the Jet Energy Scale. In reality, the course of action is a bit different: For each calorimeter cell the hardware signal is translated into a GeV value, e.g. by using test beam measurements of single calorimeter cells. Then these cells can be combined to form a proper jet with the energy E_{cells} using the cone algorithm. By applying the JES to this "raw" energy, again a jet energy at the particle level can be determined. The energy value of single cells will be important in the determination of Missing Transverse Energy, described below.
In addition to this, the Jet Energy Scale is dependent on the energy of the jet candidate and on its η value, see [39] for corresponding figures. Unfortunately, the Jet Energy Scale is also one of the major sources of uncertainty which will be discussed in detail in Section 8.4.

4.3.5 Missing Transverse Energy (MET)

Missing Transverse Energy is the signature of neutrinos and other non-interacting particles in the detector computed from the momentum imbalance of an event in the transverse plane (see Section 1.3.2). The calculation of MET combines quantities from central tracking (for primary vertex determination), calorimeter (for energy deposition) and muon detector (for the muon correction to the energy of the event). MET has to be corrected for the reconstructed objects of an event and is thus strongly dependent on this reconstruction. The missing transverse energy is the last object computed, and it uses inputs from other reconstruction algorithms, such as the Jet Energy Scale and the electromagnetic scale. Uncertainties in all these values propagate, and MET reflects this. MET is calculated in several steps [40]:

First the visible energy in the calorimeter is summed up:

$$E_{x,y}^{vis} = \sum_{cells} E_i^{x,y} \ . \tag{4.5}$$

To establish transverse momentum balance in the calorimeter, missing energy is $MET_x = -E_x^{vis}$ and $MET_y = -E_y^{vis}$ resulting in $MET = \sqrt{(MET_x)^2 + (MET_y)^2}$. It is important to stress that the calorimeter granularity is taken advantage of by using cells and primary vertex information for MET reconstruction. In this way η and φ can be determined for each cell individually, and the transverse component of the measured energy can be calculated for each single cell. As a consequence, the sum over all cells and thus $E_{x,y}^{vis}$ is more precise. In this sum, cells from the coarse hadronic (CH) calorimeter ($>$ layer 15) are not included because of the noise present. To account for jets which deposit energy in this part of the calorimeter, the clustered jets are corrected for this energy by using the coarse hadronic fraction $chf = \frac{E_{ch}}{E_{tot}}$. Then the "raw" transverse energy of the jet is subtracted from MET_x and MET_y.
In addition to this correction of "raw" deposited jet energy in the calorimeter, MET has to be corrected for the Jet Energy Scale. Only after the JES is applied to all clustered calorimeter jets, their transverse energies correspond to jet energies at the particle level and momentum balance can be applied. The same is true of the electromagnetic scale.
Finally, MET is corrected for the presence of reconstructed muons in the event, as this minimum ionizing particle would alter the MET. The momentum of muons passing a set of quality cuts ("tight" muons) is subtracted from the Missing Transverse Energy after deduction of the expected muon energy loss in the calorimeter. The final MET is given as

$$MET_{cal+muon} = \sqrt{(MET_x^{cal} - p_x^{muons})^2 + (MET_y^{cal} - p_y^{muons})^2} \ . \tag{4.6}$$

This analysis selects events with a certain amount of MET. This threshold has to be relatively large to identify MET from particle interactions, as there is a constant MET value of ≈ 10 GeV due to noise and MET-resolution:

- $MET > 30$ GeV

Chapter 5

Monte Carlo Samples

5.1 MC Generator

The PYTHIA Program

This analysis uses only Monte Carlo samples with SM-processes. All except the $t\bar{t}$-sample are generated with PYTHIA (versions: 6.2 [41]), a *leading-order parton shower generator*. What this means will be discussed in the following.
The PYTHIA program is frequently used for event generation in high-energy physics, simulating the hard interactions in e^+e^-, pp and ep colliders. Especially at a hadron collider at 2 TeV the emphasis is on multiple particle production. The best way to describe the complexity of the interaction, its higher order corrections and the hadronisation, is to factorize the problem in several components:

Initially two beam particles characterized by their *parton density functions* (pdf) approach each other head on. Only two partons enter the *hard process*, which makes up the core of the reaction. A typical example would be a $2 \rightarrow 2$ process with two initial partons and two final states. Leading order matrix element calculations are used to compute the exact kinematics. A good example would be the W-production displayed in Figure 5.1. This is the $2 \rightarrow 1$ production of a resonance followed by the leptonic decay of the gauge boson. As hadrons have a substructure of quarks and gluons, one depends on the parameterizations of pdf's which are based on experimental data (see Hera [4]). Uncertainty related to the choice of a certain pdf will appear again in Section 8.2, as it results in a cross section error.

The second component simulated by PYTHIA is modeling perturbative corrections. The problem now is that loop corrections and other higher order effects become increasingly difficult to compute in higher orders. Often one is interested in the inclusive production of particles, so the correct description of *initial- and final state radiation* is crucial. As the couplings of the strong interaction α_s and of the electromagnetic interaction α_{em} have different magnitudes with $\alpha_s \gg \alpha_{em}$, the gluon radiation dominates. Examples for these radiative corrections could be the initial radiation of a gluon by one of the incoming quarks,

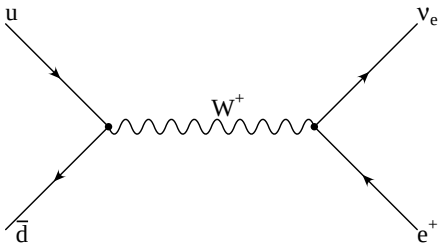

Figure 5.1: *Feynman diagram of W-production and semileptonic decay of the gauge boson.*

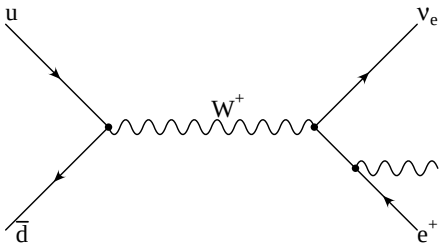

Figure 5.2: *Feynman diagram of $W \rightarrow e\nu$ with final state radiation.*

or the final state photon radiation of the electron as shown in Figure 5.2. PYTHIA solves this with the help of *parton showers*. Here branchings of one parton into several particles are described by approximating the full matrix-elements. Parton showers have the disadvantage of poorly describing multi-jet-topologies, but for the analysis of this thesis they serve as a good approximation.

Up to this point only elementary particles like quarks or gauge bosons have been considered. As QCD becomes strongly interacting at long distances, all colored partons must be transformed into colorless hadrons. This is the third component, called *hadronization*. To describe this, PYTHIA uses the model of String Fragmentation. It can be illustrated by the example of a $q\overline{q}$ final state pulling away from each other. These quarks are connected by a color flux tube which is stretched and which conserves the colorlessness. As the string stretches further, its potential energy increases until it breaks up to a new $q'\overline{q}'$ pair. This extremely simplified picture (see Figure 5.3) illustrates the "quark confinement", i.e. the fact that single isolated partons cannot be observed.
Many of the mesons and baryons produced by fragmentation are unstable and decay into the observable particles. The final component is the simulation of the branching ratios and decay modes by PYTHIA so that in the end complex objects like jets are described properly.
Besides the hard interaction, further processes must be considered to simulate real data. The partons taking part in the hard scatter leave behind *beam remnants*. Initially the proton consists of three quarks. After a hard scatter process where one quark from the proton takes part in, the remaining two quarks are color-connected to the hard interaction, which controls their fragmentation to a hadron jet. These beam remnants are observed as jets in the very forward and backward regions of the detector.
The other issue relates to *multiple interactions* in a single event arising from the multitude

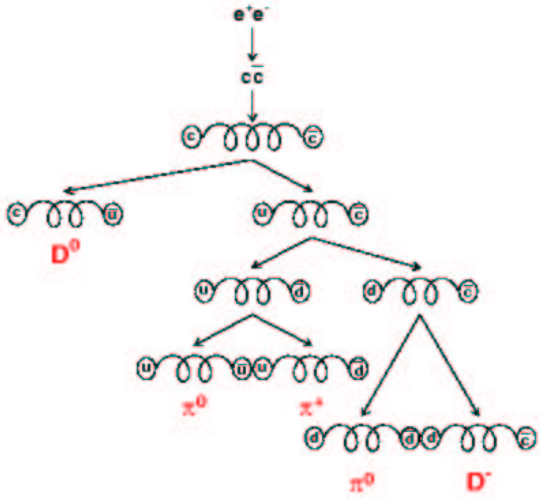

Figure 5.3: *Scheme for hadronization into a jet [42]*

of protons in the beam. In addition to the hard process of two partons, another proton anti-proton pair of the bunch can interact, e.g. with a single-diffractive scattering. In addition to this, multiple *parton* interactions have to be implemented. The spectator quarks of a hard intraction can scatter against several different partons from the other beam, so the clean signiture of the hard process will be covered with several other tracks and detected particles. Multiple interactions are simulated by underlying "minimum-bias" events in the hard scatter process. These events are data randomly taken by the detector. Due to the large cross sections most of these events are QCD $2 \rightarrow 2$ scatterings with small p_T.

The Detector Simulation

The objects discussed so far are "particle" jets and generator leptons. Of course this cannot be compared to the actual data as these contain only "detector" objects. The final step is simulating the complex interactions of the various particles with the detector material. For example, after hadronization these particle jets arrive at the calorimeter and form large hadronic showers. It is obvious that the realistic simulation of the exact structural plan of the DØ-detector and its magnetic fields is crucial for reproducing actual data. Therefore all MC used in this analysis undergo a *full simulation* by using the programs D0gstar and D0Sim [43]. The former is a casing for GEANT and determines the amount of energy deposited in the active regions of the detector and describes the particles' passage through matter. Figure 5.4 illustrates the complexity of this passage through the detector, a simulation done by D0gstar.

The latter (D0Sim) simulates the electronic signals, merges hard scatter and minimum bias events and adds noise effects of the subsystems of the detector.
This complex simulation process produces a file in a data format compatible with the standard reconstruction software D0reco, which real data also have to undergo. The only mechanisms not simulated are the triggers used to record real data. The MC samples have to be corrected for the efficiencies of the triggers.

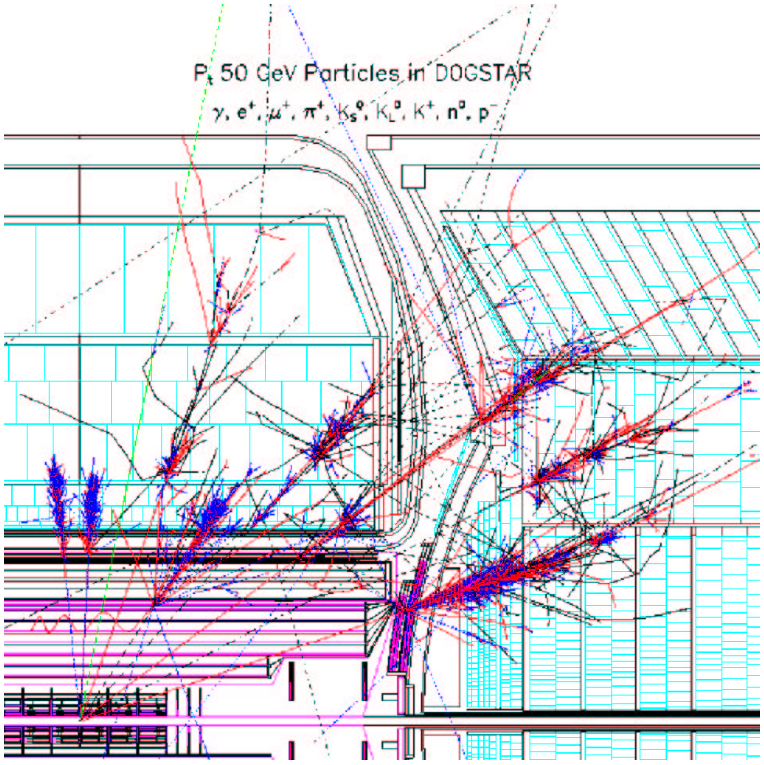

Figure 5.4: *Example for the complexity of the simulation of the passage through the detector [44].*

5.2 SM-Processes

Most MC-samples used here are "official" DØ-Monte Carlos, suggested and assembled by the New Phenomena Group (see current MC requests/status at [45]). The $t\bar{t}$ samples were produced by the Top Group and are generated by an ALPGEN/PYTHIA interface. ALPGEN does not use parton showers but, rather, exact matrix elements. This event generator calculates and produces n-parton final states, which leads to a better description of the topology. In a second step PYTHIA is used to simulate the hadronization. The diboson samples WW,WZ and ZZ are special MC samples [46]. All MC sets are inclusive in the sense that initial and final state radiation is activated, which leads to additional jets and photons. In the following, "inclusive" will be marked with "X", and "l" means any charged lepton.
As PYTHIA is only a leading order generator, the cross sections have to be corrected for next-to-leading or next-to-next-to-leading order (NNLO) effects by applying K-factors (if already available). The cross section also depends on the choice of the pdf, namely a LO- or a NLO-pdf. All DØ-Monte Carlos use CTEQ4/5L pdf's [47], which are leading order. As the determination of a NNLO pdf has not been completed yet (see [48]), the K-factor is defined as:

$$K_{NNLO} = \sigma(\text{NNLO ME, NLO pdf}) \quad / \quad \sigma(\text{LO ME, LO pdf}) \tag{5.1}$$
$$\text{with ME} = \text{Matrix Element}$$

SM-process (M_{Z/γ^*} [GeV])	σ_{MC}×BR [pb]	$\mathcal{L}_{MC}$ [pb^{-1}]	K_{NNLO}
$Z/\gamma^* \rightarrow \mu\mu$+X (5-15)	4579.7 (NNLO)	47.9	1.29
$Z/\gamma^* \rightarrow \mu\mu$+X (15-60)	450.4 (NNLO)	222.0	1.35
$Z/\gamma^* \rightarrow \mu\mu$+X (60-130)	252.7 (NNLO)	1024.9	1.39
$Z/\gamma^* \rightarrow$ ee+X (5-15)	4579.7 (NNLO)	47.0	1.29
$Z/\gamma^* \rightarrow$ ee+X (15-60)	450.4 (NNLO)	222.0	1.35
$Z/\gamma^* \rightarrow$ ee+X (60-130)	252.7 (NNLO)	977.4	1.39
$Z/\gamma^* \rightarrow \tau\tau$+X (5-15)	4579.7 (NNLO)	65.8	1.29
$Z/\gamma^* \rightarrow \tau\tau$+X (15-60)	450.4 (NNLO)	485.8	1.35
$Z/\gamma^* \rightarrow \tau\tau$+X (60-130)	252.7 (NNLO)	997.2	1.39
W $\rightarrow \mu\nu$+X	2676.9 (NNLO)	824.2	1.39
W $\rightarrow$ eν+X	2676.9 (NNLO)	832.8	1.39
W $\rightarrow \tau\nu$+X	2676.9 (NNLO)	648.4	1.39
WW $\rightarrow$ X	13.0 (NLO)	1076.9	-
ZZ $\rightarrow$ X	1.56 (NLO)	9615.4	-
WZ $\rightarrow$ X	3.96 (NLO)	4166.7	-
$t\bar{t} \rightarrow$ lν+4jets+X	2.97 (LO)	15404.0	-
$t\bar{t} \rightarrow$ 2lν+2jets+X	0.70 (LO)	13571.4	-

Table 5.1: *MC samples*

In Table 5.1 all SM-processes, as well as the NNLO cross section σ_{MC}×Branching Ratio used in the analysis, the equivalent MC-luminosity and the applied K-factor are listed. The values for the K-factors can be found in [49] for W,Z. For all other MC samples the theortical cross section was used, so no K-factor is listed in the Table. For WW, ZZ and WZ, calculations in next-to-leading-order in α_S are used taken from [50]. The leading-order $t\bar{t}$ production cross section of $\sigma = 6.82$ pb is also taken from [50]. The errors on the cross sections assumed will be discussed in Section 8.2.

At this point it should be noted that the MC-statistics are crucial for the Search Algorithm (see Section 7). In this algorithm the MC-mean is interpreted as the "true value" nature realizes. As the MC generator does not know the true mean value, the statistical error of the MC-prediction has to be considered. As a certain amount of Monte Carlo events is generated for each process, this number of events corresponds to a certain luminoisty, the MC-luminosity $\mathcal{L}_{MC}$ (see Table 5.1). This number is different from the actual data-luminosity, so MC has to be scaled down. Of course the statistical error of the MC-mean becomes negligible if the MC-luminosity exceeds the analyzed data luminosity by orders of magnitude. As a consequence, all available MC-samples were used to optimize the statistics.

5.3 Energy Scale and Smearing

A detector measures a signal of a certain height if a particle traverses. This signal must be transformed into an energy value. If a jet showers in the calorimeter, only a certain amount of the total energy can be measured because inert absorbing material which is not instrumented exists. The calorimeter, therefore, must be calibrated, which means that the measured hardware signal is multiplied by a certain factor, the *Jet Energy Scale* (see Section 4.3.4). This energy value then reflects the true physical value of the jet. Similar energy scales must be applied to all particles, even though the JES is the largest and has a significant uncertainty. A detector simulation must mimic all these energy scales. Besides this, every subsystem of the detector responsible for measuring a certain particle, e.g. the tracker or the calorimeter, has a certain intrinsic momentum or energy resolution. A detector simulation must describe these resolutions properly.
The detector simulation is often insufficient to reproduce the exact data distributions, so one must often fine-tune the MC. The analysis package Top Analyze [23] provided by the Top Group implements all smearing factors and energy scales, each of them determined in a separate analysis. Only the MET-smearing is performed self-contained within the analysis code. The following parameterizations can be found in [12], combined with the values used in the program version "Top Analyze-Stradivarius".

Electrons

The electron energy is smeared with a random Gaussian with mean zero using the following formula:

$$E' = E \cdot \alpha + E \cdot Gauss(0, \sigma) \; . \tag{5.2}$$

The energy scale is the corretion factor $\alpha = 1.007 \pm 0.001$, and the smearing factor is $\sigma = 0.042 \pm 0.004$ for fiducial electrons (i.e. energy deposition ouside the calorimeter cracks, see Section 4.3.1). For non-fiducial electrons, one obtains $\alpha = 0.971 \pm 0.012$ and $\sigma = 0.083 \pm 0.006$. The parameters are determined by comparing the Z-peak position and width in data and MC. The relative errors can be found in [51], bearing in mind that the mean parameters in this analysis have changed slightly compared to the ones stated in [51] as a new version of the analysis program Top Analyze is used.

Muons

Muon p_T is measured by determining the sagitta, which is $\propto 1/p_T$. This leads to a smearing function:

$$\frac{1}{p'_T} = \frac{1}{\alpha \, p_T} + Gauss(0, \sigma) \; . \tag{5.3}$$

Here $\alpha = 0.991 \pm 0.003$ and $\sigma = 0.0025 \pm 0.0002$ [GeV^{-1}]. Again the Z-peak serves as a calibration tool, and the errors can be found in [52].

Jets

The Jet Energy Scale is by far the largest energy scale, and, unfortunately, also one of the major sources of uncertainty. JES is applied for both data and MC independently. There may still be a difference in the scale of jets between data and MC, but no further scale correction is applied as the JES itself is too uncertain. In this analysis, the JetCorr v5.3 [39] package is used, which applies JES for data and MC and has an improved understanding of JES, leading to a smaller uncertainty. In the data the JES averages a factor of 1.6. In this analysis the uncertainty of the JES is estimated from the plots provided in [39] resulting in $\sigma_{rel} = 6\%$.
In addition to the Jet Energy Scale, jets in Monte Carlo simulations are also smeared by Top Analyze. The following parametrization of the jet-resolution from Section 2.2.2 is used (see values of N, S and C for data there):

$$\frac{\sigma(p_T)}{p_T} = \sqrt{\frac{N^2}{{p_T}^2} + \frac{S^2}{p_T} + C^2} \tag{5.4}$$

The fit parameters N, S and C are different for data and MC. For every jet in a Monte Carlo sample, $\sigma(p_T)$ is computed with the data parameters and the MC parameters. If the MC resolution is better than in the data, the MC jet is smeared once more with a Gaussian of the width :

$$\sigma(\text{smear}) = \sqrt{\sigma^2(\text{data}) - \sigma^2(\text{MC})} \tag{5.5}$$

Missing Transverse Energy (MET)

Missing Transverse Energy is reconstructed in several steps, first by being corrected for hadronic/electromagnetic calorimeter energies and then for muon momentum (see Section 4.3.5). As jets, electrons and muons are smeared in MC independently, the MET should theoretically fit to the one observed in data without further manipulations. In the case of muons this is observed to be true.
Unfortunately, the MET-distributions of processes with electrons do not match with the data also in regions where New Physics is excluded by former analyses (e.g. W-peak). One possible explanation is that electrons are objects measured by the calorimeter. Here phenomena like unclustered energy, noise (calorimeter or coherent noise) and misidentification of the primary vertex are difficult to simulate in MC. This results in the electron MC poorly describing the MET of the data, with data-distributions much broader than expected. As the reasons for these differences between data and MC are not fully understood, in this analysis MET-smearing is used only to estimate the *sytematic error*.

The following parametrization is used to explicitly smear the MET of MC events with electrons and can be found in [53]:

$$\begin{aligned} METx(smear) &= METx + Gauss_1(0,\sigma) \\ METy(smear) &= METy + Gauss_2(0,\sigma) \\ MET(smear) &= \sqrt{METx(smear)^2 + METy(smear)^2} \end{aligned} \tag{5.6}$$

$$\text{with} \quad \sigma = (2.553 \pm 0.094)\ \text{GeV} + (0.009 \pm 0.002) \cdot \sum E_{Tunclus}$$

Here $MET_{x,y}$ denotes the final Missing Tranverse Energy after all calorimeter and muon correction have been applied (see Section 4.3.5). The smearing parameters are determined by analyzing the MET of $Z \to ee$ generated events. As no MET contribution is expected here, this reflects the MET-resolution, which is primarily determined by the amount of unclustered energy $E_{Tunclus}$. Unclustered energy represents calorimeter cells which cannot be linked to physical objects like jets or electrons. These cells contribute to MET, as p_T must be balanced. In Figure 5.5 both smeared and unsmeared MC in the exclusive event class $1\nu\,1e$ are compared to data. In the MC mean distribution an excess in the peak region and a deficit in the high-p_T tail stands out. This illustrates the need for MET-smearing. The smeared MC shows a better agreement with the data, but still the data are not described perfectly.

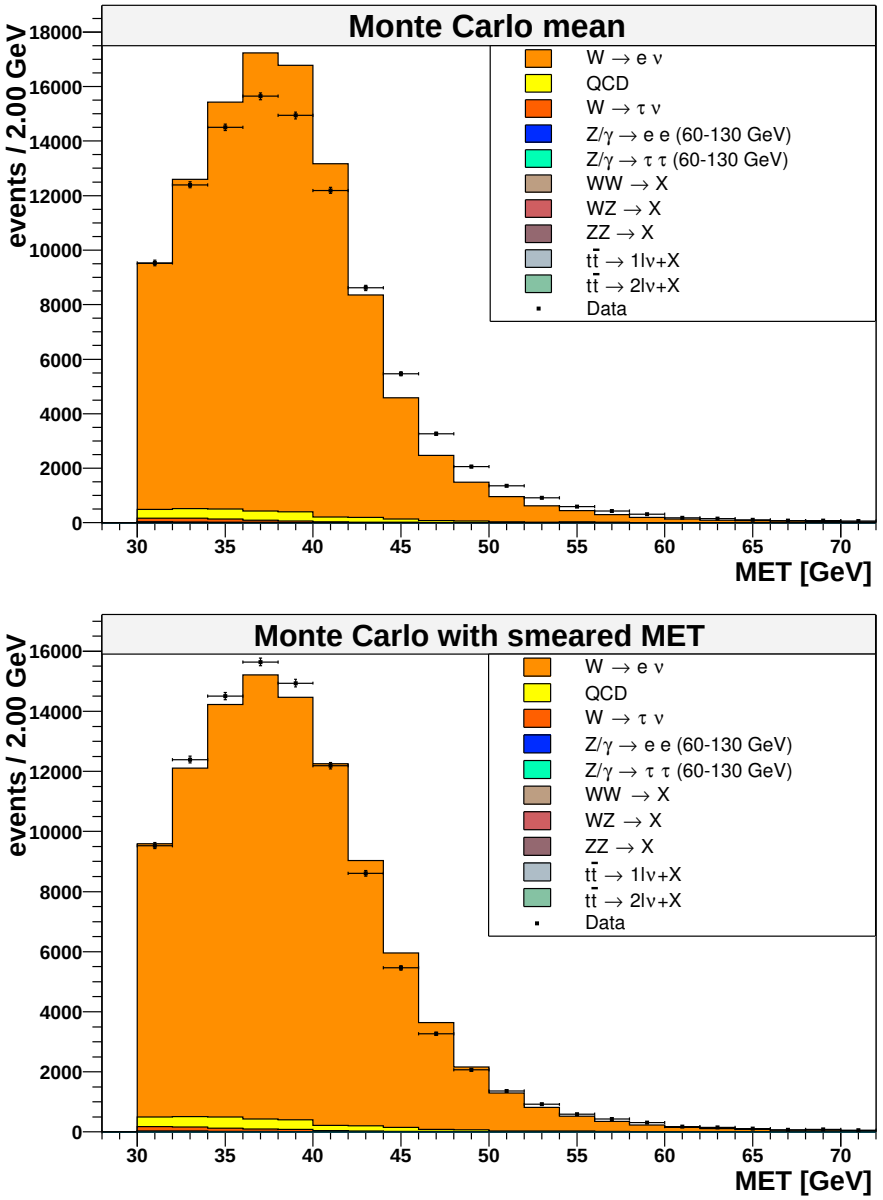

Figure 5.5: *MET-distributions of $1\nu\,1e$ exclusive event class. On the top MET is unmodified, on the bottom MET-smearing is applied*

Chapter 6

Global Data-Monte Carlo Comparison

6.1 MC-Correction-Factor

6.1.1 χ^2-Fit

The samples of Monte Carlo simulations (MC) used correspond to a certain amount of luminosity as presented in Table 5.1. In order to be comparable to the data analyzed, the MC distributions have to be scaled to the luminosity of the data $\mathcal{L}_{data}$. Besides this, the official DØ-MC do not include a trigger simulation, so an efficiency correction must be implemented.
The final correction necessary is perhaps the most difficult and uncertain one: To select events, various object identification cuts are applied both in data and in Monte Carlo, and all these different cuts have certain efficiencies which result in a general reconstruction efficiency ε^{reco}. Even though MC attempts to describe data as accurately as possible, there might be considerable differences between the efficiencies determined by data and the efficiencies extracted from MC. For an event class with only one lepton, this can be summarized as follows:

$$N_{MC}(scaled) = N_{MC}(unscaled) \times \frac{\mathcal{L}_{data}}{\mathcal{L}_{MC}} \times \varepsilon_{trig} \times \frac{\varepsilon^{reco}_{data}}{\varepsilon^{reco}_{MC}} \; . \tag{6.1}$$

This approach, however, appears technically impracticable and susceptible to errors. Determing accurate efficiencies for the multitude of cuts used in this analysis is almost impossible. One should bear in mind that efficiency studies make up a major part of a conventional analysis and often dominate the final error of the measured cross section. In addition to this, the uncertainty of the data luminosity (6.5%) also contributes to the error (see Equation 6.1).
This analysis chooses a different solution to the problem of MC-scaling by taking advantage of a certain calibration region. The idea is comparable to the mechanism of MC-smearing, where the Z-peak serves as a calibration tool as no new physics is expected there.

This analysis defines two *regions of calibration*:

- For event classes with one lepton, the 30−50 GeV p_T-range of the exclusive $1\nu\,1lepton$ events class is defined as a non-signal region
- For event classes with two muons or electrons, the $75-95$ GeV $\sum p_T$-range of the exclusive $2lepton$ events class is defined as a non-signal region

The former region is dominated by W-boson production, and the latter by Z-boson production. Both processes are regarded as basic Standard Model reactions well understood by the detector and well simulated by MC. Various analyses of various experiments have measured the Z- and W-peak without any indication of new physics, so the assumption of these two "standard candles" seems appropriate. The upper limits of 50 GeV and 95 GeV respectively still allow possible deviations in the high-p_T tail of these distributions, e.g. possible signals of exited vector bosons.

The MC-correction-factor f_{MC} is determined by summing up all MC contributions and comparing this to the data. The minimum of the χ^2

$$\chi^2 = \sum_{bin=i} \frac{(N_i^{data} - f_{MC} \cdot N_i^{MC})^2}{f_{MC} \cdot N_i^{MC}} \tag{6.2}$$

is found by plotting χ^2 as a function of f_{MC}. In the above expression N_i^{MC} is the sum of all MC in bin i after the MC has been scaled to the data luminosity. In this way f_{MC} is only the product of the efficiency factors. As a consistency check, the fraction $\frac{\varepsilon_{data}^{reco}}{\varepsilon_{MC}^{reco}}$ can be extracted and compared to rough efficiency estimations. Even though MC have been scaled to the data luminosity, the uncertainty in luminosity is still irrelevant. By scaling the MC to the data in this specific region with the determination of χ^2_{min}, any error is compensated (if $\mathcal{L}_{data}$ is twice as high as the "true" value, then f_{MC} will be ≈ 0.5).

Figure 6.1 shows the χ^2-distributions for the $1\nu\,1e$ class, the $1\nu\,1\mu$ class looks very simlar; a plot adequate to the 2 lepton class can be found in [1]. The final MC-Correction-Factors used for the rest of the analysis are:

- $f_\mu = 0.749 \pm 0.003(stat.) \pm 0.025(syst.)$ at $\chi^2_{min}/dof = 138.2/19 = 7.3$
- $f_{\mu\mu} = 0.763 \pm 0.018(stat.) \pm 0.049(syst.)$ at $\chi^2_{min}/dof = 29.7/19 = 1.56$
- $f_e = 0.894 \pm 0.003(stat.) \pm 0.022(syst.)$ at $\chi^2_{min}/dof = 142.0/19 = 7.5$
- $f_{ee} = 0.791 \pm 0.012(stat.) \pm 0.041(syst.)$ at $\chi^2_{min}/dof = 33.4/19 = 1.76$

At first view, the values of χ^2_{min} are surprisingly high. One has to bear in mind that the computation of χ^2 only includes statistical errors of the Monte Carlo simulations. An

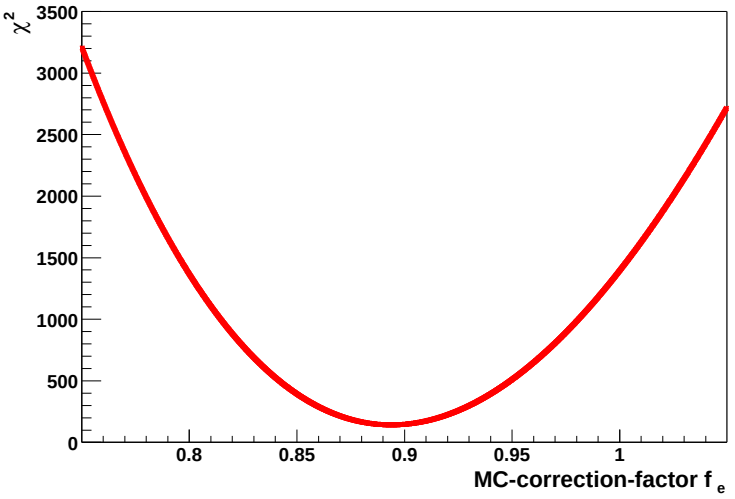

Figure 6.1: *χ^2-distribution for electron sample.*

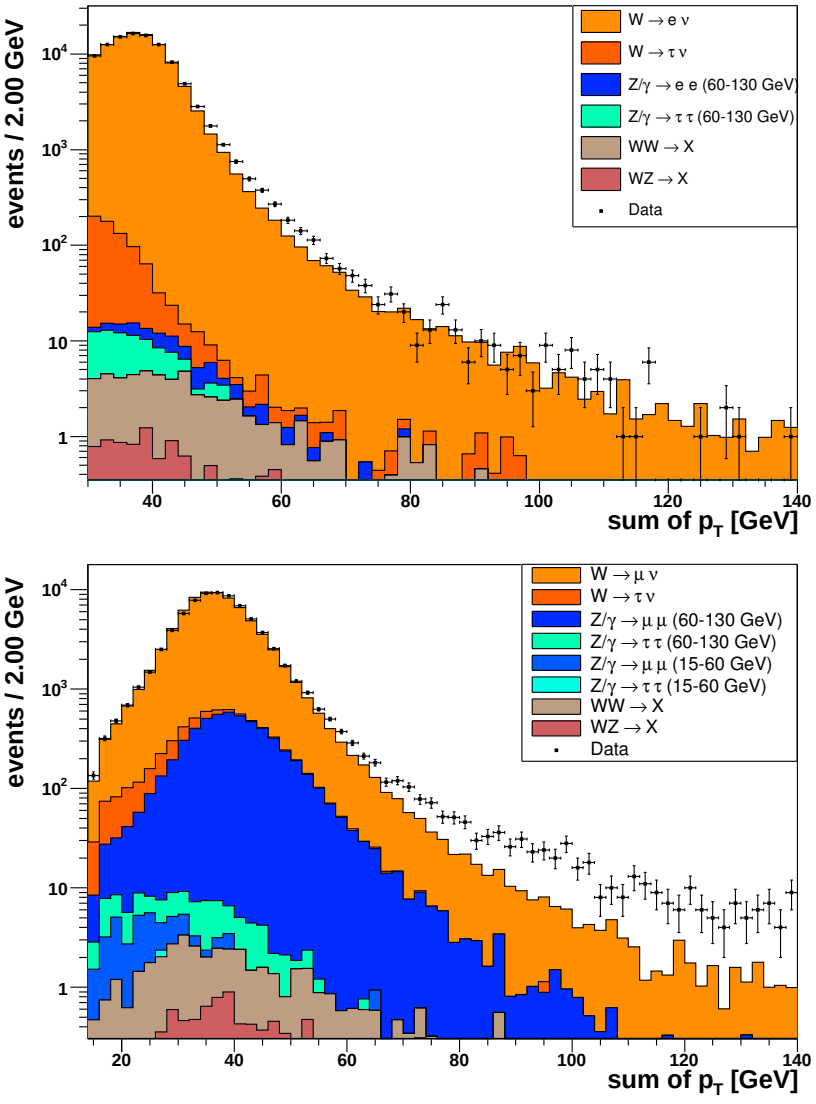

Figure 6.2: *p_T-distribution in the exclusive $1\nu\,1e$ class(top) and in the exclusive $1\nu\,1\mu$ class(bottom)*

estimation of the systematic error will be discussed below in Section 6.1.2. The MC-Correction-Factor for muons is smaller than the one of electrons. One reason for this is that the trigger efficiency for the single-muon trigger is much smaller; another is that more quality cuts to ensure a good muon measurement have been made, all of which are candidates for slight efficiency differences in data and MC.
Figure 6.2 shows the result of the scaling procedure, the order of MC processes in the legend is identical to the order in the distribution. Here, the p_T distribution of the single lepton class is shown after all MC-correction-factors have been apllied. All MC samples are included in this plot, but only the major contributors can be found in the legend. In the muon sample as well as the electron sample, data and MC agree well in the region where the scaling is done. Both distributions show an excess of data in the high-p_T tail, in the muon case this discrepancy is huge. This excess will be discussed in detail in the comparison of the total number of events (see Section 6.3) and the Search Algorithm will have to evaluate the significance of this deviation (see Section 9). In order to check the general agreement between data and scaled MC, η and φ distributions can serve as control plots. Figure 6.3 shows the φ distribution of the inclusive $1\nu\,1\mu + X$ event class and the η distribution of the inclusive $1\nu\,1e + X$ event class with all selection cuts applied (QCD which will be discussed next Section is also included, but is negligable here). Inclusive classes have been chosen, as many physical processes contribute to this class, and a general agreement should be observable without further manipulation. Both data and MC show the octants of the muon system, and the bottom hole is described well. The dips in the electron-η plot correspond to the transition areas between the calorimeter towers.

6.1.2 Systematic Uncertainty of the MC-Correction-Factor

To estimate the systematic uncertainty of the MC-correction-factor, the results for both independent non-signal regions where the scaling is performed (W-peak and Z-peak) are compared. The problem is that for an event class with 2 leptons, Equation 6.1 does not hold. Here two leptons must be reconstructed, and both could fire the single lepton trigger. For a two-lepton class the appropriate expression is:

$$N_{MC}(scaled) = N_{MC}(unscaled) \times \frac{\mathcal{L}_{data}}{\mathcal{L}_{MC}} \times \left(1 - (1 - \varepsilon_{trig})^2\right) \times \left(\frac{\varepsilon^{reco}_{data}}{\varepsilon^{reco}_{MC}}\right)^2 . \qquad (6.3)$$

If the trigger efficiency is known, than the fraction $\frac{\varepsilon^{reco}_{data}}{\varepsilon^{reco}_{MC}}$ that represents the different cut efficiencies in data and MC can be determined using two different kinematic regions of the phase space. As the trigger efficiency represents an additional uncertainty, for event classes with one lepton, f_μ and f_e is assumed as the central value, and the MC-correction-factor

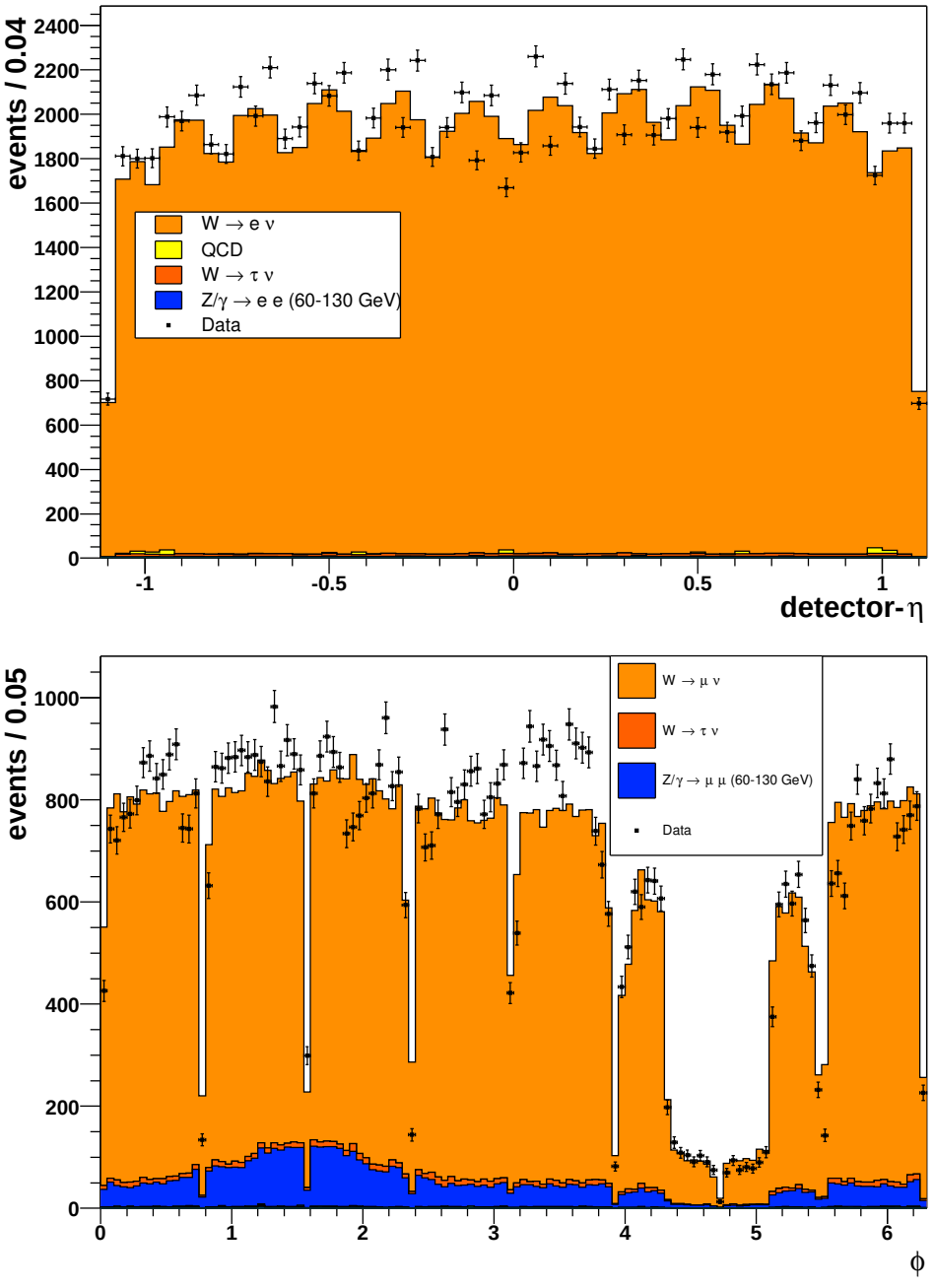

Figure 6.3: *η-distribution of electrons (top) and φ-distributions of muons (bottom)*

from the corresponding two-lepton class is used to estimate the systematic error:

$$f_\mu = \varepsilon_{trig} \times \frac{\varepsilon^{reco}_{data}}{\varepsilon^{reco}_{MC}} = 0.749 \tag{6.4}$$

$$f_{\mu\mu} = \left(1-(1-\varepsilon_{trig})^2\right) \times \left(\frac{\varepsilon^{reco}_{data}}{\varepsilon^{reco}_{MC}}\right)^2 = 0.763 \tag{6.5}$$

$$\Longrightarrow \frac{\varepsilon^{reco}_{data}}{\varepsilon^{reco}_{MC}} = \sqrt{\frac{0.763}{(1-(1-\varepsilon_{trig})^2)}} \tag{6.6}$$

$$f_\mu(Z-peak) = \varepsilon_{trig} \times \sqrt{\frac{0.763}{(1-(1-\varepsilon_{trig})^2)}} \tag{6.7}$$

Until now, trigger-efficiency has not been included. The assessment of this efficiency is only a rough estimate, and no precision measurement was performed as compared to conventional analyses. The reason here again is that the procedure of scaling MC to data using calibration peaks is more trustworthy than a separate trigger study. ε_{trig} is needed only for the purpose of estimating the systematic error.
For both skims used in this analysis, the EM1TRK-skim (electrons) and the 1MUloose-skim (muons), trigger efficiency is determined from an independent skim. The efficiency of a trigger is defined as:

$$\varepsilon_{trig} = \frac{\text{number of selected leptons + trigger fired}}{\text{number of selected leptons}} \; . \tag{6.8}$$

For combinations of triggers the separate efficiencies must be weighted according to the luminosity where the trigger is used. All electron triggers have very similar efficiencies, and the combined result is (see [31]):

$$\varepsilon_{trig}(e) = 0.980 \pm 0.013 \; . \tag{6.9}$$

The muon triggers used have considerably different efficiencies as the MUW_W_L2M3_TRK10 trigger is a major improvement on the old MU_W_L2M5_TRK10 trigger. As the new trigger was unprescaled most of the time, the trigger efficiency can be determined from the 1MUloose-skim by creating a sample where an electron trigger is required to be fired. In this way the sample is independent from the muon trigger studied as the 1MUloose-skim itself is composed without using any trigger information. By using the above expression and requiring the muon trigger to be fired, the trigger efficiency can be estimated. The result is $\varepsilon_{trig} = 0.92 \pm 0.03$(stat.). For the MU_W_L2M5_TRK10 trigger a study from [54] is used. Here the efficiency is separated into the different trigger levels: $\varepsilon_{L1} = 0.85 \pm 0.01$, $\varepsilon_{L2} = 0.87 \pm 0.01$, $\varepsilon_{L3} = 0.81 \pm 0.02$ can be combined to $\varepsilon_{trig} = 0.60 \pm 0.02$. Both muon trigger efficiencies now are weighted according to the luminosity of their usage:

$$\varepsilon_{trig}(\mu) = \frac{0.60 \cdot 29 \text{ pb}^{-1} + 0.92 \cdot 189 \text{ pb}^{-1}}{29 \text{ pb}^{-1} + 189 \text{ pb}^{-1}} = 0.88 \pm 0.03 \; . \tag{6.10}$$

Finally, the systematic error of f_μ is:

$$\sigma(syst) = |f_\mu(Z-peak) - f_\mu| = |0.774 - 0.749| = 0.025 \; . \tag{6.11}$$

For a MC-correction-factor of a di-lepton class, the procedure is identical, e.g. for the central value f_{ee}:

$$\sigma(syst) = |f_{ee}(W - peak) - f_{ee}| = |0.832 - 0.791| = 0.041 \ . \tag{6.12}$$

QCD-contribution is neglected in the determination of the MC-correction-factor and the issue of QCD-background will be discussed in Section 6.2. One could argue that this is a systematic uncertainty. Examined more closely, it appears that the disregard of QCD is well justified. The QCD-contribution within the non-signal-region of $30-50$GeV p_T-range for muons is very small, the contribution to the Z-peak in the other non-signal-region is vanishing for muons as well as for electrons. Nevertheless there is a certain contribution which could modify the MC-correction factor, especially in the electron case.
If the QCD background is added to the sum of MC and the χ^2 minimization is repeated, $f_\mu = 0.749$ does not change, and f_e becomes $f_e = 0.873$. As expected, the influence in the electron class is much greater as the QCD background is much larger. Nevertheless, the change in f_e is well within the already assumed systematic error. The estimation of the QCD background itself is not very precise. Because of that QCD background is ignored in the estimation of the MC-correction-factors. The assumed systematic error accounts for possible QCD contributions.

6.2 QCD-background

In order to analyze the total number of events in data and MC properly, the QCD processes have the be included. The estimation of the QCD-background is somewhat more complex than selecting the appropriate Standard Model Monte Carlo simulations. The set of official MC does not include a QCD-sample with sufficient statistics. There are technical reasons for this, as the production of QCD-processes mimicking events with isolated leptons is non trivial. Important processes at a hadron collider like $b\bar{b}$-production ($\sigma \approx 0.02$mb) have enormous cross sections compared to vector boson production ($\sigma_W \approx 30$nb). The events interesting for background estimation are the ones where particle decays in a QCD-jet produce an isolated muon or electron and thus contribute to one of the event classes of this analysis. For electrons another scenario is important: Most hadron jets have a large electromagnetic fraction, mostly due to pions ($\pi^0 \rightarrow \gamma\gamma$). As a consequence there is a certain probability that a jet is misidentified as an EM candidate. As jets inherit many charged particles which create tracks, e.g. $\pi^\pm$, these EM candidates can be misinterpreted as an isolated electron. As the word will be used several times in this analysis, the following definition is made:

"fake leptons from QCD": All processes that create isolated leptons in pure QCD-events

Unfortunately, most of the QCD-events generated are low-p_T "all jet" events which do not produce isolated leptons. As a consequence, an enormous number of events must be generated to develop a proper background sample. Kinematic cuts are made at the

generator level to select events with fake lepton candidates and pass only these on to the time-consuming detector simulation (with the help of the "generator level event selector" program D0_mess [55]). Finally, after a few fake-lepton events have been processed, these events provide only a very small luminosity as $N = \sigma \cdot \mathcal{L}$. When scaled to the luminosity considered in this analysis, enormous factors emerge, leading to a significant uncertainty in the background estimates.
Most DØ-analysts have decided to take a different approach by estimating the QCD-background directly from the data. This is only a "back door" approach, and it has many disadvantages, but for this analysis it appears to be the only possibility. One of the major disatvantages of estimating QCD from data is that the luminosity of the constructed background sample is the same as the data luminosity. QCD contributions estimated with upper limits cannot be reduced by generating more events (as it is possible in Monte Carlo samples) corresponding to more luminosity. This would result in a scaling-down of the upper limit.
The estimation of the QCD-background can be seperated into three steps, each of them can be illustrated with Figure 6.4:

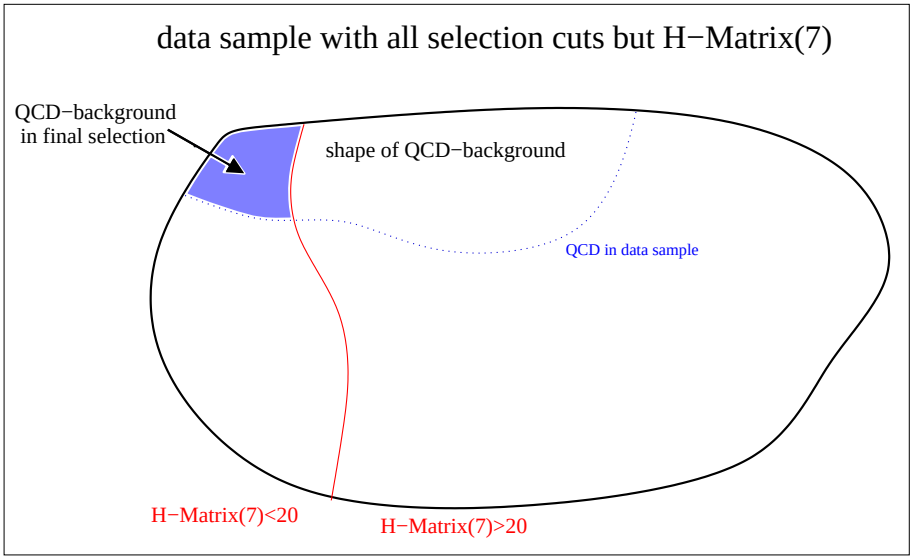

Figure 6.4: *Scheme of QCD-background estimation from the data sample.*

1) One specific cut is *inverted*, all other cuts are maintained.

In this step, the data sample is divided into two parts, illustrated by the red line in Figure 6.4 (H-Matrix(7) as an example for a cut). The blue dotted line outlines all QCD events in the data sample; some of them are in the final selection (i.e. "fake leptons from QCD"), most of them do not fullfill the last cut. By inverting the final cut, the whole right side of the data sample is selected.
The inverted cut must be a good discriminator between "real" leptons and "fake" leptons from QCD-events. In the case of electrons the H-Matrix(7) cut is inverted, as the shower shape characterizes real electrons very well:

$$\text{H-Matrix}(7) > 20\ . \tag{6.13}$$

For muons the isolation-cut is relaxed by:

$$\begin{aligned} &2.5\text{ GeV} \;<\; \text{Halo}(0.1, 0.4) < 7.5\text{ GeV} \;\text{ and }\; \text{TrkCone}(0.5) < 7.0\text{ GeV} \\ \text{or} \quad &(\,2.5\text{ GeV} \;<\; \text{TrkCone}(0.5) < 7.0\text{ GeV} \;\text{ and }\; \text{Halo}(0.1, 0.4) < 7.0\text{ GeV}\,)\,. \end{aligned} \tag{6.14}$$

Especially heavy quark jets tend to produce leptons. Most of these decay products from secondary vertices are non-isolated, but the probability of a muon from a b-jet getting a kick large enough to become isolated and therefore pass all selection criteria cannot be ignored. Again, relaxing the isolation cut produces a sample very similar to the QCD-fakes of the final sample.

The basic idea of background estimation from data is the creation of a sample comparable to the expected fake-QCD events present in the selected data. The sample with the inverted cut is expected to have similar kinematic properties (p_T and MET) to the ones from QCD-fakes of the final selection. These fakes within the final sample will behave similar to the events with inverted cuts as their only difference is the discriminating cut.

2) All possible contributions of MC with inverted cuts are removed from the data with inverted cuts. This provides the *shape* of the QCD-background.

In this way only true QCD events are selected. As regards Figure 6.4, this step selects the area "shape of QCD-background". The size of the QCD-sample is defined as:

$$N_{QCD}(\text{inverted cut}) = N_{data}(\text{inverted cut}) - N_{MC}(\text{inverted cut})\,. \tag{6.15}$$

3) The QCD-sample is normalized by scaling it to the data in a specific region.

It is important to understand that the inversion sample is not the final estimation of the QCD-background; only the *shape* of the p_T and MET distributions is used. After these distributions are obtained, a scaling factor must be determined in order to properly normalize the QCD background. Then, finally, a sample of the QCD-background in the final sample is modeled (blue area in Figure 6.4).
In this final step, the QCD sample with inverted cuts is scaled to the data of the final selection (no inverted cuts, left from red line in Figure 6.4) by minimizing a χ^2. The same χ^2-method has been applied before in the determination of the MC-correction-factor (see Section 6.1). This analysis selects events with a relatively large amount of Missing Transverse Energy, which additionally reduces the contribution of QCD fakes.
The scaling is done using the transverse mass distribution of the $1\mu(e)+MET+X$ inclusive single lepton class, which for the muon class is defined by:

$$M_T = \sqrt{(p_T(\mu) + MET)^2 - (p_x(\mu) + MET_x)^2 - (p_y(\mu) + MET_y)^2}\,. \tag{6.16}$$

QCD-contribution is expected in the low energy region, and the inclusive class is chosen as most fake lepton events have additional jets. For the estimation of QCD, the bins $0-20$ GeV of the M_T-distribution in the $1\ \mu(e)+MET+X$ inclusive class are chosen. This scaled QCD-sample can contribute to all possible event classes. The result is:

$$N_{QCD} = f_{scale} \times N_{QCD}(\text{inverted cut})\ , \tag{6.17}$$

$$f_\mu = 0.24 \pm 0.02(\text{stat.}) \pm 0.09(\text{syst.}) \quad \text{at} \quad \chi^2/dof = \frac{98}{19} = 5.2 \tag{6.18}$$

$$f_e = 1.58 \pm 0.06(\text{stat.}) \pm 0.13(\text{syst.}) \quad \text{at} \quad \chi^2/dof = \frac{88}{19} = 4.6 \tag{6.19}$$

As only statistical fluctuations are considered in the computation of χ^2_{min}, these values are relatively huge. Only by including all systematic errors a better χ^2_{min} can be achieved. Nevertheless, for determination of the scaling factor only the f-value of the minimum of χ^2 needs to be found to give a proper estimation of the QCD-background.

The systematic error is determined by repeating the χ^2 fit, once using only bins $0-10$ GeV of the M_T-distribution, and once examining the $\Delta\varphi(MET, lepton)$ distribution. The greatest difference in the scaling factor yields an estimate of the systematic error. QCD-fakes are expected to contribute at small angles in $\Delta\varphi(MET, lepton)$ as these events have fake MET contributions at small angles arising from a mis-measured muon or a jet misidentified as an electron without the proper Jet Energy Scale.
The following plots illustrate the results of the QCD-scaling and show the M_T-distributions (Figure 6.5). The $\Delta\varphi$ distributions (Figure 6.6) serve as control plots. The order of the different MC processes in the legend is identical to the one in the distribution. With the scaled QCD contribution included, data-MC comparison now fits well in the small angle region and the low energy range of M_T. The fact that in the electron case the W-peak in MC is narrower than in data is due to the problem discussed earlier, that MET needs further smearing. This will be accounted for later as a systematic uncertainty (see Section 8.5).
The sudden build-up at ≈ 60 GeV is due to the sharp W-peak in the electron data, as $W \to e\nu$ and $W \to \tau\nu$ contributions increase significantly. This build-up is the Jacobian-peak where the W-boson is produced at rest and decays into an electron and a neutrino, each carrying half of the mass m_W with their momentum p. As the W can also inherit a transverse momentum due to final state radiation, electrons with $p_T > m_W/2$ are possible which explains the high-p_T tail of the distribution. With its worse momentum resolution, the muon data do not show this behaviour.
In the electron sample, QCD-fakes are a much bigger background than in the muon sample. Misidentifying a hadron jet as an EM object is more likely than detecting isolated muons from hadron decays.

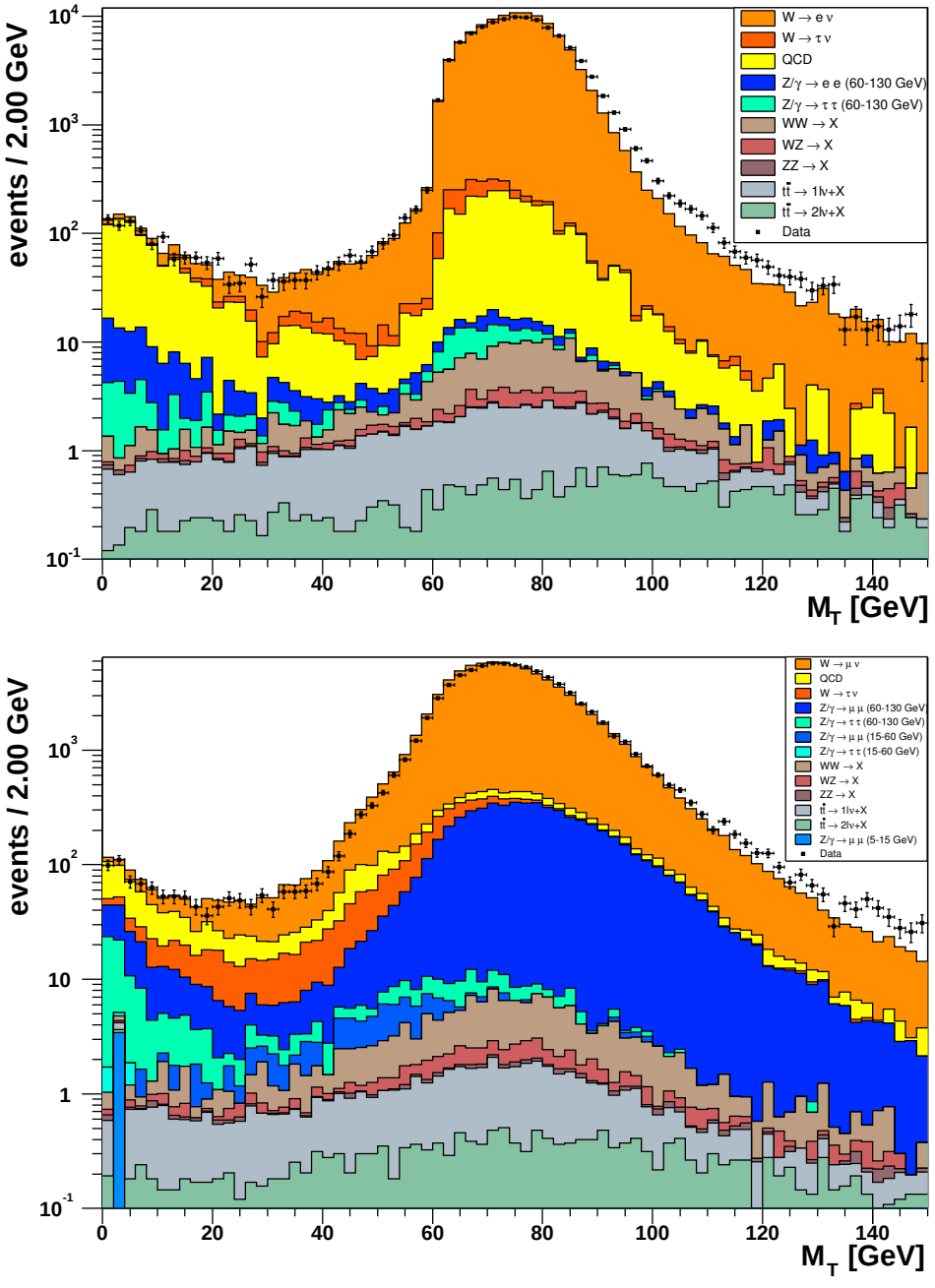

Figure 6.5: *Transversal mass of electron-data(top) and muon-data(bottom)*

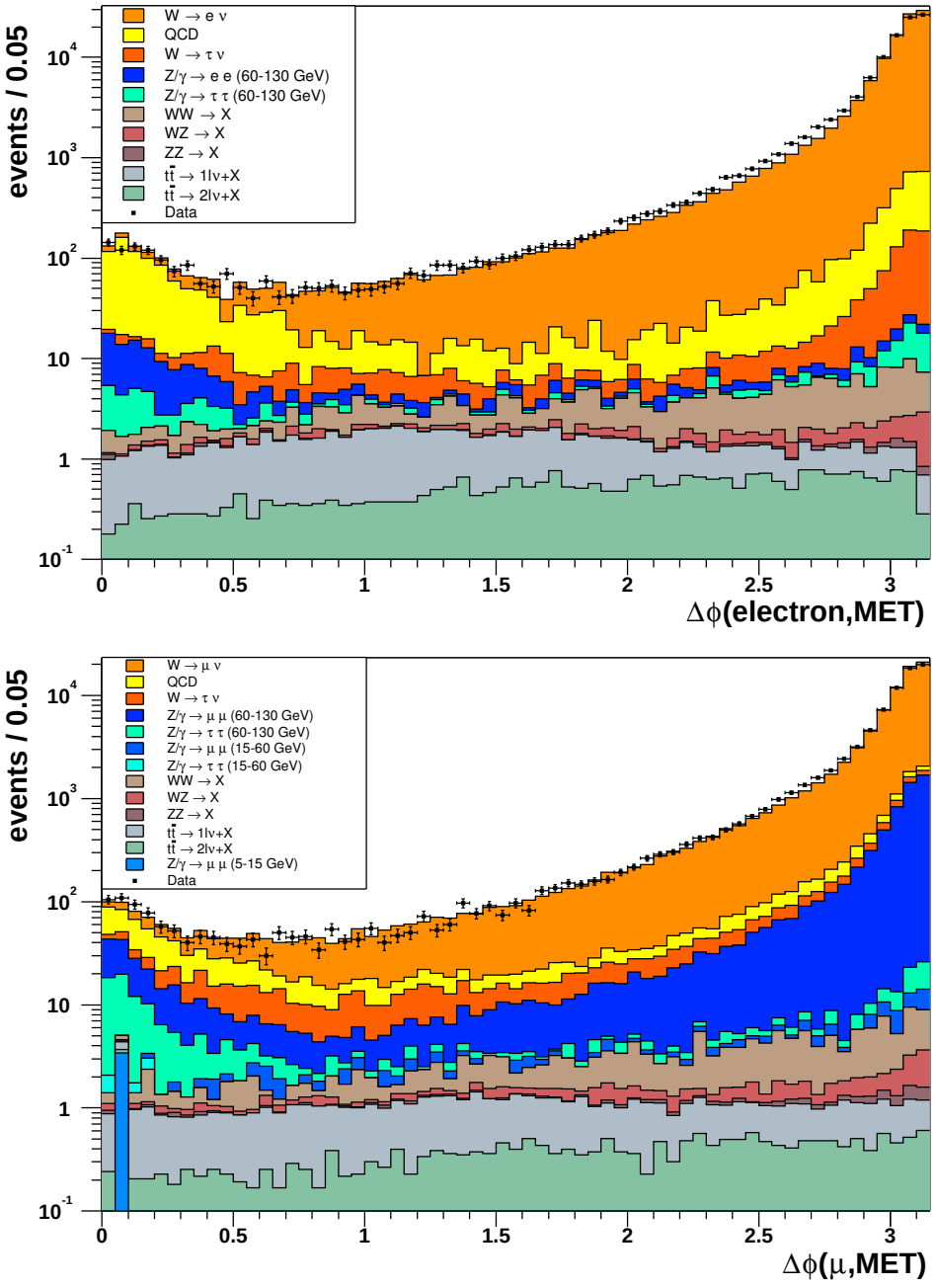

Figure 6.6: *Δφ-distribution as control plot for QCD-background (top:electrons, bottom:muons)*

6.3 Comparison of the Total Number of Events

The following Tables 6.2-6.9 represent the sum of all events per event class in data and Monte Carlo simulations (MC)found in the muon data and in the electron data. All exclusive and inclusive classes where data were recorded are compared to the sum of the relevant MC contributions.
For all MC contributions, only the statistical error is accounted for. This general comparison is only meant to give a rough estimate of the agreement between selected data and Standard Model predictions. Systematic errors are discussed and included later for the Search Algorithm. Nevertheless, the systematic uncertainties can be estimated at the $\sim 3\%$ level (see syst. error of the MC-correction-factor in Section 6.1). As a consequence, all deviation smaller than this should not be regarded as a problem. The statistical error of MC is defined as:

$$\sigma(N_{MC}) = \sqrt{N_{MC}(unscaled)} \cdot \frac{\mathcal{L}_{data}}{\mathcal{L}_{MC}} \cdot f_{lepton} \ . \tag{6.20}$$

If the statistics are low, the unscaled MC contributions are not Gaussian distributed, i.e. $N_{MC}(unscaled) < 5$. In this case the 68% confidence level upper limits are taken:

$N_{MC}(unscaled)$	upper limit
0	1.15
1	2.36
2	3.52
3	4.65
4	5.77

Table 6.1: *Upper limits calculated for 68% confidence level.*

All Gaussian statistical errors are added in quadrature and upper limits are added linearly. As an example, one MC contribution with 3 ± 0.5, one with < 1.5 and one with 7 ± 1 are combined to:

$$\text{Gaussian error:} \quad \sqrt{0.5^2 + 1^2} = 1.12 \tag{6.21}$$

$$\text{upper limits:} \quad < 1.5 \tag{6.22}$$

$$\Longrightarrow 10^{+2.62}_{-1.12} \tag{6.23}$$

In contrast to the Monte Carlo samples, the QCD-events estimated from data are assumed to be Gaussian distributed with a width defined by the total error of the QCD-scale-factor. As this systematic error dominates, statistical errors can be neglected. Event classes where the QCD estimation from data does not provide a single event are assumed to be not contaminated by any QCD, and upper confidence levels are considered for the processes simulated by MC only.

In some event classes, certain SM contributions are treated differently in the sense that no upper limits are used for them. This is the case when the MC process obviously cannot contribute to the event class under consideration, and taking upper limits would be too

conservative for the MC contribution. Different arguments for neglecting certain MC can be assembled. In the event class tables, the reason is stated by using abbreviations:

s: The "s"ignature of the process is completely different from the event class studied, and the MC can only contribute if some particles are misidentified. If the probability of this fake rate is very small, contributions can be neglected. Another reason could be that the invariant mass of the intermediate boson is too small to produce leptons fullfilling the selection cuts.
Example: $Z \to \mu\mu$ can hardly produce $1\nu\,1\mu\,1e\,1j$

b: A possible contribution is suppressed by an additional "b"ranching ratio further decreasing the cross section. The specific MC does not produce events in a superordinate class, so upper confidence levels are too conservative.
Example: Upper CL are used for $Z \to \tau\tau$ and the class $1\nu\,1\mu\,3j$, so for the $1\nu\,1\mu\,4j$ class the additional gluon radiation producing a jet leads to a vanishing contribution

It should be emphasized that no SM contribution is neglected. All processes are included, but the upper confidence level is used only for selected contributors.

For all classes the MC-correction-factors identified in Section 6.1 have been applied. For classes with one specific lepton f_l is used, for classes with two leptons of the same flavour f_{ll} is used, respectively. For classes with lepton mixtures, e.g. $1\nu\,1\mu\,1e$ in the muon data sample, f_μ is applied multiplied by $\frac{f_e}{\varepsilon_{trig}(e)}$ to account for the reconstruction efficiency correction of the additional electron.

6.3.1 Muon Sample

This section discusse the overall data-MC comparison of the inclusive and exclusive event classes of the muon skim (see Table 6.2 to Table 6.5). Before the diffences between data and MC are analyzed in detail, first a few remarks about the good agreement in general should be made. To summarize, exclusive and inclusive event classes show similar behaviour: Event classes with considerable differences between data and Monte Carlo in one sample show the same deviations in the other sample. This indicates that the deviations are caused by the exclusive event classes and the error propagates to the inclusive samples, where additional physical objects do not balance the observed discrepancy. In general the agreement between data and MC is satisfactory. There are no striking discrepancies, data and MC are always of the same order of magnitude. Most deviations can be explained by detector effects or by the absence of systematic errors in the sum of MC.

Beginning with the first class $1\nu\,1\mu(+X)$, data and sum of Monte Carlo are well within the error ranges, which are dominated by the relative uncertainty of the QCD-background ($\approx 38\%$ of σ_{QCD}). This agreement is forced as the MC-correction-factors are determined from this class. The factor determined from the limited calibration region $30 - 50$ GeV seems to describe the rest of the distribution very well.
All event classes with photons show no significant differences between data and MC, so the

initial and final state radiation of photons seems to be well simulated by the MC generator PYTHIA. No signal from non-SM processes like $\mu' \to \mu + \gamma$ can be observed. In addition to this, the identification of photons with the detector seems to be well understood for real data and well implemented in the detector simulation.
Most of the classes with exotic combinations like $1\nu\,1\mu\,1e + X$ or $1\nu\,2\mu\,1\gamma$ are consistent for data and MC. The differences are either well within the error ranges, or the statistics of the event class are so small that the Poisson law explains the observed deviations, e.g. $1\nu\,1\mu\,1e\,2j$ with data=4 and MC=$1.9^{+0.9}_{-0.1}$. A very interesting event is the single data point seen in $1\nu\,1\mu\,1e\,3j$. With an expectation of 0.2 events from MC, again the deviation is not disturbing. Looking at the different MC contributions, it becomes obvious that $t\bar{t} \to 2lX$ is the ideal candiate for this event. Even though the probability for a top-production with an additional gluon jet is not huge, and even tough the properties of the event are impressive ($p_T(e) = 136$ GeV, $p_T(\mu) = 30$ GeV, $MET = 84$ GeV, $p_T(jet1) = 95$ GeV, $p_T(jet2) = 93$ GeV, $p_T(jet3) = 49$ GeV), this event is one of the official top-candidates (see [56]). Event displays of this event with the number 8710859 from Run 174901 can be found in the diploma thesis by O. Kraff, see [1].

Deviations can be found in the event classes with n jets, $1\nu\,1\mu\,nj(+X)$. Reviewing these differences between data and MC, one must remember that systematic errors except for QCD are not included in these tables as yet. Especially in classes with jets, the uncertainty in the Jet Energy Scale of $\approx 6\%$ is a major source of error which puts the deviations into perspective.
In the $1\nu\,1\mu\,1j(+X)$ class an excess of Monte Carlo events can be observed. An excess of MC is primarily a sign for problems in the simulation, as negative interferences in data are very improbable. The presumption that MC does not model gluon radiation properly is enforced by Figure 6.7, where the $1\nu\,1\mu\,1j$ class of data and $W \to \mu\nu$ MC are compared to each other. In this plot the $|\eta|$-cut for jets was lowered from 1.5 to 0.5 in order to exclude the Intercryostat region of the detector, which is not well understood. Even though other contributions like QCD or $Z \to \mu\mu$ are neglected, this single Monte Carlo sample already fills the major part of the data distribution. Especially in the peak region, where other contributions like QCD or Z-MC cannot be neglected, MC almost exceeds the data. The uncertainty in the MC-correction-factor and in JES could reduce this discrepancy, but still, the PYTHIA Monte Carlos seem to overestimate the amount of single gluon radiation.
In contrast to this, exclusive and inclusive event classes with two and three jets show a significant excess of data. Whereas MC simulation seems to overrate the single gluon radiation, multiple gluon emission resulting in two or three jets does not describe the data properly. This fact is not surprising, as PYTHIA is a parton shower generator which does not use exact matrix element calculation (see Section 5.1). Initial- and final state radiation appears to be not well described by this parton shower approximation. Nevertheless, one must remember that systematic uncertainties can also explain parts of these deviations, and conclusions should not be drawn before considering the final results of the Search Algorithm in Section 9. In the event classes $1\nu\,1\mu\,4j(+X)$, the agreement between data and MC is very good. Here, the lack of statistics seems to hide any possible deviations.

The last obvious difference between data and MC can be found in the $1\nu\,2\mu(+X)$ event classes. Before an observation of New Physics can be claimed, first possible explanations refering to problems in the MC simulations must be discussed.

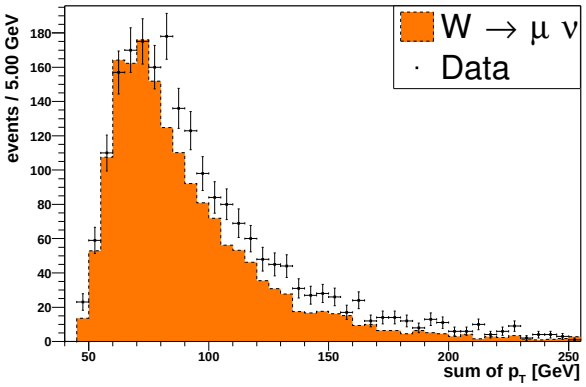

Figure 6.7: *Exclusive event class $1\nu\,1\mu\,1j$ of data and $W \to \mu\nu$ Monte Carlo simulation. In this particular plot only jets with $|\eta| < 0.5$ are considered.*

In the sum of the Monte Carlos processes, this class is dominated by $Z \to \mu\mu$ (60-130 GeV). Of course, Missing Transverse Energy is not a physical part of a Z decay. The only explanation of a contribution from Z-production is the mis-measurement of the muon p_T, resulting in a "faked" MET. It might be expected that the excess in data is due to the fact that MC assumes p_T-resolutions which are too optimistic. Mis-measured muons appear more often in the data than expected. In principle the smearing of MC, which is calibrated using the Z-peak (see Section 5.3), should describe these mismeasured muons. The discrepancy between data and Monte Carlo events in the $1\nu\,2\mu(+X)$ class shows that there are effects in the muon-p_T measurement which are not well understood and not well simulated by simply smearing the MC.
Figure 6.8 supports the hypothesis of an excess in data caused by mismeasured Z-events. This does not mean that Z-production is a process not well understood with the detector. In this particular event class, mis-measured Z-events are explicitly selected. The left plot shows that the muons are generally back-to-back, a behaviour typical of Z-decays. The right plot shows the angle between either of the two muons and the Missing Transverse Energy. One can see the excess of data at low angles and in the back-to-back region. Both of them indicate that the MET of the event is caused by a mis-measurement of the muon p_T. The generated Z-events shows a similar behaviour, but in real data such poor measurements seem to appear more often than assumed in the simulation.

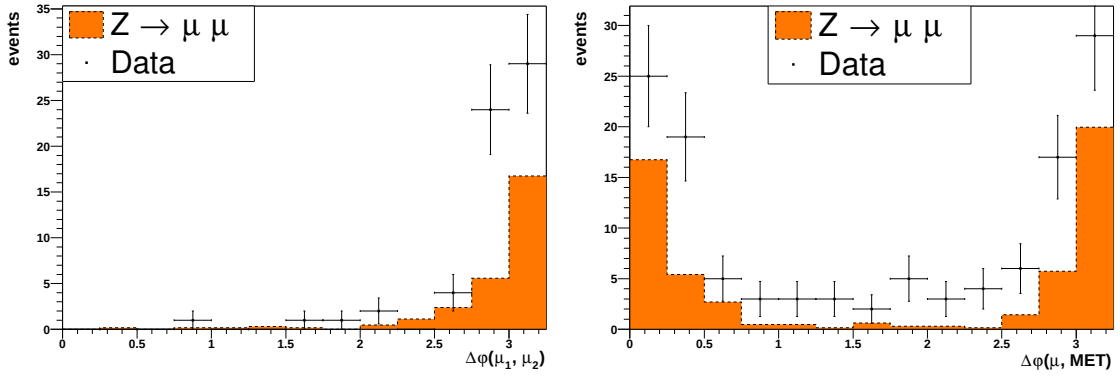

Figure 6.8: *Phi between the two muons (left) and phi between each muon and MET are shown for the $1\nu\,2\mu$ exclusive event class, data and generated Z-events in comparison.*

inclusive event classes of the muon data: part 1							
Class	Data	sum of MC	W$\rightarrow \mu\nu$	W$\rightarrow \tau\nu$	Z$\rightarrow \mu\mu$ ($5-15$ GeV)	Z$\rightarrow \mu\mu$ ($15-60$ GeV)	Z$\rightarrow \mu\mu$ ($60-130$ GeV)
1ν 1μ X	82500	82900±600	73600±100	1260±20	s	41±4	5990±30
1ν 1μ 1j X	5359	5650±160	4320±30	273±8	b	14±2	416±8
1ν 1μ 1e X	30	32^{+3}_{-2}	16±2	<0.3	b	s	2.6±0.6
1ν 1μ 1γ X	227	237^{+9}_{-7}	188±6	3.5±0.9	b	<0.8	28±2
1ν 1μ 1e 1j X	7	$6.2^{+1.8}_{-0.7}$	1.3±0.5	b	b	b	<0.3
1ν 1μ 1γ 1j X	25	25^{+5}_{-2}	15±2	<0.9	b	<0.8	3.7±0.8
1ν 2μ X	87	54^{+6}_{-3}	<0.5	<0.3	b	<1.3	46±3
1ν 2μ 1j X	24	21^{+4}_{-2}	b	b	b	<1.3	18±2
1ν 2μ 1γ X	1	$0.1^{+1.2}_{-0.0}$	b	b	b	<0.4	<0.4
1ν 1μ 2j X	603	445^{+30}_{-28}	244±7	17±2	b	<1.6	29±2
1ν 1μ 1e 2j X	5	$2.1^{+1.1}_{-0.1}$	<0.6	b	b	b	b
1ν 1μ 1γ 2j X	6	$3.3^{+2.3}_{-0.5}$	1.4±0.5	b	b	b	<0.9
1ν 2μ 2j X	2	$3.9^{+0.8}_{-0.7}$	b	b	b	b	2.8±0.7
1ν 1μ 3j X	71	54^{+7}_{-4}	12±2	<0.9	b	<0.8	1.3±0.4
1ν 1μ 1e 3j X	1	$0.2^{+0.4}_{-0.0}$	b	b	b	b	b
1ν 1μ 4j X	9	$9.7^{+2.1}_{-0.3}$	<0.9	<0.3	b	b	<0.2

Table 6.2: *Data-MC comparison and first MC-contributions for inclusive event classes.*

inclusive event classes of the muon data: part 2									
Class	Z$\rightarrow \tau\tau$ ($5-15$ GeV)	Z$\rightarrow \tau\tau$ ($15-60$ GeV)	Z$\rightarrow \tau\tau$ ($60-130$ GeV)	WW$\rightarrow$X	ZZ$\rightarrow$X	WZ$\rightarrow$X	$t\bar{t} \rightarrow$1lX	$t\bar{t} \rightarrow$2lX	QCD
$1\nu\ 1\mu$ X	s	1.7±0.8	136±5	104±4	4.6±0.3	21±1	46±1	23±1	1620±620
$1\nu\ 1\mu$ 1j X	b	<1.9	76±4	63±3	2.6±0.2	11±1	45±1	21±1	410±160
$1\nu\ 1\mu$ 1e X	b	<0.4	1.6±0.5	5.0±0.8	<0.1	1.3±0.2	<0.1	4.1±0.2	0.7±0.3
$1\nu\ 1\mu\ 1\gamma$ X	b	<0.4	3.8±0.8	2.0±0.5	0.1±0.0	0.4±0.1	1.3±0.1	1.2±0.1	8.3±3.2
$1\nu\ 1\mu$ 1e 1j X	b	s	0.9±0.4	<0.5	<0.1	<0.1	<0.1	3.8±0.2	0.2±0.1
$1\nu\ 1\mu\ 1\gamma$ 1j X	b	s	1.6±0.5	<0.5	<0.1	<0.1	1.3±0.1	1.1±0.1	2.4±0.9
$1\nu\ 2\mu$ X	b	<0.4	<1.0	1.9±0.5	0.7±0.1	1.2±0.2	<0.0	1.8±0.2	1.9±0.7
$1\nu\ 2\mu$ 1j X	b	b	<1.0	<0.4	0.3±0.1	0.3±0.1	<0.0	1.7±0.1	0.7±0.3
$1\nu\ 2\mu\ 1\gamma$ X	b	b	<0.2	<0.2	<0.0	<0.0	<0.0	0.1±0.0	0.0
$1\nu\ 1\mu$ 2j X	b	<0.8	7.7±1.1	19±2	0.9±0.1	3.4±0.4	42±1	12.8±0.4	70±27
$1\nu\ 1\mu$ 1e 2j X	b	b	<0.4	b	<0.0	<0.0	<0.0	1.9±0.1	0.2±0.1
$1\nu\ 1\mu\ 1\gamma$ 2j X	b	b	<0.6	<0.2	<0.0	<0.1	1.0±0.1	0.7±0.1	0.2±0.1
$1\nu\ 2\mu$ 2j X	b	b	b	b	0.1±0.0	<0.1	<0.0	1.0±0.1	0.0
$1\nu\ 1\mu$ 3j X	b	<0.4	<0.9	0.9±0.4	0.1±0.0	0.3±0.1	28±1	2.2±0.2	9.3±3.6
$1\nu\ 1\mu$ 1e 3j X	b	b	<0.4	b	<0.0	<0.0	<0.0	0.2±0.0	0.0
$1\nu\ 1\mu$ 4j X	b	b	b	<0.4	<0.0	<0.0	8.3±0.3	0.2±0.0	1.2±0.5

Table 6.3: *The rest of MC-contributions for inclusive event classes.*

exclusive event classes of the muon data: part 1							
Class	Data	sum of MC	W$\rightarrow \mu\nu$	W$\rightarrow \tau\nu$	Z$\rightarrow \mu\mu$ ($5-15$ GeV)	Z$\rightarrow \mu\mu$ ($15-60$ GeV)	Z$\rightarrow \mu\mu$ ($60-130$ GeV)
$1\nu\ 1\mu$	76854	77000±500	69100±100	980±20	s	27±3	5520±30
$1\nu\ 1\mu$ 1j	4713	5170±140	4060±30	255±8	b	12±2	369±8
$1\nu\ 1\mu$ 1e	23	24^{+3}_{-2}	15±2	<0.3	b	s	2.3±0.6
$1\nu\ 1\mu\ 1\gamma$	201	211^{+8}_{-7}	173±6	3.0±0.9	b	<0.4	24±2
$1\nu\ 1\mu$ 1e 1j	2	$3.5^{+1.4}_{-0.5}$	0.9±0.4	b	b	b	<0.3
$1\nu\ 1\mu\ 1\gamma$ 1j	19	21^{+4}_{-2}	14±2	<0.9	b	<0.8	3.0±0.7
$1\nu\ 2\mu$	62	32^{+4}_{-2}	<0.5	<0.3	s	<0.4	28±2
$1\nu\ 2\mu$ 1j	22	17^{+4}_{-2}	b	b	b	<1.3	15±2
$1\nu\ 2\mu\ 1\gamma$	1	<1.0	b	b	b	<0.4	<0.4
$1\nu\ 1\mu$ 2j	520	381^{+27}_{-25}	230±7	16±2	b	<1.2	24±2
$1\nu\ 1\mu$ 1e 2j	4	$1.9^{+0.9}_{-0.1}$	<0.6	b	b	b	b
$1\nu\ 1\mu\ 1\gamma$ 2j	6	$2.8^{+2.2}_{-0.5}$	1.4±0.5	b	b	b	<0.9
$1\nu\ 2\mu$ 2j	2	$3.7^{+0.9}_{-0.7}$	b	b	b	b	2.8±0.7
$1\nu\ 1\mu$ 3j	61	43^{+7}_{-4}	11±2	<0.9	b	<0.8	1.3±0.4
$1\nu\ 1\mu$ 1e 3j	1	$0.2^{+0.4}_{-0.0}$	b	b	b	b	b
$1\nu\ 1\mu$ 4j	9	$8.7^{+2.3}_{-0.5}$	<0.9	<0.3	b	b	<0.2

Table 6.4: *Data-MC comparison and first MC-contributions for exclusive event classes.*

exclusive event classes of the muon data: part 2									
Class	Z→ $\tau\tau$ ($5-15$ GeV)	Z→ $\tau\tau$ ($15-60$ GeV)	Z→ $\tau\tau$ ($60-130$ GeV)	WW→X	ZZ→X	WZ→X	$t\bar{t}$ →1lX	$t\bar{t}$ →2lX	QCD
1ν 1μ	s	<0.8	58±3	33±2	1.5±0.2	8.5±0.6	0.3±0.1	0.8±0.1	1210±460
1ν 1μ 1j	b	<1.6	65±3	43±3	1.5±0.2	7.1±0.5	3.3±0.2	5.5±0.3	343±132
1ν 1μ 1e	b	<0.4	0.7±0.3	4.7±0.8	<0.0	0.6±0.2	<0.0	0.3±0.1	0.4±0.2
1ν 1μ 1γ	b	<0.4	2.1±0.6	1.7±0.9	0.1±0.0	0.3±0.1	<0.1	0.1±0.0	6.2±2.4
1ν 1μ 1e 1j	b	s	0.7±0.3	<0.5	<0.1	<0.0	<0.0	1.9±0.1	0.0
1ν 1μ 1γ 1j	b	s	1.3±0.5	<0.5	<0.0	<0.1	0.2±0.0	0.3±0.1	2.2±0.8
1ν 2μ	b	<0.4	<0.2	1.7±0.5	0.4±0.1	0.7±0.2	<0.0	0.2±0.0	1.2±0.5
1ν 2μ 1j	b	b	<1.0	<0.4	0.2±0.1	0.2±0.1	<0.0	0.6±0.1	0.7±0.3
1ν 2μ 1γ	b	b	b	<0.2	<0.0	<0.0	<0.0	<0.0	0.0
1ν 1μ 2j	b	<0.8	6.7±1.0	18±2	0.7±0.1	3.0±0.3	13.5±0.4	7.2±0.3	61±24
1ν 1μ 1e 2j	b	b	<0.2	b	<0.0	<0.0	<0.0	1.7±0.1	0.2±0.1
1ν 1μ 1γ 2j	b	b	<0.6	<0.2	<0.0	<0.0	0.6±0.1	0.6±0.1	0.2±0.1
1ν 2μ 2j	b	b	b	b	<0.1	<0.1	<0.0	0.9±0.1	0.0
1ν 1μ 3j	b	<0.4	<0.8	0.8±0.3	<0.1	0.2±0.1	19.1±0.4	1.7±0.1	8.2±3.1
1ν 1μ 1e 3j	b	b	<0.4	b	<0.0	<0.0	<0.0	0.2±0.0	0.0
1ν 1μ 4j	b	b	b	<0.4	<0.0	<0.0	7.5±0.3	0.2±0.0	1.0±0.4

Table 6.5: *The rest of MC-contributions for exclusive event classes.*

6.3.2 Electron Sample

Again, inclusive and exclusive classes show similar behaviour with respect to the data↔Monte Carlo agreement. Nevertheless, the classes with good agreement and the ones with deviations are not the same as in the muon sample, so an independent discussion of Table 6.6 to Table 6.9 is necessary. Again total event numbers in data and Monte Carlo are similar throughout all event classes. However, the deviations are more distinct than in the muon sample. In order to properly quantify the amount of agreement and pinpoint problems in the Monte Carlo simulations, results from the Search Algorithm must be awaited.

What strikes first is an excess of Monte Carlo of $\approx 1\%$ in the single lepton class $1\nu\,1e(+X)$ where the MC-correction-factor is extracted from. If one assumes a systematic error of $\approx 3\%$, this discrepancy is not incompatible with the Standard Model. Considering the different contributions to the sum of all MC, the QCD background is the most promising candidate to explain the deviation. Figure 6.9 shows the electron p_T-distribution of the inclusive event class $1\nu\,1e\,X$ for data and all MC. The excess of Monte Carlo events in the W-peak region can be seen, which is caused primarily by the QCD-background estimated from data. It seems surprising that data and MC do not match in the region where the χ^2-fit is carried out (see Section 6.1). The reader should keep in mind that QCD was not included in the determination of the MC-correction-factor. The tranversal mass of the electron-data in Figure 6.5 justifies this decision: Here one can see considerable contributions of QCD in the W-mass region ($60 - 120$ GeV), which also contribute to the excess seen in Figure 6.9. QCD is not expected to contribute to high transversal masses, so it is more likely that these QCD events have a different explanation. In the extraction of QCD from data, the Monte Carlo samples with the inverted cuts are subtracted from the QCD-sample. It seems as if MC does not describe the data with inverted cuts properly, so more data events with inverted cuts are observed in the W-region than a W-Monte Carlo with inverted cuts can model. These QCD-contributions in the W-region play only a role in the $1\nu\,1e(+X)$ event classes. As the determination of the MC-correction-factor does not include them, and as the Search Algoritm excludes the W-region, this slight discrepancy in the general data and MC comparison should not be a problem for the rest of the analysis.

Unlike the muon sample, event classes with photons ($1\nu\,1e\,1\gamma(+X)$, $1\nu\,1e\,1\gamma\,1j(+X)$, $1\nu\,1e\,1\gamma\,2j(+X)$) show a persistent deficit of MC. Based on the numbers from the muon sample, it was concluded that the radiation of inital- and final state gammas is well understood and simulated by PYTHIA. The problem seems to be located elsewhere. As the only difference between photons and electrons is the veto on a trackmatch (see Section 4.3.2), it is possible that this discriminating cut is not well simulated in Monte Carlo. In real data, the probability of a genuine electron being misidentified as a photon is larger than the probability assumed by the detector simulation, i.e. the trackmatch efficiency is lower in the data than in MC. In principle, this efficiency difference should be adjusted by the MC-correction-factor, but slight differences seem to persist which cause the observed deviation in these specific event classes. Poorly measured $Z \rightarrow ee$ events are candidates for such a scenario. One of the electrons can be misidentified as a photon, and a poor p_T measurement can fake MET. In particular, non-fiducial EM candidates (energy deposition

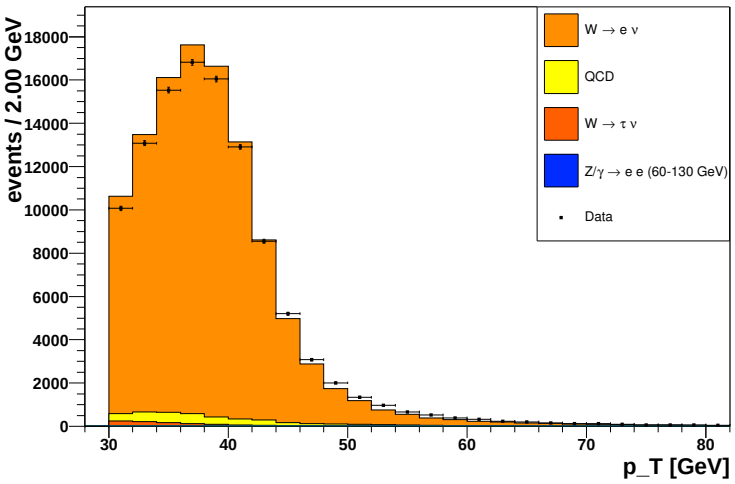

Figure 6.9: *p_T-distribution of the inclusive event class $1\nu\,1eX$ for data and all MC cnotributions.*

in the calorimeter cracks), which are not excluded in this analysis, could cause the excess observed in the data, as their energy cannot be determined accurately and as the matching between calorimeter entry and track is more complicated. Looking at the details of the $1\nu\,1e\,1\gamma$ class, one finds that 24 out of the 125 data events contain gammas which head for the regions of the calorimter cracks (non-fiducial EM-candidates). This hypothesis is supported by the fact that the most recent Photon ID excludes non-fiducial EM candidates. Figure 6.10 shows the angle φ between the electron and the gamma in the $1\nu\,1e\,1\gamma$ event class; only W-MC is considered as it dominates. In the plot one can see the excess of data in the back-to-back region, arguing for Z-events with one electron misidentified as a photon. In addition to this, more data than MC are observed in the small angle region. Here one could imagine a scenario of $W + jet$, where the jet is misidentified as a photon. It seems as if the simulation of photons by Monte Carlo simulation itself works, but the misidentification probabilities of a jet or an electron are too optimistic in MC.

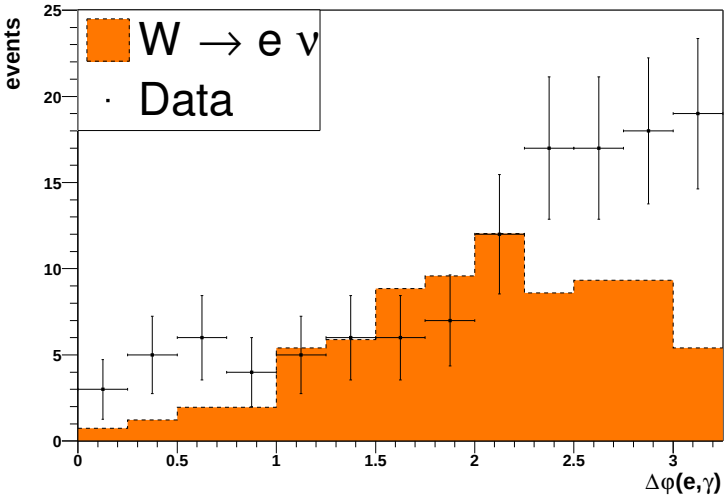

Figure 6.10: *Phi between the electrons and the gamma shown for the $1\nu\,1e\,1\gamma$ exclusive event class, data and generated W-events in comparison.*

Just like the muon sample, the electron skim shows a large excess of data in the $1\nu\,2e(+X)$ event class. This fact is very important as a signal of non-SM particles would be expected in both lepton samples, assuming lepton universality. On the other hand, a similar detector effect could also cause the discrepany in the both lepton samples, and must be discussed. The reason for this deviation is believed to be similar to the muon case. Z-events in which one electron is poorly measured could result in a MET passing the 30 GeV cut. In addition to this, jets with a large electromagnetic component ($\pi^0 \rightarrow \gamma\gamma$) which are misidentified as electrons can contribute to this "exotic" event class. This could also explain the fact that the deviation in the electron sample is even bigger than the one in the muon data. In Figure 6.11, $\Delta\varphi$ between the two electrons can be seen. At large angles (back-to-back) an enormous excess of data stands out. In the muon sample (Fig. 6.8) at least some MC-events contribute, whereas the $Z \rightarrow ee$ MC does not model the data properly. The real data seem to be not understood well, as some uncertainties in the electron energy measurement seem to be not simulated in MC.

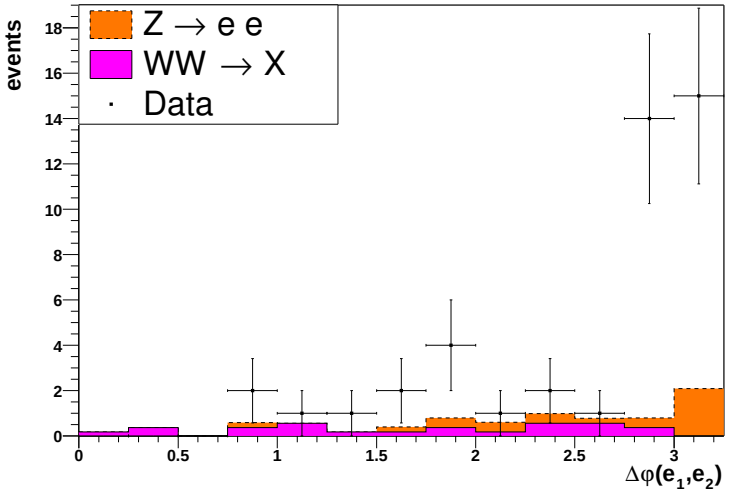

Figure 6.11: *Phi between the two electrons shown for the $1\nu\,2e$ exclusive event class, data and Z-MC + WW-MC in comparison.*

Problems with the simulation of gluon radiation also manifest themselves in the electron sample. In contrast to the muon skim, the $1\nu\,1e\,1j(+X)$ event classes in data and MC agree well. Only a slight excess of MC can be observed in the exclusive class, showing similar behaviour to the overestimation of single gluon radiation in the muon MC. Figure 6.12 shows the exclusive jet class $1\nu\,1e\,1j$ of data and generated $W \rightarrow e\nu$ events. This plot corresponds to Figure 6.7 of the muon sample, with the difference that in the electron case, the η cut has not been tightened. Again the W-MC already fills the major part of the area below the data points. Looking at Table 6.8 one can see that the only relevant contribution missing is QCD. This fact could explain why in the electron case data and MC are in better agreement. In both samples gluon radiation seems to be overestimated as both W-MC show huge contributions. In the muon sample several other MC-processes also contribute to the event class ($W \rightarrow \tau\nu$, $Z \rightarrow \mu\mu$). If we assume an overestimated single gluon radiation in MC, then all these contributions will be overrated, resulting in the observed excess. In the electron case, QCD dominates next to $W \rightarrow e\nu$ and the MC excess is reduced.

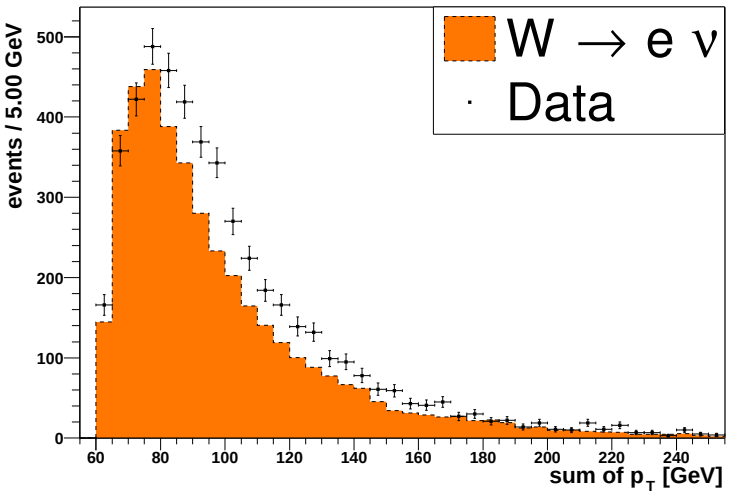

Figure 6.12: *Exclusive event class $1\nu\,1e\,1j$ of data and simulated $W \to e\nu$ events. In this plot all jets with $|\eta| < 1.5$ are considered.*

The event classes with multiple jets ($1\nu\,1e\,2j(+X)$ and $1\nu\,1e\,3j(+X)$) again show the excess of data previously discussed in the muon skim. PYTHIA is not able to describe high jet-multiplicities properly, but again large systematic uncertainties (e.g. JES) also contribute to this deviation. For the four and five jet classes, data and MC agree well.

In the rest of the event classes, data and Monte Carlo simulation are well within the error ranges. The exotic ones with both an electron and a muon show a continuous excess of data, but as statistics are very low, none of them is cause for concern.

inclusive event classes of the electron data: part 1							
Class	Data	sum of MC	W$\rightarrow e\nu$	W$\rightarrow \tau\nu$	Z$\rightarrow ee$ ($5-15$ GeV)	Z$\rightarrow ee$ ($15-60$ GeV)	Z$\rightarrow ee$ ($60-130$ GeV)
1ν 1e X	109126	110600±300	106000±200	768±16	s	s	136±5
1ν 1e 1j X	5704	5660±90	4370±30	104±6	b	b	96±5
1ν 1e 1γ X	169	119^{+7}_{-5}	87±5	1.6±0.7	b	b	5.9±1.1
1ν 1e 1γ 1j X	43	22^{+4}_{-2}	7.1±1.3	<0.4	b	b	3.6±0.9
1ν 2e X	68	23^{+4}_{-2}	<0.8	<0.3	b	b	13±2
1ν 2e 1j X	24	13^{+3}_{-1}	<0.5	b	b	b	9.1±1.3
1ν 2e 1γ X	1	<0.8	<0.3	b	b	b	<0.2
1ν 1e 2j X	718	537±15	281±8	8.8±1.7	b	b	14±2
1ν 1e 1γ 2j X	11	$2.7^{+1.7}_{-0.5}$	<0.6	b	b	b	1.0±0.5
1ν 2e 2j X	5	$1.6^{+1.5}_{-0.1}$	b	b	b	b	<1.1
1ν 1e 3j X	98	69^{+4}_{-2}	12±2	<1.1	b	b	1.3±0.5
1ν 1e 4j X	14	14^{+2}_{-1}	<0.6	b	b	b	<0.5
1ν 1e 5j X	2	$0.8^{+0.2}_{-0.1}$	b	b	b	b	b
1ν 1μ 1e X	15	11^{+2}_{-1}	<0.2	<0.3	b	b	<0.2
1ν 1μ 1e 1j X	6	$5.0^{+0.9}_{-0.2}$	b	b	b	b	b
1ν 1μ 1e 2j X	6	$1.6^{+0.2}_{-0.1}$	b	b	b	b	b
1ν 1μ 1e 3j X	1	$0.1^{+0.0}_{-0.0}$	b	b	b	b	b

Table 6.6: *data-MC comparison and first MC-contributions for inclusive event classes*

inclusive event classes of the electron data: part 2									
Class	Z→ $\tau\tau$ ($5-15$ GeV)	Z→ $\tau\tau$ ($15-60$ GeV)	Z→ $\tau\tau$ ($60-130$ GeV)	WW→X	ZZ→X	WZ→X	$t\bar{t}$ →1lX	$t\bar{t}$ →2lX	QCD
1ν 1e X	s	<0.5	79±4	125±5	3.1±0.3	22±1	65±1	32±1	3350±300
1ν 1e 1j X	b	b	29±2	67±4	1.1±0.2	10±1	64±1	30±1	880±80
1ν 1e 1γ X	b	b	<1.2	2.3±0.7	0.4±0.1	<0.3	1.3±0.1	1.9±0.2	19±2
1ν 1e 1γ 1j X	b	b	<1.0	<0.4	0.1±0.1	<0.1	1.2±0.1	1.7±0.2	7.9±0.7
1ν 2e X	b	b	<0.8	3.9±0.8	0.9±0.1	1.7±0.3	0.1±0.0	3.5±0.2	0.0
1ν 2e 1j X	b	b	<0.6	<0.6	0.3±0.1	0.4±0.1	0.1±0.0	3.1±0.2	0.0
1ν 2e 1γ X	b	b	b	<0.2	<0.0	<0.0	<0.0	<0.1	0.0
1ν 1e 2j X	b	b	5.9±1.1	23±2	0.4±0.1	3.6±0.4	57±1	17±1	127±12
1ν 1e 1γ 2j X	b	b	<0.2	<0.2	<0.1	<0.1	0.9±0.1	0.8±0.1	0.0
1ν 2e 2j X	b	b	b	b	<0.1	<0.1	<0.1	1.6±0.1	0.0
1ν 1e 3j X	b	b	<0.5	1.5±0.5	<0.1	<0.3	36±1	2.5±0.2	16±1
1ν 1e 4j X	b	b	b	<0.2	<0.0	<0.1	10.2±0.4	0.2±0.0	3.2±0.3
1ν 1e 5j X	b	b	b	b	<0.0	<0.1	0.8±0.1	<0.0	0.0
1ν 1μ 1e X	b	b	<0.4	3.9±0.8	<0.1	0.7±0.2	<0.1	3.6±0.2	3.2±0.3
1ν 1μ 1e 1j X	b	b	b	<0.4	<0.1	<0.1	<0.1	3.4±0.2	1.6±0.1
1ν 1μ 1e 2j X	b	b	b	b	<0.0	<0.0	<0.1	1.6±0.1	0.0
1ν 1μ 1e 3j X	b	b	b	b	<0.0	<0.0	<0.0	0.1±0.0	0.0

Table 6.7: *rest of MC-contributions for inclusive event classes*

exclusive event classes of the electron data: part 1							
Class	Data	sum of MC	W$\rightarrow e\nu$	W$\rightarrow \tau\nu$	Z$\rightarrow ee$ ($5-15$ GeV)	Z$\rightarrow ee$ ($15-60$ GeV)	Z$\rightarrow ee$ ($60-130$ GeV)
1ν 1e	103244	104800±300	101600±200	663±15	s	s	33±3
1ν 1e 1j	4935	5090±80	4080±30	95±6	b	b	70±4
1ν 1e 1γ	125	98^{+6}_{-5}	80±4	1.6±0.7	b	b	2.3±0.7
1ν 1e 1γ 1j	32	18^{+4}_{-2}	6.9±1.3	<0.4	b	b	2.5±0.7
1ν 2e	43	$9.4^{+2.4}_{-1.2}$	<0.5	<0.3	b	b	3.9±0.8
1ν 2e 1j	19	10^{+3}_{-1}	<0.5	b	b	b	8.3±1.2
1ν 2e 1γ	1	<0.7	<0.3	b	b	b	<0.2
1ν 1e 2j	599	461±13	268±8	8.2±1.6	b	b	11±2
1ν 1e 1γ 2j	11	$2.5^{+1.7}_{-0.5}$	<0.6	b	b	b	1.0±0.5
1ν 2e 2j	5	$1.4^{+1.5}_{-0.1}$	b	b	b	b	<1.1
1ν 1e 3j	83	54^{+4}_{-2}	12±2	<1.1	b	b	1.0±0.5
1ν 1e 4j	12	$12.7^{+1.8}_{-0.4}$	<0.6	b	b	b	<0.5
1ν 1e 5j	2	$0.7^{+0.2}_{-0.1}$	b	b	b	b	b
1ν 1μ 1e	9	$6.0^{+1.9}_{-0.8}$	<0.2	<0.3	b	b	<0.2
1ν 1μ 1e 2j	5	$1.5^{+0.3}_{-0.1}$	b	b	b	b	b
1ν 1μ 1e 3j	1	0.1±0.0	b	b	b	b	b

Table 6.8: *data-MC comparison and first MC-contributions for exclusive event classes*

exclusive event classes of the electron data: part 2									
Class	Z→ $\tau\tau$ ($5-15$ GeV)	Z→ $\tau\tau$ ($15-60$ GeV)	Z→ $\tau\tau$ ($60-130$ GeV)	WW→X	ZZ→X	WZ→X	$t\bar{t}$ →1lX	$t\bar{t}$ →2lX	QCD
1ν 1e	s	<0.5	49±3	47±3	1.0±0.1	9.1±0.7	0.6±0.1	1.1±0.1	2460±220
1ν 1e 1j	b	b	22±2	44±3	0.4±0.1	6.3±0.6	6.5±0.3	8.2±0.3	750±70
1ν 1e 1γ	b	b	<0.5	2.1±0.6	0.2±0.1	<0.3	0.1±0.0	0.2±0.1	11±1
1ν 1e 1γ 1j	b	b	<1.0	<0.4	<0.1	<0.1	0.3±0.1	0.8±0.1	7.9±0.7
1ν 2e	b	b	<0.4	3.5±0.8	0.6±0.1	1.0±0.2	<0.0	0.4±0.1	0.0
1ν 2e 1j	b	b	<0.6	<0.6	0.2±0.1	0.3±0.1	<0.0	1.5±0.1	0.0
1ν 2e 1γ	b	b	b	<0.2	<0.0	<0.0	<0.0	<0.0	0.0
1ν 1e 2j	b	b	5.7±1.1	21±2	0.2±0.1	3.3±0.4	21±1	9.9±0.4	112±10
1ν 1e 1γ 2j	b	b	<0.2	<0.2	<0.1	<0.1	0.7±0.1	0.8±0.1	0.0
1ν 2e 2j	b	b	b	b	<0.1	<0.1	<0.1	1.4±0.1	0.0
1ν 1e 3j	b	b	<0.5	1.5±0.5	<0.1	<0.3	25±1	2.1±0.2	13±1.1
1ν 1e 4j	b	b	b	<0.2	<0.0	<0.1	9.3±0.3	0.2±0.0	3.2±0.3
1ν 1e 5j	b	b	b	b	<0.0	<0.1	0.7±0.1	<0.0	0.0
1ν 1μ 1e	b	b	<0.4	3.7±0.8	<0.0	0.5±0.1	<0.0	0.2±0.1	1.6±0.1
1ν 1μ 1e 2j	b	b	b	<0.2	<0.0	<0.0	<0.0	1.5±0.1	0.0
1ν 1μ 1e 3j	b	b	b	b	<0.0	<0.0	<0.0	0.1±0.0	0.0

Table 6.9: *rest of MC-contributions for exclusive event classes*

Chapter 7

Search Algorithm

In Chapter 6, the general agreement between data and Monte Carlo simulation (MC) was examined. The Search Algorithm presented in this chapter represents a more detailed data-MC comparison and is the actual core of this analysis. Its principles are based on the procedures of a Model Independent Search performed at the H1 experiment (see [19]). This algorithm can be divided into two separate steps: First, the distributions are systematically scanned for the greatest deviation, then, in the next step, the statistical significance of this deviation is evaluated in order to separate discoveries from fluctuations typical for the specific event class.

7.1 Region of Interest

As stated above, this part of the algorithm has the task of identifying for every event class the region of greatest deviation between data and MC in the variable $\sum p_T$ or the MET distribution.

The word "region" must be put in context here. Every distribution is characterized by its binning. A region is no more than a *connected combination of bins*. The data-MC comparison is executed for all possible regions of the distribution, so, for example, bin 3 is examined, bin $3-4$, bin $3-5$, ..., bin 3–last bin. The algorithm is thus sensitive to deviations present in a single bin as well as to a continuous excess of data or MC in a wide combination of bins. A narrow resonance would result in a deviation present in only a few bins; a signal spread over a large energy region could be identified by analyzing a combination of numerous bins. This reflects the model independence as super massive X-particles with extremely short lifetimes are given the same attention as more stable non-Standard Model particles. Thus, the Search Algorithm can detect deficit regions, excess regions and single outstanding events. The task is to identify any possible deviation from the SM.
As the amount of possible bin combinations decreases with a broader binning, [5 GeV] bins where chosen for all distributions in order to speed up the algorithm. Compared to the momentum resolution of the central tracker for muons of $\Delta(p_T)/p_T^2 \approx 0.002\ \mathrm{GeV}^{-1}$ and of the calorimeter for electrons of $\Delta(p_T)/p_T \approx 0.04$ (see Section 2.2), this binning should

offer sufficient sensitivity. A binning much finer than the error of the p_T measurement does not make sense.

After defining all relevant combinations of bins, the actual strategy for comparison must be identified. For a single region, two numbers have to be compared: The amount of data in the region N_{data} and the sum of all MC contributions in the same region N_{MC}. The extent of deviation is determined by calculating a *p-value*:

p-value: The probability for a given expectation value N_{MC} to measure a fluctuation greater than the one observed with N_{data} when repeating the experiment.

As the number of events observed in data can be too small to assume a Gaussian distribution, Poisson statistics are applied. Given a Poisson distribution with the expectation value N_{MC}, the p-value is simply the area of this graph from the point N_{data} onwards, defined by:

$$p = \begin{cases} \sum_{i=N_{data}}^{\infty} \frac{\exp(-N_{MC})(N_{MC})^i}{i!} & \text{if} \quad N_{data} \geq N_{MC} \\ \sum_{i=0}^{N_{data}} \frac{\exp(-N_{MC})(N_{MC})^i}{i!} & \text{if} \quad N_{data} < N_{MC} \end{cases} . \tag{7.1}$$

This is the probability for N_{MC} to fluctuate up to N_{data} and further, given the statistical error of the measurement.

In this analysis the calculation of the p-value is modified by including the statistical error of the Monte Carlo mean value N_{MC} itself and the systematic uncertainties of the measurement. The incorporation of these errors is accomplished by considering a probability density function made up by a "convolution" between the Poisson distribution and a Gaussian of the width δN_{MC}. All errors are interpreted as uncertainties in the expectation value N_{MC}; data are fixed numbers without any errors. In this way the final *estimator* p for each considered region, based on the procedure of the H1 analysis [19], is defined by:

$$p = \begin{cases} \sum_{i=N_{data}}^{\infty} A \cdot \int_0^{\infty} \mathrm{d}b \, \exp\left(\frac{-(b-N_{MC})^2}{2(\delta N_{MC})^2}\right) \cdot \frac{e^{-b}\, b^i}{i!} & \text{if} \quad N_{data} \geq N_{MC} \\ \sum_{i=0}^{N_{data}} A \cdot \int_0^{\infty} \mathrm{d}b \, \exp\left(\frac{-(b-N_{MC})^2}{2(\delta N_{MC})^2}\right) \cdot \frac{e^{-b}\, b^i}{i!} & \text{if} \quad N_{data} < N_{MC} \end{cases} . \tag{7.2}$$

The constant factor A ensures the normalization of the probability density function to unity.
In this computation N_{MC} is assumed to be the *true* value, the one nature actually realizes. Of course simulated Monte Carlo samples are also only the result of generating a variable numerous times with a given probability density function. Only if the MC-statistics is

infinite, N_{MC} can become the true value. As we have only limited MC-statistics, there is an intrinsic MC statistical error for a specific MC i which contributes to δN_{MC}:

$$\sigma_{stat} = \sqrt{N_{MC_i}(unscaled)} \times \frac{\mathcal{L}_{data}}{\mathcal{L}_{MC_i}} \times f_{lepton} = \sqrt{N_{MC_i}} \times \sqrt{\frac{\mathcal{L}_{data}}{\mathcal{L}_{MC_i}}} \times \sqrt{f_{lepton}} \,. \quad (7.3)$$

Here f_{lepton} denotes the MC-correction-factor and N_{MC} are the actual bin entries of each distribution. Besides this, all systematic errors contribute to δN_{MC}. In the detector simulation and the MC-smearing, many assumptions have to be made, and many systematic uncertainties remain. These systematic uncertainties assume that the true value is Gaussian distributed around N_{MC}. All these errors are assumed to be uncorrelated and are added in quadrature. The individual contributions will be discussed in detail in Section 8. Ultimately this is expressed as:

$$\delta N_{MC} = \sqrt{\sigma_{stat}^2 + \sum_i \sigma_i(syst)^2} \,. \quad (7.4)$$

The true value N_{MC} is expected to be Gaussian distributed. This assumption is very delicate for small unscaled MC-contributions ($N_{MC}(unscaled) < 5$) as here Poisson's law must be applied. The problem is that the algorithm also deals with single bin regions where (especially in the high-p_T tail) zero MC contributions are very common. As only limited Monte Carlo statistics are available, it is not unlikely that some or all MC do not contribute at all, or very little, to a certain event class or a certain region.
A solution similar to the overall event class comparison in Section 6.3 is selected. If in one simuated Monte Carlo process the contribution is only $N_{MC}(unscaled) < 5$, then the 68% confidence level upper limit is taken (see Table 6.1). These upper limits are then considered a new expectation value where a Gaussian distribution can be assumed. As in the case of the overall event class comparison, it is only for some MC-contributions that this upper confidence level is considered. If a MC is not expected to contribute to the total event class, it is also not expected to contribute to any region of a distribution (see "s" and "b" in Tables 6.2-6.9). In addition to this, the upper confidence level is only used as a bin entry if there is a data contribution in the same bin. Otherwise, upper confidence levels would be summed up in regions where neither data nor MC are expected, resulting in unphysical MC excess in some regions. For the special case $N_{MC}(unscaled) = 0$ the upper estimate is applied only if there is not a single Monte Carlo or QCD event contributing to the bin. If some MC contribute to the bin entry, then the sum of these MC is regarded as the expectation value of the bin and the data are compared to this sum. In this case no confidence levels for $N_{MC}(unscaled) = 0$ are applied.
For the QCD-background estimated from data, again no upper confidence levels are applied and only the total error of the QCD-scaling-factor is included in the algorithm.

The computation of the p-value can be illustrated with Figure 7.1. The black curve represents the Gaussian distribution of the true value N_{MC} with its width δN_{MC}. The blue curve is the Poisson distribution for a specific value b of the convolution. The integral of this blue curve from N_{data} onwards is the probability of the fluctuation when N_{data} is

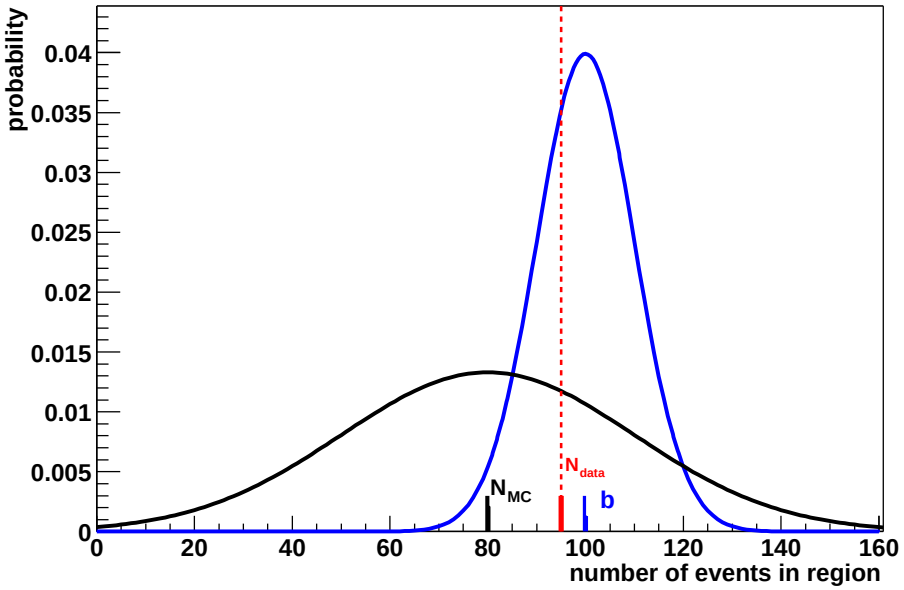

Figure 7.1: *Schematic view illustrating the computation of the statistical estimator p.*

measured and b is expected. In this case this integral is very large, which shows that the p-value rises with the width δN_{MC}. If N_{MC} is itself uncertain because of systematic errors, then a small difference between N_{MC} and N_{data} is very common, and the p-value is close to 0.5.

After the p-value is defined it is calculated for all possible regions in all event class distributions. The *Region of Interest* is simply the region with the smallest p-value p_{min}^{data}. This region represents the greatest deviation of data from the SM prediction in the specific event class and its MET or $\sum p_T$ distribution. If the p-value is very small, this deviation cannot be explained by statistical or systematic errors, and the region is a potential candidate for physics beyond the Standard Model. As this p-value is a convolution of a Gaussian and a Poisson distribution, its value cannot be directly translated into terms of sigmas corresponding to the deviation.
It has to be stressed that a small p-value itself does not suffice to identify a region of possible New Physics. The p_{min} obtained is related to a *single specific region* in the distribution. It estimates how probable a deviation is in the specific region. Besides this, the significance of this deviation has to be tested by considering how probable such a deviation is in any of the connected regions. This is what the second step of the algorithm evaluates.

7.2 General Significance of the Deviation

7.2.1 Definition of the Estimator

As stated before, the p-value itself only identifies the region of greatest deviation. It computes the probability of the observed fluctuation for the specific Region of Interest. In order to determine whether the fluctuation is a signal of New Physics, one would expect that this deviation is *localized* to this specific region and can be observed at the same position when the experiment is repeated. To quantify the significance of the deviation, it is necessary to calculate the probability $\widetilde{P}$ of observing a deviation with a p-value smaller than p_{min}^{data} in *any region* of the distribution.
If the systematic errors of a certain event class are very large, then small p-values are very likely to appear at different regions when the experiment is repeated several times. Then the deviation would not be significant. If a real signal of New Physics is present in one event class, and the deviation is well outside the error bands, then a p-value as small as observed in the specific data region will be found in no other region but the original Region of Interest when the experiment is repeated.

Of course, the repetition of the entire experiment is only a "thought experiment". What can be done is to simulate numerous measurements based on the SM prediction by *generating Hypothetical Data Histograms* (HDH). For every bin of a distribution the true value is given by N_{MC}. Hypothetical data are generated, i.e. "diced", by using the probability density function defined in Equation 7.2. The exact procedure of dicing requires further discussion (see Section 7.2.2).

For every bin hypothetical data are generated according to the statistical and systematical errors, so a real measurement is simulated, and a very large number of Hypothetical Data Histograms can be created (limited only by the CPU-power). All these distributions are now compared to the original true value, N_{MC}, by repeating step 1 of the algorithm and identifying the Region of Interest of the diced data, i.e. computing p_{min}^{SM}. The estimator of the General Significance of the Deviation per event class is defined as:

$$\widetilde{P} = \frac{\text{number of HDH with } p_{min}^{SM} \leq p_{min}^{data}}{\text{total number of HDH}} \; . \tag{7.5}$$

This estimator provides information about how likely a deviation as large as the one observed in data is for the specific event class, assuming the Standard Model. It thus characterizes the general properties of the event class. In a class with enormous systematic errors and few statistics, deviations anywhere in the distribution are very common. Nevertheless it may well be possible that all these deviations present at different locations of the distribution are much smaller than the one observed in the localized region of the data-MC comparison. In this case the p_{min}^{data} value is significant, and $\widetilde{P}$ is very small.

7.2.2 The Principle of Dicing Data

Testing the significance of the deviation is done by dicing Hypothetical Data Histograms, i.e. one changes the true value N_{MC} slightly to reflect the inherent statistical and sytematic errors. As all error contributions are included, this is the best way to simulate a real measurement. Of course, the assumption that these errors are both well understood and realistic is crucial for this procedure.

The basis for the dicing is the probability density function in Equation 7.2, even though the actual dicing process is divided into several parts with respect to all error contributions. It is essential that contributions which are *statistically independent* can be *decoupled* and diced seperately. There are three main dicing-contributions for each Hypothetical Data Histogram:

- assumed systematic errors as part of the Gaussian convolution
- statistical error of MC as part of the Gaussian convolution
- Poisson probability to account for the actual measurement

The systematic errors are the most complicated aspect as they consist of numerous different factors. They will be discussed in more detail in Section 8. In principle, these different factors are assumed to be statistically independent, and for each contribution a temporary correction factor is generated according to a Gaussian with the width of the corresponding relative error. This means that for a contribution with a relative error of 5% one HDH could be created with a temporary correction of 7% (e.g. if a cross section has a 5% uncertainty, one HDS could have a cross section 7% lower than the mean value). Great care has to be taken as the *bins* are *correlated* for all systematic contributions; for some of them even the different simulated Monte Carlo processes are correlated.
The best example is the Jet Energy Scale: If the JES simulated in MC is 2% too low, then all jets lose the same 2% of energy, so all bins and all MC are correlated.
As stated above, only statistically independent contributions are diced seperately, so for one Hypothetical Data Histograms there will be a single random number multiplied by all jet-energies of all MC, diced only *once*.

The statistical error of the MC can be simulated much easier as all MC and all bins are uncorrelated. For every bin, the statistical errors of the MC processes i can be added quadratic resulting in $\sigma^2(\sum_i N_{MC_i}) = \sum_i \left(N_{MC_i} \times \frac{\mathcal{L}_{data}}{\mathcal{L}_{MC_i}} \times f_{lepton} \right)$, see Equation 7.3. As the bins are uncorrelated, for each of them a new bin entry is generated using a Gaussian with the width $\sigma(\sum_i N_{MC_i})$, where bin_{old} is the bin entry already corrected for sytematic uncertainties:

$$bin_{new} = Gauss\left(bin_{old},\ \sigma(\sum_i N_{MC_i}) \right) \ . \tag{7.6}$$

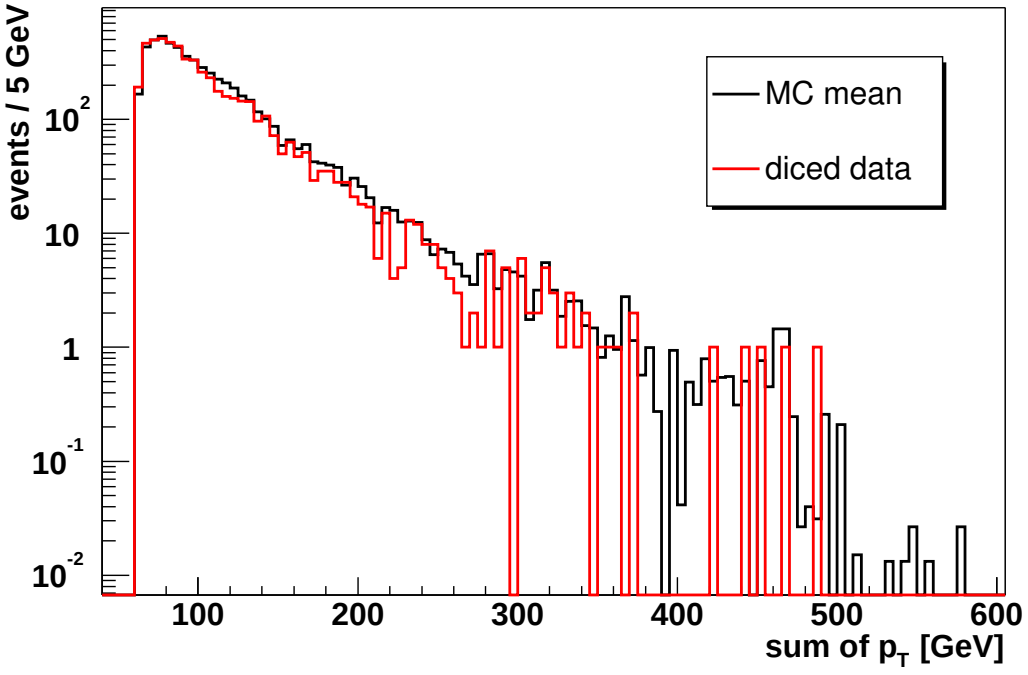

Figure 7.2: *Comparison between MC mean value and diced data for* $1\nu\,1e\,1j$ *inclusive event class.*

Similarly, the Poisson probability simulating the actual measurement is applied for all bin entries separately. The "input" is the bin entry after all previous corrections have been completed; the "output" is the new bin entry which will be compared to the true value N_{MC} using the p-value.

Figure 7.2 shows the comparison between the MC mean distribution of the $1\nu\,1e\,1j$ inclusive event class and a single Hyphothetical Data Histogramm, created by using the dicing mechanism described above. This mechanism creates only integer numbers for the diced data, which can be seen in the high-p_T tail of the distribution where only some bins contain single diced data. These two histograms would now be passed over to the Region of Interest algorithm to spot the region of greatest deviation. Even with simple inspection one can easily observe differences between the MC mean and the diced data, differences similar to the ones observed in the real data.

Chapter 8

Systematic Uncertainties

This chapter discusses all systematic errors implemented in the Search Algorithm. The size of each contribution is estimated. This is crucial, as too conservative errors affect the sensitivity of the algorithm, whereas errors neglected by mistake lead to unrealsitic deviations mimicking New Physics. All separate contributions discussed herein are assumed to be statistically independent. Each error must be investigated for correlations, and a technique to implement the error in the Search Algorithm must be developed.
As stated before, the integration of systematic errors in the statistical estimator p has been inspired by the H1 analysis (see [19]). The technical implementation of these errors in the Search Algorithm has been developed independently. This chapter gives a detailed prescription and could provide valuable information for future Model Independent Searches.

8.1 MC-Correction-Factor

One of the major systematic uncertainties of this analysis relates to the scaling of simulated Monte Carlo samples (MC) to the data using calibration regions. The advantage of this procedure is that the systematic errors of numerous efficiency determinations and the luminosity error drop out.
The technique of estimating this error was presented in Section 6.1.2. What is still missing is the discussion of correlations. As this MC-correction-factor is determined by comparing the sum of all MC to the data in a specific region (W/Z-peak), and as this factor is used for all event classes, it is obvious that all bins and all Monte Carlos are correlated with respect to this error. If the real correction-factor is 2% lower than the one extracted from data, than all MC and all bin entries share this 2% error.
For a specific region defined by the bins i, by the contributing MC j and by the resulting bin content N_i^j, assuming a relative error of the correction-factor σ_{rel}, the total error σ_f caused by the MC-correction-factor can be computed as:

$$\sigma_f^2 = \sigma_{rel}^2 \cdot \left(\sum_{bin\ i} \sum_{MC\ j} N_i^j \right)^2 . \tag{8.1}$$

For the Region of Interest this error is computed for every region individually and added in quadrature to the overall error of the Gaussian in Equation 7.2.

In the case of dicing data, for each Hypothetical Data Histogram a random number g_f is diced once using a Gaussian with the width σ_{rel} and a mean 1. In this way a new MC-correction-factor f_{new} active for a single HDH is found. As all bins and all MC are correlated, each bin entry consisting of the sum of all MC contributions without any efficiency correction is multiplied by this new factor to form the new "diced" bin entry:

$$f_{new} = f_{mean} \cdot g_f \ , \tag{8.2}$$

$$N_i^j = f_{new} \cdot N_i^j \ . \tag{8.3}$$

8.2 MC Cross Section Uncertainty

Before the MC are scaled to data using specific calibration regions, there is another implicit scaling uncertainty, as each MC is scaled to data luminosity before any MC-correction-factor is applied. This scale factor depends on the cross section of the individual MC process. The uncertainty in the cross section (uncertainties in the theoretical computations+error of pdf) is assumed to be statistically independent from the MC-correction-factor. Of course, an inaccurate W-cross section also affects the MC-correction-factor, but it is believed that other systematic uncertainties in the procedure of scaling MC in calibration regions dominate. These uncertainties are best estimated by consulting two different kinematic regions (W-peak and Z-peak). To be conservative, the cross section error and the MC-correction-factor error are both regarded and treated separately.

With respect to the cross section error, all bins again are correlated. In contrast to the MC-correction-factor, only some Monte Carlo samples are statistically dependent, because an error in the $t\bar{t}$ cross section has no impact on $W \rightarrow \mu\nu$. An example of two processes which are correlated is W/Z. The computation of their cross sections using similar Feynman diagrams links these two processes.
The following cross section systematic errors are applied in the algorithm:

- $\sigma_{rel}(W/Z) = 3.5\%$, taken from [49]
- $\sigma_{rel}(t\bar{t}) = 5.0\%$, taken from [57]
- $\sigma_{rel}(WW/ZZ/WZ) = 4.0\%$, taken from [50] when different pdf's are regarded

For the Region of Interest the computation of the error is identical to Equation 8.1 with the minor modification that the sum over the MC j includes only the correlated ones (e.g. WW/ZZ/WZ).

When dicing data, three random numbers are generated once, using a Gaussian with the individual width σ_{rel} and a mean 1. All bins of each MC-contribution N_i^j(scaled to luminosity) are then multiplied by the corresponding factor.

8.3 QCD-Background

As in the scaling procedure of the MC-correction-factor, the determination of the QCD-scaling-factor is subjected to systematic uncertainties which have already been evaluated in Section 6.2. As the QCD-scaling-factor is applied for the whole event class, all bins are correlated, and the implementation in the Search Algorithm is identical to the previously discussed systematic errors. For each region the error contribution is:

$$\sigma_{QCD}^2 = \sigma_{rel}^2 \left(\sum_{bin\ i} N_i^{QCD} \right)^2 . \tag{8.4}$$

For the Hypothetical Data Histograms, again a random number is produced once with respect to σ_{rel} of electrons or muons. Then all QCD bin entries are multiplied by this factor, and the new bin entry is added to the already modified sum of MC events.
The relative error for the electron skim amounts $\sigma_{rel} = 9\%$, whereas the uncertainty in the muon case with $\sigma_{rel} = 38\%$ is much greater as the QCD contamination is much smaller, and the scaling procedure is more inaccurate.

8.4 Energy Scale

The energy scale was discussed and motivated in Section 5.3 , and the parametrizations for muons, electrons and jets have also been presented. Each energy scale factor is also characterized by an uncertainty:

- $\sigma_{rel}(muon) = 0.3\%$
- $\sigma_{rel}(electron) = 0.1\%$ for fiducial (no energy deposition in φ-cracks)
- $\sigma_{rel}(electron) = 1.2\%$ for non-fiducial
- $\sigma_{rel}(jet) = 6.0\%$

The error of the Jet Energy Scale obviously dominates and is one of the most important uncertainties in this analysis.
All statistically independent energy scale factors α correct the p_T of the corresponding objects, so the error can be implemented easily using error propagation:

$$\sum p_t = \alpha_e \cdot p_T(electron) + \alpha_{jet} \cdot p_T(jet) , \tag{8.5}$$

$$\Longrightarrow \sigma^2 \left(\sum p_t \right) = p_T^2(electron) \cdot \sigma^2(\alpha_e) + p_T^2(jet) \cdot \sigma^2(\alpha_{jet}) . \tag{8.6}$$

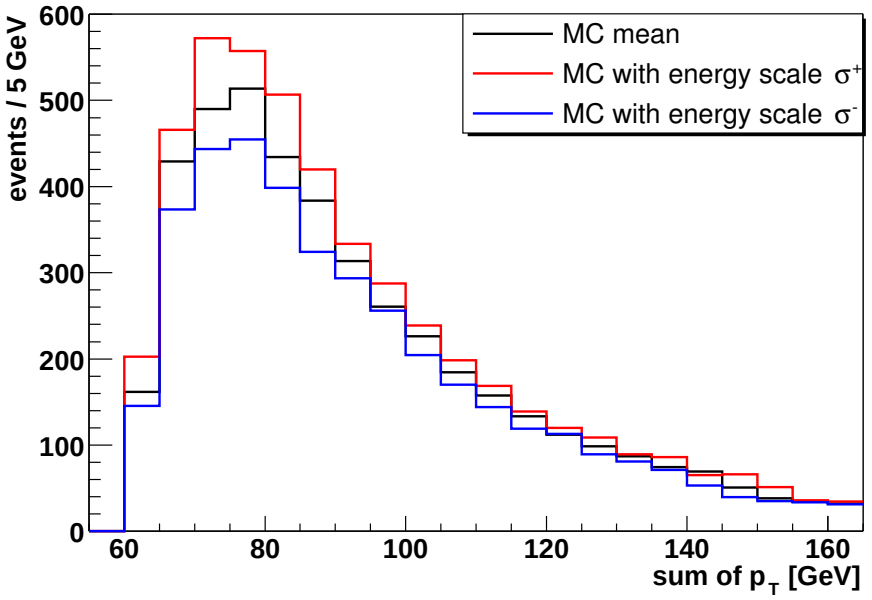

Figure 8.1: *$1\nu\,1e\,1j$ exclusive class of simulated $W \rightarrow e\nu$ events, unmodified MC and MC with energy scale errors are plotted*

The error in the energy scale affects the momentum of each particle in each event. This information cannot be used in the algorithm as only bin entries $N_i^j \pm \sigma(N_i^j)$ can be implemented. In order to estimate the effect of the energy scale errors, the $\sum p_t$ of each event is raised and lowered by one sigma:

$$\sum p_t{}' = \sum p_t \pm \sigma\left(\sum p_t\right) \ . \tag{8.7}$$

As all bins are correlated, each MC sample is processed once with the error added for all bins ond once with the error subtracted. In this way one gets two new histograms for each event class defining the sigma bands of the energy scale. This is illustrated in Figure 8.1, where the $1\nu\,1e\,1j$ exclusive class of the simulated $W \rightarrow e\nu$ MC events is shown. In addition to the MC mean, the two histograms containing the up- and down-energy scale errors are plotted. In this event class the large JES-error dominates and leads to considerable differences in the bin entries. In the case of the up error all jets gain 6% of energy, so many more events fullfill the selection cuts and contribute to this class. The down error is the exact opposite as fewer events are seen. Because of the fact that the muon p_T is added to the jet energy, the entire distribution is not only shifted to either side, but all bins lose or gain events.

The energy scale error can be included in the Region of Interest, bearing in mind that all bins and all MC are statistically dependent. If, for example, the JES is underestimated compared to the "real" value, then all MC and all bins are affected by this error in the same direction. For a specific region with i bins and j MC-contributions, the error in the positive direction is defined by:

$$\sigma_{scale}(+) = \left| \sum_{bin\ i} \sum_{MC\ j} N_i^j(\sigma^+) - \sum_{bin\ i} \sum_{MC\ j} N_i^j(\text{mean}) \right| \ . \tag{8.8}$$

Here $N_i^j(\sigma^+)$ is the bin entry of the histogram with the energy scale raised, $N_i^j(\text{mean})$ is the corresponding bin entry of the unmodified MC. A second error, $\sigma^2_{scale}(-)$, can be determined using the histogram with the negative direction $N_i^j(\sigma^-)$. As the algorithm requires symmetrical errors, the maximum of these two σ_{scale} is used for each region.

For generating Hypothetical Data Histograms special care has to be taken that the direction of the energy scale error is preserved for all bins. For each HDH a random number is diced once using a Gaussian with mean= 0 and $\sigma = 1$. The sign of this number fixes the direction of the energy scale error (up or down) in this specific turn of dicing data, the absolute value determinines the magnitude of the error.
Assuming a random number +0.5, the error can be determined for each bin i separately by computing

$$\sigma_{scale} = \sum_{MC\ j} N_i^j(\sigma^+) - \sum_{MC\ j} N_i^j(\text{mean}) \ . \tag{8.9}$$

The new bin entry is then defined by:

$$bin_{new} = \left(\sum_{MC\ j} N_i^j(\text{mean}) \right) + 0.5 \cdot \sigma_{scale} \ . \tag{8.10}$$

In this way direction and magnitude of the error are preserved for all bins and all MC-contributions.

8.5 Smearing

In Section 5.3 the need for smearing events simulated by Monte Carlo was introduced and the smearing functions for the different physical objects were stated. All of these smearing parametrizations are determined using fit procedures. Of course these smearing parameters inherit certain errors, which represent a sytematic error in the analysis. If, for example, the smearing parameters are underestimated, then peaks in MC will tend to be more narrow than in the data.
In order to estimate the possible contributions of this error, the smearing parameters for muons and electrons are varied at one sigma. For jets the smearing error can be neglected as the JES uncertainty dominates. MET is not oversmeared in this analysis, as in principle the lepton smearing propagates to the Missing Transverse Energy. As stated before this holds in the muon case, but there are large deviations in MET distributions with electrons. As the physical reasons for this disagreement are not well understood, this analysis does not smear MET from the first. The smeared electron MET is only used to estimate a contribution to the systematic error of an MET distribution.

In principle the variation of the smearing parameters should go in both directions. In practice this is technically very time-consuming, so a simpler solution is chosen: The smeared Monte Carlo mean distributions (E' for electrons) are further smeared (E'') to simulate a lepton smearing parameter $\sigma_{\text{up}} = \sigma_{\text{mean}} +$ error. As the width of two added

Gaussians is the quadratic sum of both individual widths, this can be incorporated by additionally smearing the electrons and muons with a Gaussian of the width σ_{add}:

$$
\begin{aligned}
E' &= E + E \cdot Gauss(0, \sigma_{\text{mean}}) && (8.11)\\
E'' &= E + E \cdot Gauss(0, \sigma_{\text{up}}) && (8.12)\\
&= E' + E' \cdot Gauss(0, \sigma_{\text{add}}) && (8.13)\\
&= E + E \cdot Gauss(0, \sigma_{\text{mean}}) + E \cdot Gauss(0, \sigma_{\text{add}}) + && (8.14)\\
&\quad \underbrace{E \cdot Gauss(0, \sigma_{\text{mean}}) \cdot Gauss(0, \sigma_{\text{add}})}_{\approx 0} \\
\Longrightarrow \sigma^2_{\text{add}} &= \sigma^2_{\text{up}} - \sigma^2_{\text{mean}} \,. && (8.15)
\end{aligned}
$$

Independent of this, the electron MET is smeared using Equation 5.7. Ultimately this results in two histograms for muons (mean and add-smeared), and three histograms for electrons (mean, add-smeared and MET-smeared). Implementing the error in the Search Algorithm is similar to the energy scale case. Again all bins and all MC are correlated as the smearing parameters hold for all leptons and all MET energy ranges of all MC.

For the retrieval of the Region of Interest the smearing error is computed for each region seperately:

$$
\sigma^2_{\text{smear}} = \left(\sum_{bin\ i} \sum_{MC\ j} N_i^j(\text{add-smeared}) - \sum_{bin\ i} \sum_{MC\ j} N_i^j(\text{mean}) \right)^2 . \qquad (8.16)
$$

The error of MET in the electron case can be computed in the same way.

Also in the test of General Significance, special care has to be taken to preserve the direction of the smearing error throughout the whole distribution as all bins are statistically dependent. In contrast to the energy scale, the down error is not available. To account for possibly smaller smearing parameters, the error determined from the difference between the add-smeared histogram and MC mean histogram is simply symmetrised. Again, a random number g can be diced once using a Gaussian with mean= 0 and $\sigma = 1$, where the sign fixes the direction of the smearing error (up or down) and the absolute value determinines the magnitude of the error.
For each bin a new entry can then be generated using:

$$
bin_{new} = \left(\sum_{MC\ j} N_i^j(\text{mean}) \right) + g \cdot \left(\sum_{MC\ j} N_i^j(\text{add-smeared}) - \sum_{MC\ j} N_i^j(\text{mean}) \right) . \qquad (8.17)
$$

8.6 Summary of the Different Contributions

Some final remarks concerning the dominant contributions to the systematic uncertainty should be made.
As most event classes are very different from each other, the systematic error varies much from one class to another. In the case of the multiplicative factors, the magnitudes are easy to estimate, e.g for a single muon class : MC-correction-factor $\sigma_{rel} \approx 3\%$, MC cross section uncertainty $\sigma_{rel} \approx 3.5\%$ as W-production domiantes and error on QCD-background $\sigma_{rel} \approx 40\%$.
The other two systematic uncertainties (energy scale and smearing) are more difficult to figure as they shift the whole distribution. Again, the dominate contribution strongly depends on the specific event class: In a class with jets, the smearing errors can be neglected with respect to the Jet Energy Scale uncertainty; in a MET distribution with electrons, the MET-smearing will be the dominant error, next to the scale factor uncertainties.
To conclude, the systematic uncertainties are different for every event class and for every Region of Interest. In the event classes with huge statistics, $1l(+X)$ and $1l\,1j(+X)$, all systematic errors combined are $\lesssim 10\%$. In the exotic classes and the high-p_T tails of the distributions, they can also reach values of $\approx 50\%$, due to the fact that here smearing and energy scale uncertainty redistribute the bin entries (e.g. unsmeared bin entry: $N_{MC} = 4$, smeared bin entry: $N_{MC} = 2$).

Chapter 9

Results and Interpretation

Before the event classes with the largest and most interesting deviations are discussed and the general agreement between data and Monte Carlo simulations (MC) is examined, the output of the Search Algorithm is explained in more detail.
As mentioned in Section 7, only the combination of a Region of Interest (p) and the General Significance ($\widetilde{P}$) does provide meaningful information on the agreement in each event class. Looking at the p-values in Tables 9.1 to 9.4 one can see that the majority of the event classes has very small p-values. As a comparison, a one sigma deviation corresponds to a pure Poisson p-value (no convolution with Gaussian systematic error) of $p = 0.17$. A three sigma deviation in terms of Poisson probability corresponds to $p < 0.002$. In the Search Algorithm this pure Poisson probability is convoluted with a Gaussian to account for systematic errors. This convolution only increases the p-value, so that a comparison with a pure Poisson p-value should provide a good reference point.
The meaning of the $\widetilde{P}$-value can be illustrated using the following "thought experiment": Suppose one can turn on New Physics and measure a peak in a specific p_T-distribution, corresponding to a $p \ll 0.002$. Then one turns off New Physics again and repeats the experiment several times, only assuming the Standard Model; this is what the generation of Hypothetaical Data Histograms (HDH) actually is. If one observes a three or more sigma effect ($p \ll 0.002$) caused by SM only in about ten percent of the hypothetical experiments ($\widetilde{P} \approx 10\%$), then one would not claim the discovery of New Physics in the first experiment. This should be kept in mind when reviewing the $\widetilde{P}$-values of the event classes. In addition to their values, the number of generated Hypothetical Data Histograms is also given; e.g. $\widetilde{P} = \frac{0}{500} < 0.002$ where no smaller p-value was found in 500 HDHs, so an upper limit is given.
A final general remark should be made regarding the predominant location of the Region of Interest: Most of the bin combinations with the greatest deviation in inclusive and exclusive event classes contain only a few events and a small MC mean value ($N < 5$), thus Poisson statistics have to be applied. Most of the event classes show discrepancies in the low statistic regions of the p_T- or MET-spectrum, often in the high-p_T tails. This underlines the importance of the accurate implementation of upper limits and Poisson statistics in the Search Algorithm (see Section 7.1). Only in this way signals of New Physics with small cross sections can be detected.

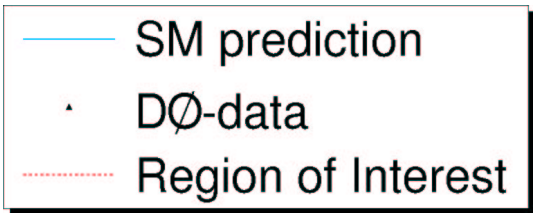

Figure 9.1: *Legend for Figures 9.2 to 9.12.*

Tables 9.1 to 9.4 summarize the results of the Search Algorithm for all event classes, Figures 9.2 to 9.12 show the distributions with significant deviations, the legend in Figure 9.1 holds for all these plots.

9.1 Muon Sample

In the inclusive sample, 10 out of the 32 event classes show a deviation in p of more than three sigmas; in the exclusive sample, one can find 7 out of the 32 classes with such a statistically significant deviation (see Tables 9.1 to 9.2). This yields that most of the event classes are well within the error ranges of the MC expectation values, even though the p-values seem surprisingly small at first sight. One should also remenber that the algorithm considers all possible regions for each distribution. It is very likely that at least some of these show discrepancies with more than just one sigma.
This conclusion is supported by the results of the General Significance test: Most of the event classes with $0.002 < p < 0.1$, i.e. a moderate deviation, show relatively large $\widetilde{P}$-values (> 0.2). This fact also proves the good performance of the generation of hypothetical data: If the deviation, quanitified by the p-value, is within statistical fluctuations (less than three sigmas), then the significance of the deviation is low and many Hypothetical Data Histograms show similar differences between data and MC. Only event classes with small p-values *and* small $\widetilde{P}$-values will be treated as interesting deviations.
In the following, classes with

$$p < 0.002 \quad \text{and} \quad \widetilde{P} < 0.2 \tag{9.1}$$

are marked as "Classes of Interest" and then discussed in detail.

$1\nu\,1\mu(+X)$ inclusive and exclusive event class:
In this event class the p_T distributions as well as the MET distributions show a highly significant excess of data in the high-p_T tail. The discrepancy is huge with all $p < 10^{-43}$. This discrepancy has not been noticed in previous analyses like W-cross section measurements (e.g. [35]) or Model Independent Searches in Run I (see [18]), which arouses suspicion that a detector effect rather than New Physics causes this deviation. The results of the General Significance test ($\widetilde{P} < 0.004$) indicate that this excess of data can not be explained by any of the assumed systematic or statistical uncertainties. Inclusive as well as exclusive event classes seem to be incompatible with the Standard Model. Figure 9.2 shows the continuous excess of high-p_T events in the data. However, this discrepancy between data

and MC was anticipated in these regions (see Section 4.3.3). The muon p_T measurement is very imprecise for large momenta and Monte Carlo simulation underestimates the mismeasurement of muons. Then there is an irreducible background of large momenta muons from cosmic ray showers. As a consequence, various quality cuts were performed in order to ensure a proper muon momentum measurement. It seems as if these selection cuts do not suffice to suppress this high-p_T background. For example, this analysis does not perform a tight dca-cut (distance-of-closest-approach) which is the best way to identify poor muon p_T measurements. This decision was made in order to remain sensitive to decays of non-SM particles in secondary vertices and in order to increase the statistics of "exotic" event classes. This anlysis is not interested in a precision measurement of the W-peak as other conventional analyses cover this task (see [35]).
Nevertheless, this poor agreement between data and SM-prediction in the high-p_T tail of the W-peak is unsatisfying as MC should describe the systematic errors of the muon measurement properly, even without a dca-cut. High-p_T muons are of major interest in most searches for New Physics and should therefore be well understood. In addition to this, such an excess of data could also indicate the excistence of new massive vector bosons (W'). To conclude, a separate analysis should focus on this particular event class in order to optimize the muon description of the Monte Carlo simulations and in order to search for the existence of W' and other possible new phenomena in DØ Run II data.

$1\nu\,1\mu\,1j(+X)$ inclusive and exclusive event class:
In this event class several distributions show deviations, see Figure 9.3. Exclusive and inclusive MET distributions both show a broad excess of data in the region $100\text{ GeV} < MET \lesssim 300\text{ GeV}$ resulting in $p_{incl.} = 3 \cdot 10^{-7}/\widetilde{P}_{incl.} < 0.003$ and $p_{excl.} = 1 \cdot 10^{-5}/\widetilde{P}_{excl.} < 0.003$. In order to decide whether this is a signal for New Physics, one should examine first if the discrepancy can be explained by problems with the MC. The excess of events with high MET in the data could be caused by mis-measured muons. A measured muon p_T which is much larger than the real p_T results is a considerable contribution to MET, thus leading to the observed excess. MET seems to be not well modeled in MC and the smearing of MET should be investigated further in the future.
The sum of p_T distributions show deviations in different parts of the spectrum. The inclusive class shows a data excess in a single bin at 775 GeV. This is also stressed by the small $\widetilde{P}$-value of 0.02 which proves the significance of this deviation. Two things are important for this distribution: First, the MC entries are not ditributed smootly in this high-momentum region. The reader must be reminded that upper limits are only considered for a bin if data are present, too. Secondly, the fact that in the Region of Interest the MC mean is much smaller than in the other bins could be caused by a non-vanishing MC-contribution in this bin. In this case, the upper limit for $N_{MC}(unscaled) = 0$ is not applied (see Section 7.1). With respect to the low statistics in this distribution and with respect to the momentum resolutions 775 GeV, which are much poorer than a 5 GeV binning, this deviation is well compatible with the Standard Model.
Only in the exclusive p_T distribution the Region of Interest found by the Search Algorithm is located near the data deficit which has already been detected in Section 6.3. The p-value found here is by far the largest one with $p = 0.006$. This shows the good performance of the Search Algorithm: As all systematic uncertainties like the Jet Energy Scale are included, the deviation which attracts the eye most is rated minor significant with $\widetilde{P}_{excl.} = 0.40$.

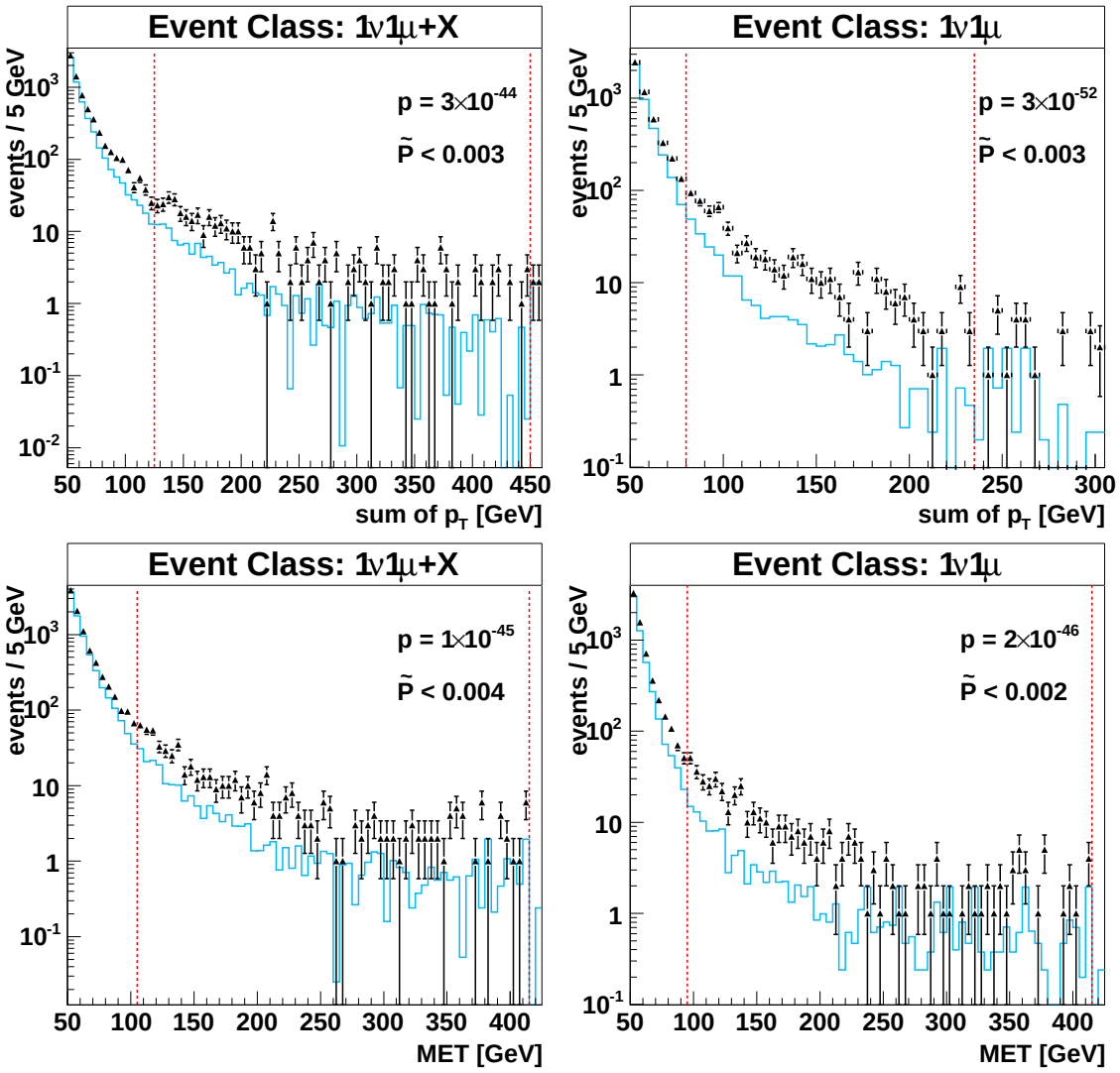

Figure 9.2: *Results of the Search Algorithm in the $1\nu\,1\mu(+X)$ inclusive and exclusive event classes.*

Only the rising flank of the distribution is part of the Region of Interest. This is due to the fact that here the systematic errors are smaller than in a region extended up to 100 GeV. The deviations in the other distributions are much more interesting and should draw the attention of future analyses.

$1\nu\,1\mu\,1\gamma(+X)$ inclusive and exclusive event class:

The inclusive as well as the exclusive p_T-distribution show an excess of MC events in a 15 GeV wide range ($p_{incl.} = 0.001$ and $\widetilde{P}_{incl.} = 0.10$, $p_{excl.} = 0.002$ and $\widetilde{P}_{excl.} = 0.14$). The corresponding distributions can be found in Figure 9.4. One can see a similar behaviour of the exclusive and inclusive class. Both show an excess of MC in the rising flank at low momenta. In the overall data-MC comparison in Section 6.3 it was concluded that data and MC agree well and only the SM processes of the initial- and final state radiation of photons can be seen. These processes seemed to be modeled well by PYTHIA. The Search Algorithm modifies this conclusion as there is a certain discrepancy. Monte Carlo simulation seems to underestimate the energy of the photons as events with a small sum of p_T are

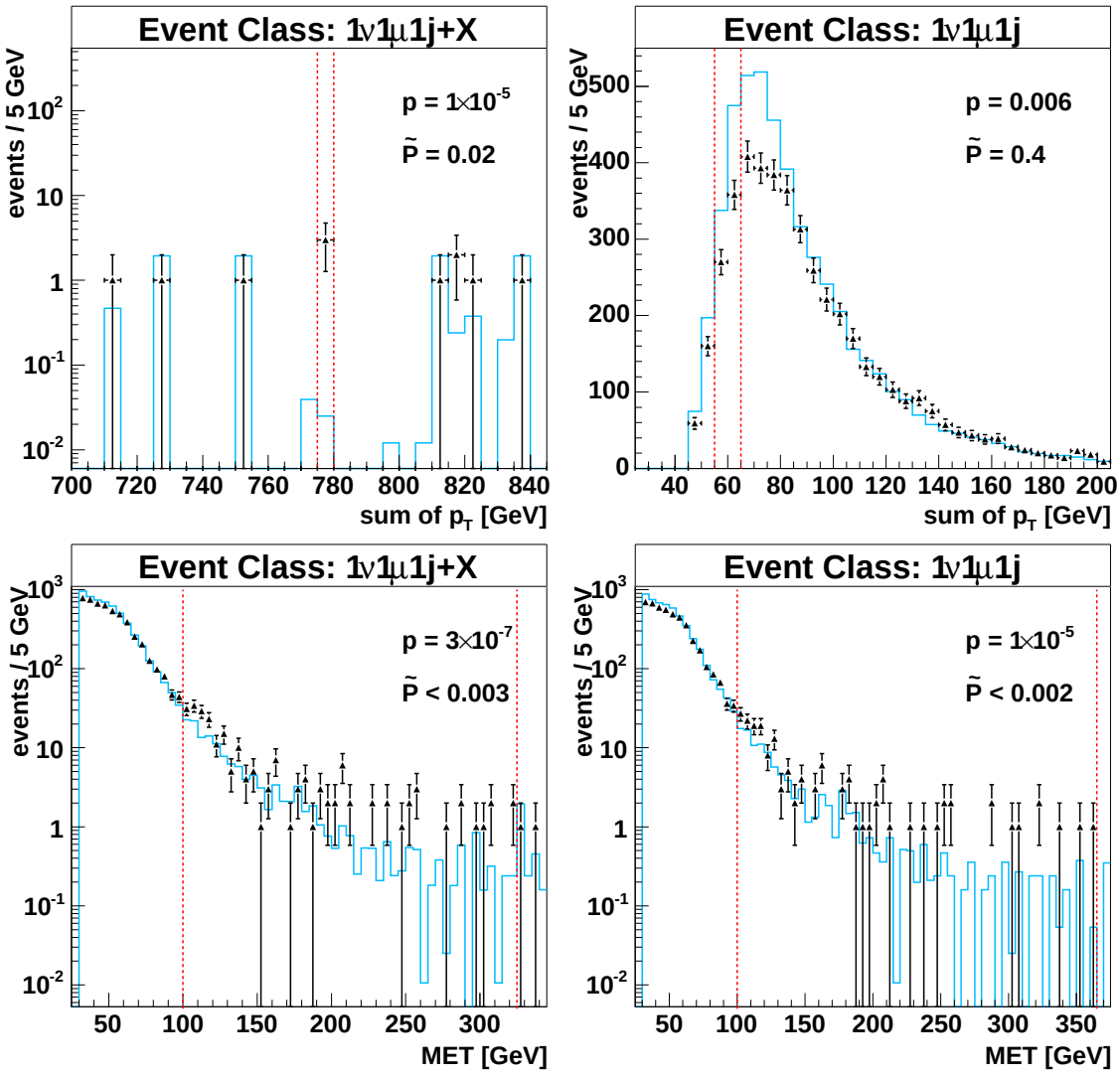

Figure 9.3: *Results of the Search Algorithm in the $1\nu\,1\mu\,1j(+X)$ inclusive and exclusive event classes.*

enhanced, whereas data show more events with larger tranverse momenta. Nevertheless, both General Significance values are above 10%, so the deviation observed in the data can be assessed as a fluctuation, eihter statistical or sytematic. An effect completely incompatible with the Standard Model would have a much smaller $\widetilde{P}$-value, which supports the assumption of MC not describing common SM-processes in every detail.

$1\nu\,2\mu(+X)$ inclusive and exclusive event class:

Consistent with the overall comparison in Section 6.3, a significant deviation is found in all distributions of this class, exclusive as well as inclusive and sum of p_T as well as MET. This deviation could be caused by an unmodeled detector effect or by a signal beyond the Standard Model. Especially the exclusive p_T distribution shows a huge deviation with $p = 1 \cdot 10^{-6}$. This deviation is highly significant as not a single HDH produces a comparable p-value ($\widetilde{P} < 0.001$). The top of Figure 9.5 illustrates this significant deviation in one exclusive and one inclusive distribution. The corresponding other distributions look very similar. In the exclusive as well as in the inclusive p_T-distribution one can see a

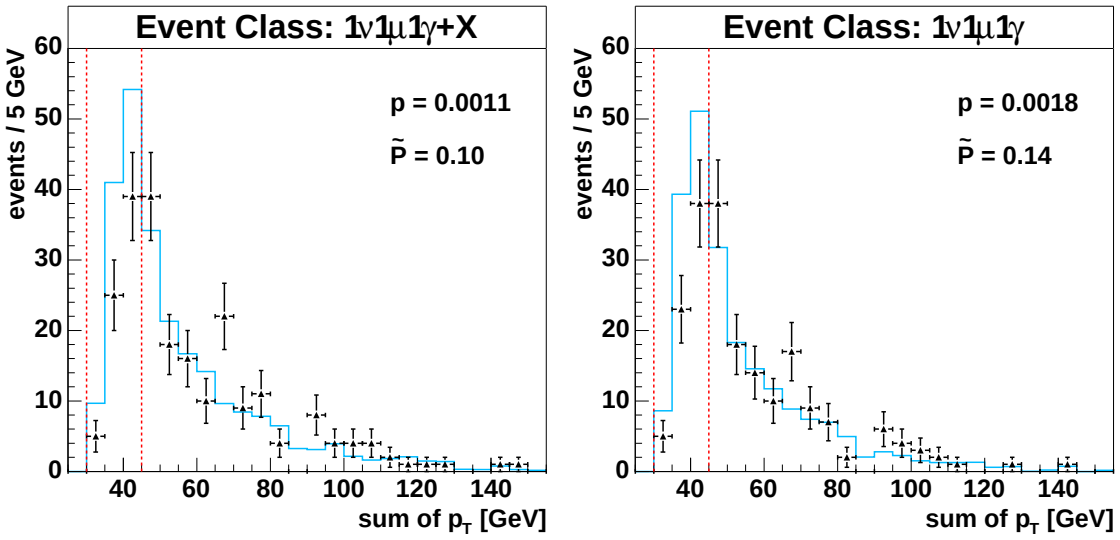

Figure 9.4: *Results of the Search Algorithm in the $1\nu\,1\mu\,1\gamma(+X)$ inclusive and exclusive event classes.*

distinct peak between 80 GeV and 100 GeV. Here again, the excellent performance of the Search Algorithm can be seen in the way this peak in the data is spotted and estimated as a significant deviation.
In order to decide whether this is a detector effect or a trace of New Physics, the MET-plots can provide additional information. Here the Region of Interest is very broad, not favouring a certain amount of MET. It is difficult to relate the excess in the p_T-distribution to any excess in the MET-distribution, there is just a general trend that there are more data than expected in the complete energy range.
In the comparison of the total event numbers the disagreement between data and MC was explained with mis-measured $Z \rightarrow \mu\mu$ events as most muons were more or less back-to-back and as MET was mainly parallel or anti-parallel to the muons. Considering the location of the Region of Interest, one concludes that this is exactly where Z-events would be expected (around 90 GeV). This fact does not match with the assumption of mis-measurement, as a muon momentum too high, which could mimic MET, would result in a sum of p_T outside the Z-range. Only the improbable scenario of one muon being measured with a p_T too low and one with a p_T equally too high could result in a sum of p_T in the Z-region and a MET at the same time. The bottom of Figure 9.5 shows two $\Delta\varphi$ plots of data events in the Region of Interest between 80 GeV and 100 GeV of the $1\nu\,2\mu$ exclusive event class. $\Delta\varphi$ between both muons again shows the typical back-to-back behavior of Z-events, $\Delta\varphi$ between each muon and MET also supports the hypothesis of mis-measured muons. MC contributions are not shown, but their distribution should be similar to Figure 6.8.
A better explanation of the observed phenomena is difficult in the context of this Model Independent Search. MIS identifies and quantifies the deviation. Then an independent analysis should look at the specific event class in detail in order to either claim a discovery or to learn more about detector effects.

$1\nu\,1\mu\,2j(+X)$ inclusive and exclusive event class:
In this event class all inclusive and exclusive distributions show a Region of Interest with a slight excess of data. With $\widetilde{P} = 0.33$ the deviation in the exclusive sum of p_T spectrum is

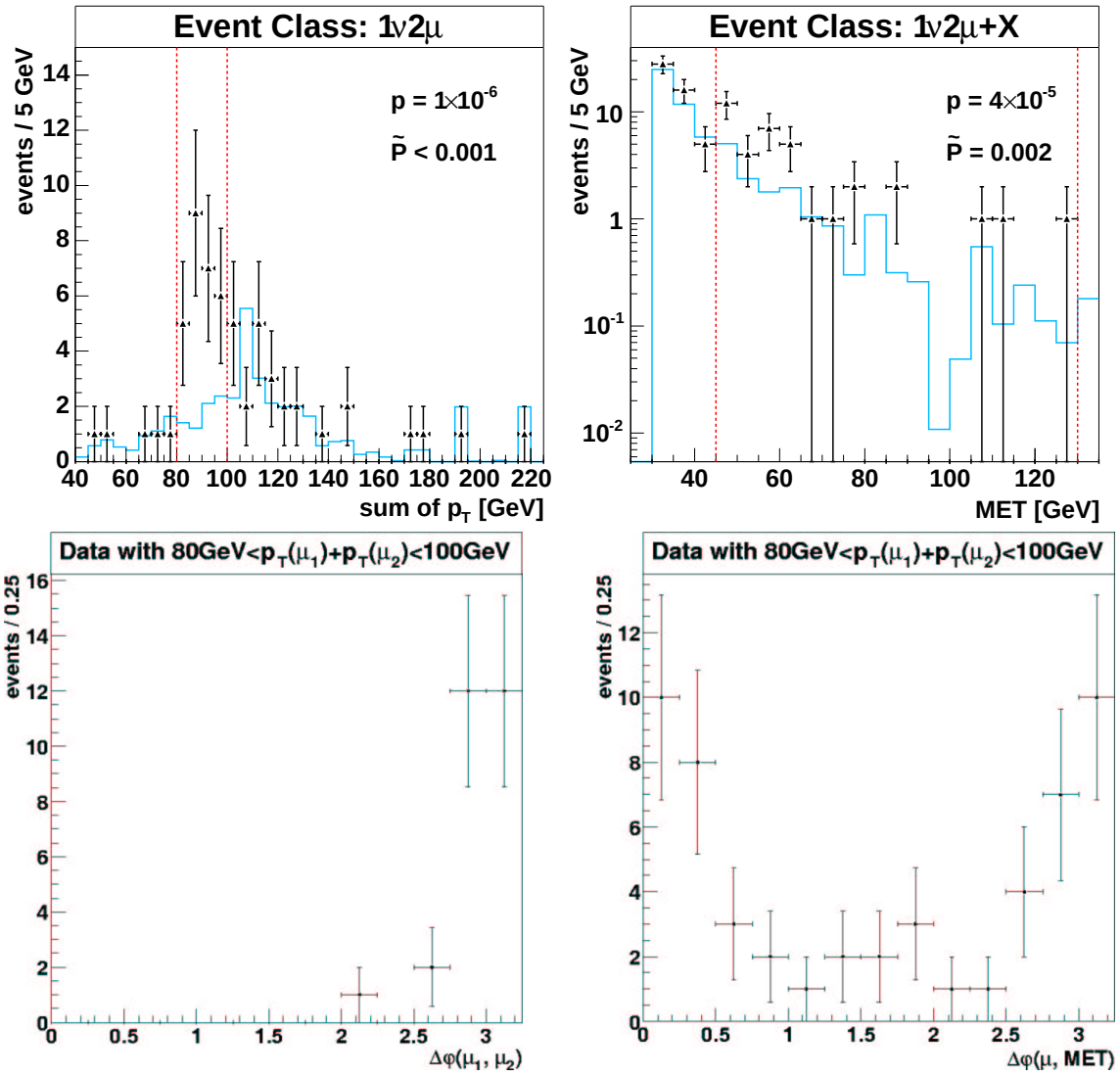

Figure 9.5: *Results of the Search Algorithm in the $1\nu\,2\mu(+X)$ inclusive and exclusive event classes (top) and phi between different objects from the Region of Interest in the $1\nu\,2\mu$ exclusive class (bottom).*

not significant, all other plots are shown in Figure 9.6. What stands out is that all Region of Interest are in the high-p_T tail of the distribution, e.g. $495 - 555$ GeV in the inclusive sum of p_T or $190 - 240$ GeV in the exclusive MET plot. Especially the MET classes show a sigificant deviation with $p_{incl.} = 1 \cdot 10^{-5} / \widetilde{P}_{incl.} = 0.006$ and $p_{excl.} = 5 \cdot 10^{-6} / \widetilde{P}_{excl.} < 0.001$. There are several data points in a region where one expects much less events from Monte Carlo simulations.

As the region of data excess in the sum of p_T distribution is at very large energies, both jets must own much energy ($\gtrsim 100$ GeV). It was mentioned before several times that PYTHIA simulates multiple gluon radiation only incompletely as it underestimates multiplicities higher than one (see Section 6.3). In addition to this, it seems as if PYTHIA also underestimates the transverse momentum of the emitted gluons, leading to an excess of data in the high-p_T region.

The deviation in the MET distributions is more significant and more difficult to explain.

An attempt can be made at this point, further investigations should clarify this discrepancy found by the Search Alsorithm. Again, only a distinct mis-measurement of the muon or of the jet E_T could lead to "fake" MET of this magnitude (≈ 200 GeV). As the selection cuts for both objects (muons and jets) are very tight, a mis-measurement of this amount is very unlikely and thus worrying. The Monte Carlo simulation should model these quality selection cuts properly.

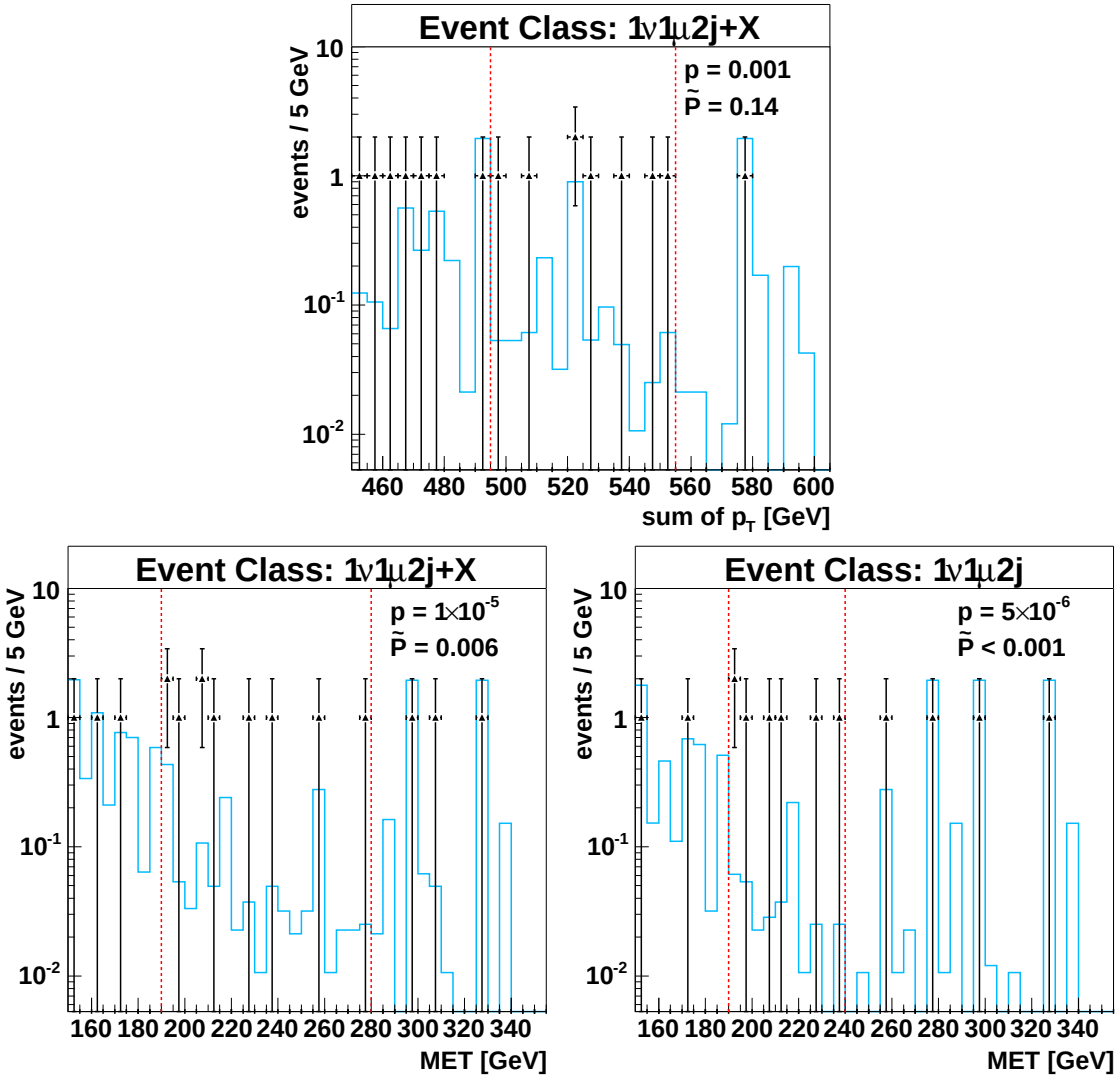

Figure 9.6: *Results of the Search Algorithm in the $1\nu\,1\mu\,2j(+X)$ inclusive and exclusive event classes.*

$1\nu\,1\mu\,1e\,2j(+X)$ inclusive and exclusive event class:

In this event class both inclusive distributions and the exclusive MET distribution show an data excess of three events in a narrow region (≈ 30 GeV), illustrated in Figure 9.7. This excess is very interesting as $t\bar{t} \rightarrow 2l\,2jets$ is the most important contributor in this "exotic" event class. Next to this, QCD is the only Standard Model process able to produce such a combination of particles. SUSY particles are also expected to produce lepton combinations with unlike flavour, i.e. electron and muon, so this event class is promising for the search for New Physics. The p-values of these three distributions are all very low ($p < 0.005$) , but only the inclusive sum of p_T distribution is below the three sigma line. All General

Significances are $\approx 5\%$. These small numbers indicate that the deviation is worth to be studied further. One explanation could be that either the generated $t\bar{t}$ sample is insufficient, or that mismeasured QCD-events are more often in the data than expected. In order to prove the presence of a non-SM signal more statistics are needed.

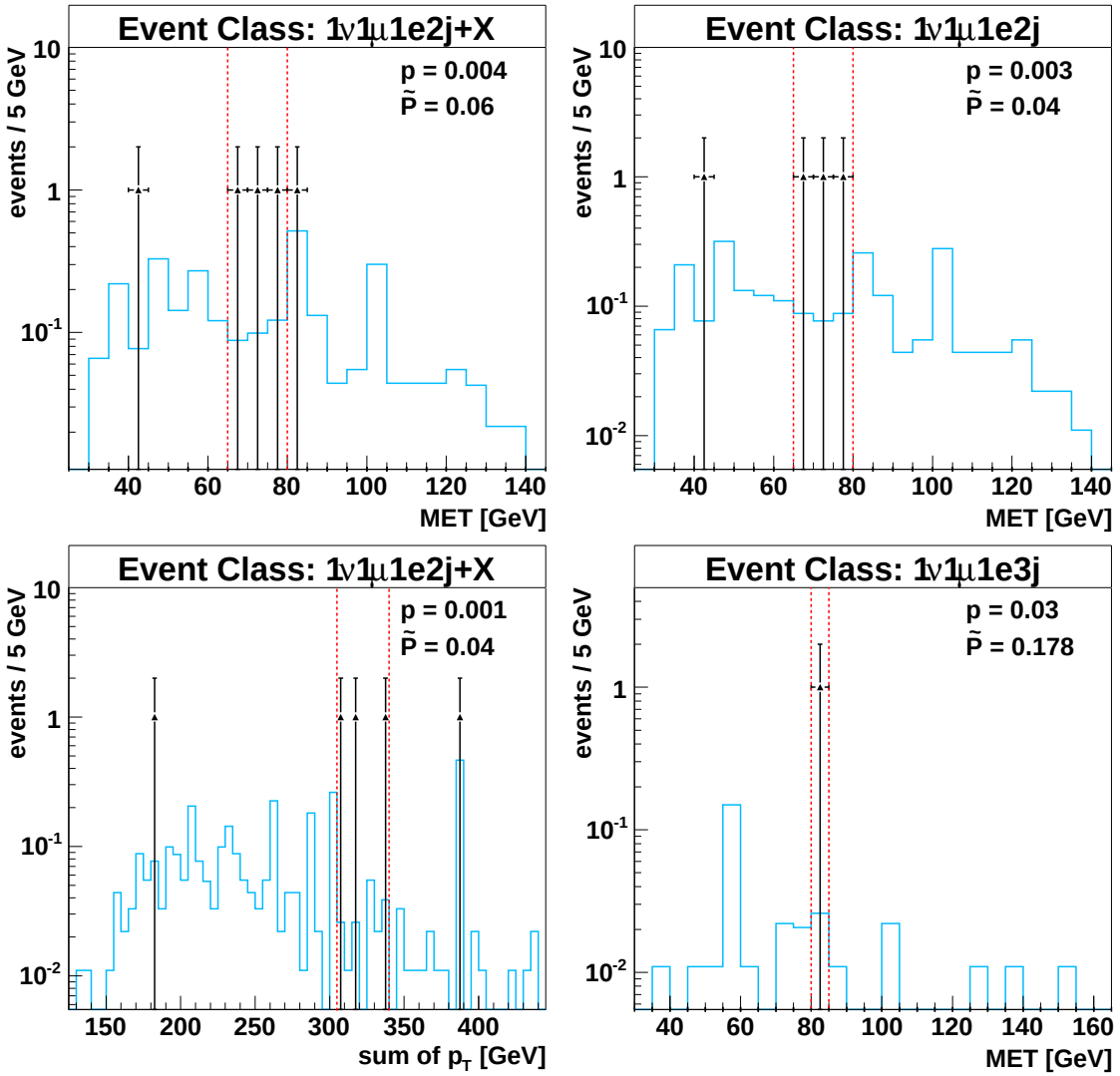

Figure 9.7: *Results of the Search Algorithm in the $1\nu\,1\mu\,1e\,2j(+X)$ inclusive and exclusive event classes and in the $1\nu\,1\mu\,1e\,3j$ exclusive event class.*

$1\nu\,1\mu\,1e\,3j$ exclusive event class:

This event class (see Figure 9.7) corresponds to the interesting $t\bar{t}$ candidate discussed in Section 6.3. The event display can be found in [1]. In the overall event number comparison this single event was not rated as a significant deviation. Because of the lack of statistics, the Search Algorithm agrees with this decision. This illustrates the correct evaluation of the Search Algorithm, insofar as a single event found in one distribution is not a huge and significant deviation. The single data event can be regarded as a rare $t\bar{t} \rightarrow 2l + 3j$ event. This conclusion is supported by the fact that the event actually is one of the official top-candidates (see [56]).

To conclude, 5 event classes with significant deviations are found in the muon data sample:

$$1\nu\,1\mu(+X),\ 1\nu\,1\mu\,1j(+X),\ 1\nu\,2\mu(+X),\ 1\nu\,1\mu\,2j(+X) \text{ and } 1\nu\,1\mu\,1e\,2j(+X)$$

For most of these deviations, problems with the Monte Carlo event generator or the detector simulation could serve as an explanation. In order to provide evidence for the existence of New Physics, further studies are needed to either claim a discovery or to improve the Monte Carlo simulation.
The rest of the 16 event classes agrees with the Standard Model predictions which shows that, in general, the DØ detector is well understood and that Monte Carlo predictions describe most of the data. In the other Model Independent Search analysis of DØ RunII (see [1]), significant deviations are observed in the single muon classes $1\mu + (X)$ and in the di-muon classes $2\mu(+X)$. This result agrees well with this analysis. By requiring MET, additional interesting event classes can be detected.

Results from Search Algorithm for inclusive event classes of muon data							
Class	Dis	p	$\widetilde{P}$	Region [GeV]	N_{data}	N_{MC}	δN_{MC}
1ν 1μ X	$\sum p_T$	$3 \cdot 10^{-44}$	$\frac{0}{300} < 0.003$	125–450	380	125.9	9.8
	MET	$1 \cdot 10^{-45}$	$\frac{0}{260} < 0.004$	105–415	566	210.4	14.9
1ν 1μ 1j X	$\sum p_T$	$1 \cdot 10^{-05}$	$\frac{4}{200} = 0.02$	775–780	3	0.025	0.022
	MET	$3 \cdot 10^{-07}$	$\frac{0}{320} < 0.003$	100–325	221	141.6	8.4
1ν 1μ 1e X	$\sum p_T$	0.0045	$\frac{112}{500} = 0.22$	130–170	4	0.62	0.13
	MET	0.0084	$\frac{127}{500} = 0.25$	75–80	4	0.71	0.23
1ν 1μ 1γ X	$\sum p_T$	0.0011	$\frac{50}{500} = 0.10$	30–45	69	104.8	6.8
	MET	0.0077	$\frac{125}{500} = 0.25$	30–40	97	130.0	8.1
1ν 1μ 1e 1j X	$\sum p_T$	0.014	$\frac{218}{500} = 0.44$	275–295	2	0.17	0.06
	MET	0.059	$\frac{370}{500} = 0.74$	70–80	2	0.39	0.07
1ν 1μ 1γ 1j X	$\sum p_T$	0.0087	$\frac{245}{500} = 0.49$	75–100	2	9.4	1.5
	MET	0.073	$\frac{452}{500} = 0.90$	35–55	9	15.9	2.3
1ν 2μ X	$\sum p_T$	0.0006	$\frac{11}{500} = 0.02$	80–105	36	17.4	2.5
	MET	$4 \cdot 10^{-05}$	$\frac{3}{1500} = 0.002$	45–130	37	16.2	1.8
1ν 2μ 1j X	$\sum p_T$	0.0081	$\frac{131}{500} = 0.26$	60–150	1	8.5	2.0
	MET	0.0082	$\frac{80}{500} = 0.16$	45–60	11	4.2	0.88
1ν 2μ 1γ X	$\sum p_T$	0.58	$\frac{500}{500} = 1.$	190–215	0	0.18	0.71
	MET	0.63	$\frac{500}{500} = 1.$	35–155	1	1.30	0.53
1ν 1μ 2j X	$\sum p_T$	0.001	$\frac{70}{500} = 0.14$	495–555	8	1.6	0.53
	MET	$1 \cdot 10^{-05}$	$\frac{3}{500} = 0.006$	190–280	10	1.5	0.34
1ν 1μ 1e 2j X	$\sum p_T$	0.001	$\frac{35}{1000} = 0.035$	305–340	3	0.18	0.05
	MET	0.0044	$\frac{31}{500} = 0.062$	65–80	3	0.31	0.07
1ν 1μ 1γ 2j X	$\sum p_T$	0.040	$\frac{334}{500} = 0.67$	205–290	5	1.8	0.41
	MET	0.078	$\frac{354}{500} = 0.71$	105–110	1	0.081	0.031
1ν 2μ 2j X	$\sum p_T$	0.11	$\frac{404}{500} = 0.81$	210–510	0	2.4	0.63
	MET	0.093	$\frac{316}{500} = 0.63$	45–50	1	0.098	0.038
1ν 1μ 3j X	$\sum p_T$	0.0094	$\frac{342}{500} = 0.68$	380–385	3	0.32	0.22
	MET	0.020	$\frac{341}{500} = 0.68$	75–105	4	11.1	1.4
1ν 1μ 1e 3j X	$\sum p_T$	0.23	$\frac{268}{500} = 0.54$	350-355	1	0.25	0.17
	MET	0.028	$\frac{114}{500} = 0.23$	80–85	1	0.026	0.017
1ν 1μ 4j X	$\sum p_T$	0.038	$\frac{443}{500} = 0.89$	240-285	6	2.3	0.45
	MET	0.038	$\frac{335}{500} = 0.67$	75–160	0	3.5	0.62

Table 9.1: *Summary of results for inclusive muon classes*

Results from Search Algorithm for exclusive event classes of muon data							
Class	Dis	p	$\widetilde{P}$	Region [GeV]	N_{data}	N_{MC}	δN_{MC}
1ν 1μ	$\sum p_T$	$3 \cdot 10^{-52}$	$\frac{0}{390} < 0.003$	80–235	596	205.2	15.4
	MET	$2 \cdot 10^{-46}$	$\frac{0}{430} < 0.002$	95–415	447	132.5	13.2
1ν 1μ 1j	$\sum p_T$	0.0056	$\frac{198}{500} = 0.40$	55–65	628	812.6	67.7
	MET	$1 \cdot 10^{-05}$	$\frac{0}{300} < 0.003$	100–365	164	106.6	6.9
1ν 1μ 1e	$\sum p_T$	0.080	$\frac{468}{500} = 0.94$	45–60	2	6.0	1.0
	MET	0.018	$\frac{157}{500} = 0.31$	75–80	3	0.47	0.22
1ν 1μ 1γ	$\sum p_T$	0.0018	$\frac{71}{500} = 0.14$	30–45	66	99.0	6.5
	MET	0.0096	$\frac{132}{500} = 0.26$	40–45	65	45.5	3.7
1ν 1μ 1e 1j	$\sum p_T$	0.18	$\frac{494}{500} = 0.99$	170–175	1	0.06	0.23
	MET	0.053	$\frac{298}{500} = 0.60$	35–145	0	3.1	0.5
1ν 1μ 1γ 1j	$\sum p_T$	0.0097	$\frac{195}{500} = 0.39$	70–100	3	11.1	1.7
	MET	0.031	$\frac{253}{500} = 0.51$	35–55	6	13.9	2.2
1ν 2μ	$\sum p_T$	$1 \cdot 10^{-06}$	$\frac{0}{1000} < 0.001$	80–100	27	7.1	1.5
	MET	$1 \cdot 10^{-04}$	$\frac{3}{1000} = 0.003$	45–130	23	8.1	1.4
1ν 2μ 1j	$\sum p_T$	0.012	$\frac{170}{500} = 0.34$	60–150	1	7.5	1.6
	MET	0.017	$\frac{158}{500} = 0.32$	45–60	10	4.1	0.9
1ν 2μ 1γ	$\sum p_T$	0.56	$\frac{500}{500} = 1.$	150–155	1	0.87	0.46
	MET	0.56	$\frac{500}{500} = 1.$	20–45	1	0.87	0.46
1ν 1μ 2j	$\sum p_T$	0.0032	$\frac{167}{500} = 0.33$	190–225	61	37.4	4.9
	MET	$5 \cdot 10^{-05}$	$\frac{0}{1000} < 0.001$	190–240	7	0.48	0.21
1ν 1μ 1e 2j	$\sum p_T$	0.0086	$\frac{123}{500} = 0.25$	315–340	2	0.13	0.05
	MET	0.0026	$\frac{39}{1000} = 0.04$	65–80	3	0.25	0.06
1ν 1μ 1γ 2j	$\sum p_T$	0.028	$\frac{279}{500} = 0.56$	205–290	5	1.5	0.5
	MET	0.078	$\frac{348}{500} = 0.70$	105–110	1	0.08	0.03
1ν 2μ 2j	$\sum p_T$	0.12	$\frac{436}{500} = 0.87$	210–510	0	2.3	0.6
	MET	0.072	$\frac{252}{500} = 0.50$	45–50	1	0.07	0.03
1ν 1μ 3j	$\sum p_T$	0.0070	$\frac{272}{500} = 0.54$	380-385	3	0.30	0.19
	MET	0.010	$\frac{203}{500} = 0.41$	135–140	2	0.15	0.04
1ν 1μ 1e 3j	$\sum p_T$	0.23	$\frac{226}{500} = 0.45$	350–355	1	0.25	0.17
	MET	0.028	$\frac{89}{500} = 0.18$	80–85	1	0.026	0.017
1ν 1μ 4j	$\sum p_T$	0.029	$\frac{409}{500} = 0.82$	240–285	6	2.19	0.40
	MET	0.041	$\frac{341}{500} = 0.68$	160–165	1	0.04	0.03

Table 9.2: *Summary of results for exclusive muon classes*

9.2 Electron Sample

Again, a general trend can be seen that classes with large p-values ($p \gg 0.002$) also have large $\widetilde{P}$-values ($\widetilde{P} > 0.2$). In the inclusive event classes, 15 out of the 34 distributions show deviations in p more significant than three sigmas ($p < 0.002$); in the exclusive event classes, the agreement is better as only 7 distributions can be found with such a disrepancy (see Tables 9.3 to 9.4). The fact that exclusive event classes show less discrepancies was also found in the muon data. In the electron sample the description of inclusive data seems to be more problematic than in the muon sample, as more classes with "small" p- and $\widetilde{P}$-values are found. In order to concentrate on the event classes with the most interesting deviations, a tighter requirement is chosen an classes with

$$p < 0.002 \quad \text{and} \quad \widetilde{P} < 0.1 \tag{9.2}$$

are marked as "Classes of Interest" and then discussed in detail.

$1\nu\,1e(+X)$ inclusive and exclusive event class:
In the muon sample, the single lepton class showed a significant deviation in the high-p_T regions of the distributions. The electron sample only shows significant deviations in the MET distributions ($p_{incl.} = 1 \cdot 10^{-6}/\widetilde{P}_{incl.} < 0.003$ and $p_{excl.} = 6 \cdot 10^{-6}/\widetilde{P}_{excl.} < 0.003$). In Figure 9.8 one can see the broad excess of data above 60 GeV. The sum of p_T distributions do not reveal this excess, so the presence of a signal might be doubted. It was stated before that the description of Missing Transverse Energy in MC is not well understood as MC distributions are more narrow than the data. In this analysis, MET smearing was only considered as part of the sytematic uncertainties (see Section 8.5). The excess of data in these distributions shows that MET smearing should be further investigated. Only by smearing the MC, the deviation found by the Search Algorithm could be eliminated. Studies are needed to explain why the electron MET, unlike the muon MET, needs additional smearing.
The fact that both sum of p_T distributions (see Figure 9.8) agree with the Standard Model prediction (the excess of data in the single bin is just a fluctuation, see $\widetilde{P} = 0.14$) supports the hypothesis, that the excess of data found in the muon data is caused by muon momentum mis-measurements. "Lepton universality" would demand a similar behaviour in the electron data. Nevertheless, an independent analysis should examine both single lepton classes, to either find and solve the problems of the MC description of the data or to establish a possible signal pointing to New Physics (e.g. to set the first limits of Run II for W').

$1\nu\,1e\,1j + X$ inclusive event class:
In this event class, only the inclusive MET distribution shows a significant deviation with $p_{incl.} = 0.0003$ and $\widetilde{P}_{excl.} = 0.05$. The rest of the class agrees well with the Standard Model. As in the exclusive MET distribution a similar excess of data cannot be found in the same Region of Interest, this single discrepancy is assessed as compatible with the Standard Model.

$1\nu\,1e\,1\gamma\,1j + X$ inclusive event class:
In this event class only the inclusive MET distribution shows a significant deviation with

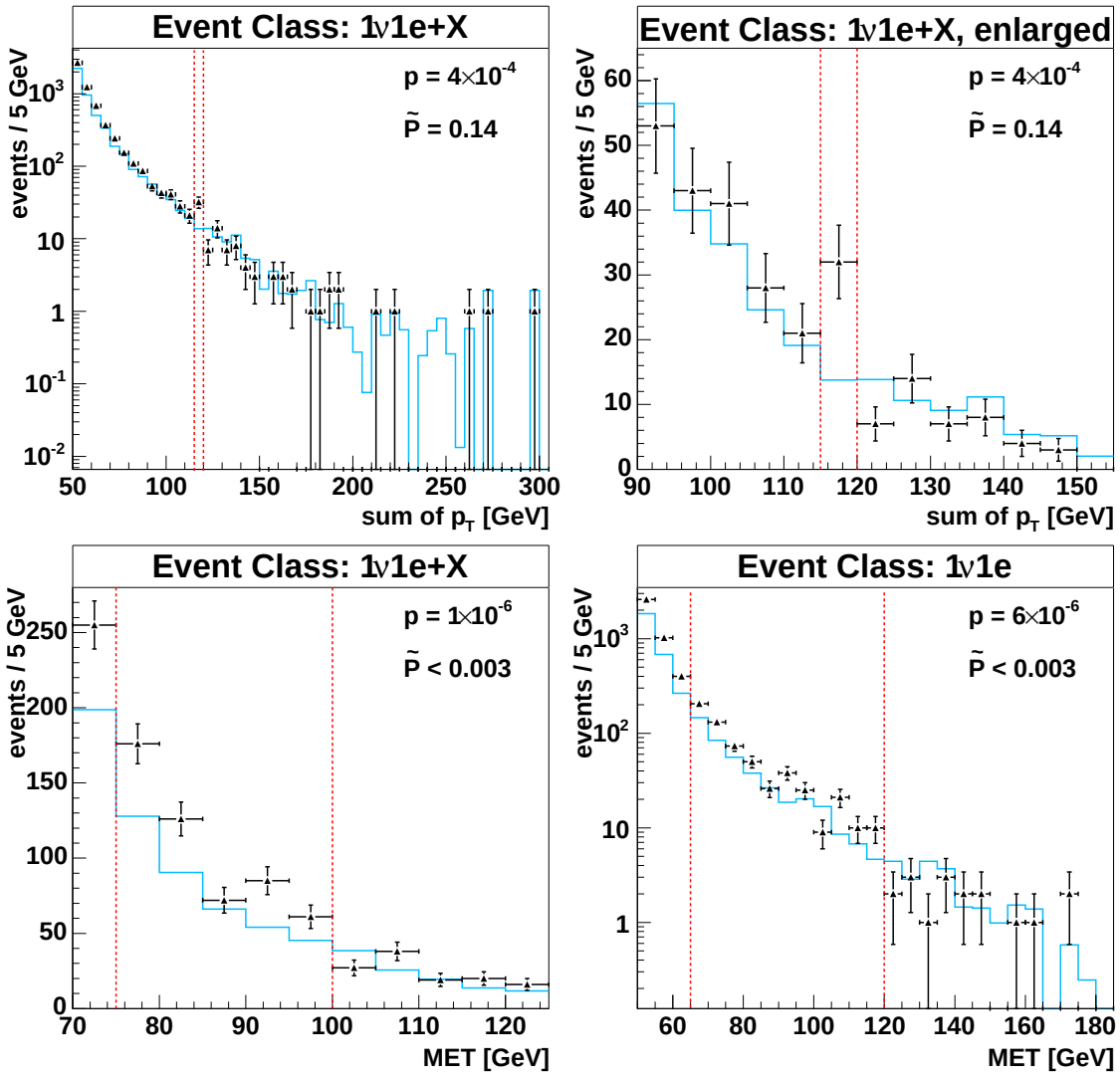

Figure 9.8: *Results of the Search Algorithm in the $1\nu\,1e(+X)$ inclusive and exclusive event classes.*

$p_{incl.} = 0.0015$ and $\widetilde{P}_{excl.} = 0.03$. Figure 9.9 displays this distribution and the excess of three data events. The inclusive sum of p_T distribution has a similar p-value, but with a General Significance of 12% the deviation is compatible with the Standard Model. One could argue that the sytematic uncertainties may be underestimated in the MET distribution. The photon and the jet can lead to additional systematic uncertainties in MET, thus explaining the deviation. In addition to this, the statistics of both, data and MC, in Figure 9.9 are very low. The deviation found here is caused by a MC distribution which is not smooth. For this class more MC samples are needes to increase the MC-statistics. As no other distribution shows significant traces of a possible signal, it is too early to talk of an evidence of New Physics.

$1\nu\,2e(+X)$ inclusive and exclusive event class:

Just like in the muon data, all exclusive distributions and one inclusive distribution show significant deviations with $p < 2 \cdot 10^{-6}$ and $\widetilde{P} < 0.003$. A more moderate, but still significant discrepancy between data and Mont Carlo simulations ($p_{incl.} = 0.0002$ and

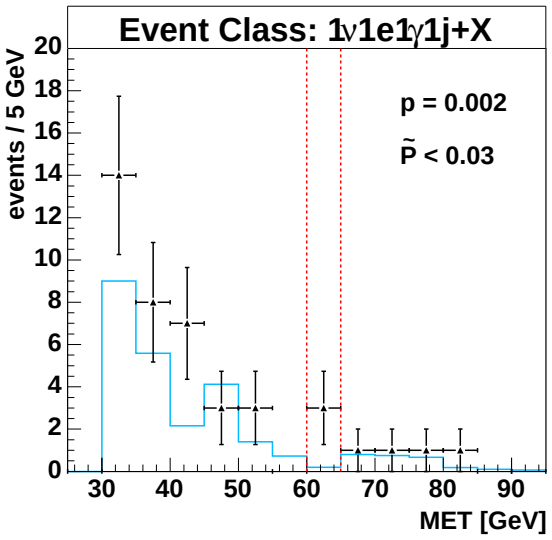

Figure 9.9: *Results of the Search Algorithm in the* $1\nu\,1e\,1\gamma + X$ *inclusive event class.*

$\widetilde{P}_{excl.} = 0.02$) is found in the inclusive MET spectrum. Figure 9.10 shows all four plots. In Table 6.6 and Table 6.8 one can see that in the integral of all events, $sum_{data} > sum_{MC}$ in this event class. If the MC distributions would be normalized to the data, the agreement would be much better as the shapes of data and MC resemble.
The Regions of Interest show a behavior similar to the muon data: They are in both sum of p_T distributions around the Z-region (≈ 90 GeV), and in both MET plots the excess of data is spread over a large momentum range. This fact intensifies the importance of the observed deviation as both, muon and electron data, show a consistent behavior. A signal of New Physics would be expected in muon data and electron data, assuming "lepton universality". On the other hand, if a detector effect is present in the muon sample, a similar process can easily cause a deviation in the electron case, e.g. mis-measured $Z \rightarrow ll$ events. Compared to the muon momentum, the electron p_T is measured much more precisely at high momenta (see Section 2.2). This reduces the probability of mis-measured electrons. On the other hand, a large number of Z-events is produced at the Tevatron, so many mis-measured electrons which could contribute to this event class exist. In order to determine whether this is a detector effect or a signal, further independent studies are needed.

$1\nu\,1e\,1\gamma\,2j(+X)$ inclusive and exclusive event class:
The inclusive and the exclusive sum of p_T distributions show significant deviations with $p_{incl.} = 0.0008/\widetilde{P}_{incl.} = 0.04$ and $p_{excl.} = 0.0010/\widetilde{P}_{excl.} = 0.04$. As both dostributions look very similar, only the exclusive class is shown in Figure 9.11. The excess of four data events is around 210 GeV. The MET spectra also show slight deviations, so the whole event class is not described properly by MC. Table 6.6 and Table 6.8 show that the integral, too, shows an excess of data. The General Significance of 4% is not as low as in other event classes, so one might argue that these deviations are compatible with the Standard Model. As in this class electrons, photons and jets are required, a Monte Carlo simulation not describing the data properly cannot be excluded. All these objects are detected with the calorimeter and misidentification probabilities could be different in data and MC. This combination of particles needs to be examined in detail in order to verify any existence of New Physics.

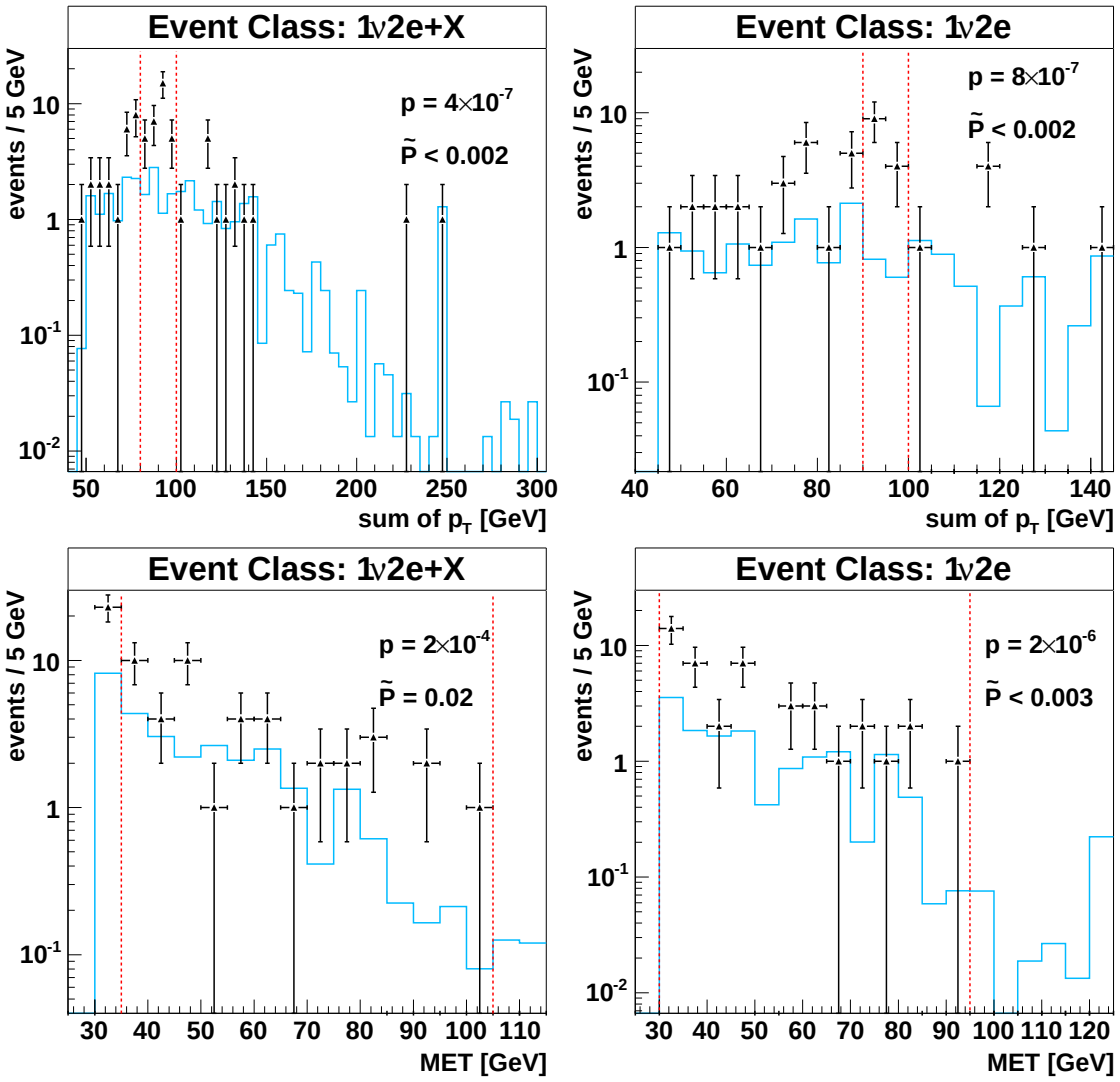

Figure 9.10: *Results of the Search Algorithm in the $1\nu\,2e(+X)$ inclusive and exclusive event classes.*

$1\nu\,1\mu\,1e(+X)$ inclusive and exclusive event class:

This event class shows significant deviations in all inclusive distributions ($\sum p_T$: $p_{incl.} = 0.0005/\widetilde{P}_{incl.} = 0.02$ and MET: $p_{incl.} = 4 \cdot 10^{-5}/\widetilde{P}_{excl.} = 0.001$) and in the exclusive MET spectrum ($p_{excl.} = 0.001/\widetilde{P}_{incl.} = 0.02$), see Figure 9.12. In the muon sample (see Tables 9.1 to 9.2) no discrepancy was found in this event class. This demonstrates the effect of the different sytematic uncertainties implemented in the Search Algorithm as the Regions of Interest are similar for both data samples: For muons these sytematic errors seem to be larger and thus lead to data compatible with the Standard Model. In addition to this, in the muon sample no electron triggers are used and the electrons are only required to have $p_T > 15$ GeV. This is only half of the p_T-cut in the electron sample, so a direct comparison between both samples is difficult.

Regarding the electron data, this deviation is very intersting as only a few Standard Model processes contribute to this event class (WW, $t\bar{t}$, QCD). Many SUSY-signal also have the signature of leptons with different flavor ($e + \mu$), so signals of New Physics seem possible. Nevertheless, the statistics here are very low (6 events and less). More statistics could

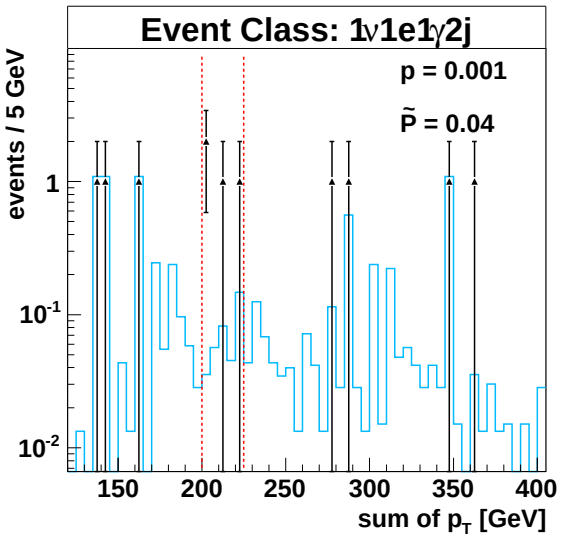

Figure 9.11: *Results of the Search Algorithm in the $1\nu\,1e\,1\gamma\,2j(+X)$ inclusive and exclusive event classes.*

answer the question if these data events are caused by jets identified as electrons and by isolated muons from b-jets, or if a signal is present.

$1\nu\,1\mu\,1e\,1j(+X)$ inclusive event class:
Only in the inclusive MET distribution a significant deviation can be found. As no exclusive events were found in this class and as the $1\nu\,1\mu\,1e\,2j(+X)$ class shows the same Region of Interest, further discussion is continued in the $1\nu\,1\mu\,1e\,2j(+X)$ class.

$1\nu\,1\mu\,1e\,2j(+X)$ inclusive and exclusive event class:
In the muon data (Tables 9.1 to 9.2), a significant deviation was found in this class. Here, too, both inclusive distributions show discrepancies ($\sum p_T$: $p_{incl.} = 0.0007/\widetilde{P}_{incl.} = 0.04$ and MET: $p_{incl.} = 6 \cdot 10^{-5}/\widetilde{P}_{incl.} = 0.002$). In the exclusive MET distribution, data and SM prediction do not agree with $p_{excl.} = 0.0004/\widetilde{P}_{excl.} = 0.01$. All can be found in Figure 9.12. The MET distributions look very similar to the corresponding ones of the muon data. In both, an excess of data around 80 GeV is marked as the Region of Interest. The excess of data in the inclusive electron sum of p_T spectrum is much broader.
The discrepancy in this event class is very interesting as several distributions in both data samples show an excess of data. Also in the integrals of the distributions more data than expected are observed (see Tables 6.2, 6.4, 6.6 and 6.8), but statistics are low. $t\bar{t}$ is the only Standard Model process that contributes to this event class. In order to exclude a problem with the top-MC, an independent $t\bar{t}$ sample should be used and the event class should be investigated in detail. As the measurement of the top mass and the top cross section is a major task of the DØ collaboration, this excess in the data excess should be understood completely. Also, a class with so little Standard Model "background" is an ideal candidate for the discovery of New Physics.

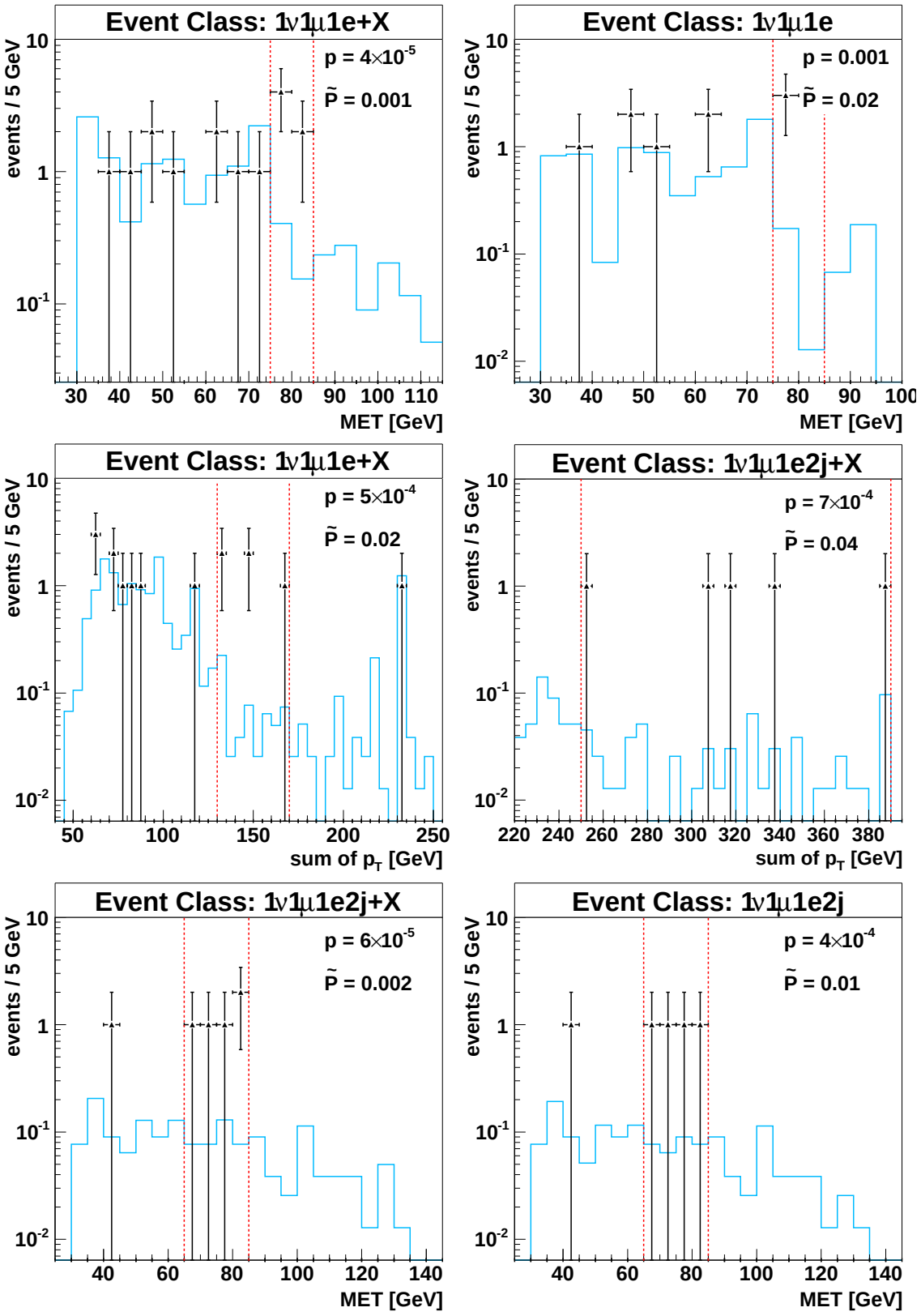

Figure 9.12: *Results of the Search Algorithm in the $1\nu\,1\mu\,1e(+X)$ inclusive and exclusive event classes (first three plots) and in the $1\nu\,1\mu\,1e\,2j(+X)$ inclusive and exclusive event classes (last three plots).*

To conclude, 4 classes with significant deviations are found in the electron data sample:

$$1\nu\,1e(+X),\ 1\nu\,2e(+X),\ 1\nu\,1\mu\,1e(+X) \text{ and } 1\nu\,1\mu\,1e\,2j(+X)$$

What strikes about three of those classes is, that the corresponding muon classes were also valued as not compatible with the MC prediction. Further studies should investigate the origins of these deviations in both lepton classes. In the context of this Model Independent Search, no indications of New Physics can be claimed. For most deviations imperfections in the Monte Carlo simulations can be blamed. Only with the help of more statistics and more elaborate selection cuts, a hint of New Physics could be found in these classes.
The rest of the event classes agrees well with the Standard Model prediction, but a general trend can be observed that MET in the electron data is not well understood. This should be solved for future analyses.
The other Model Independent Search analysis of DØ RunII (see [1]) observes significant deviations in the high-p_T tails of the single electron classes $1e(+X)$ and in the di-electron classes $2e(+X)$, as well. This supports the conclusion that parts of the distributions in these event classes, with or without MET, are not well understood.

Results from Search Algorithm for inclusive event classes of electron data							
Class	Dis	p	$\widetilde{P}$	Region [GeV]	N_{data}	N_{MC}	δN_{MC}
1ν 1e X	$\sum p_T$	$4 \cdot 10^{-4}$	$\frac{14}{100} = 0.14$	115–120	32	13.8	2.6
	MET	$1 \cdot 10^{-6}$	$\frac{0}{300} < 0.003$	75–100	520	383.8	18.7
1ν 1e 1j X	$\sum p_T$	0.011	$\frac{246}{400} = 0.62$	90–100	826	690.0	51.8
	MET	$3 \cdot 10^{-4}$	$\frac{16}{300} = 0.05$	75–100	298	230.0	11.4
1ν 1e 1γ X	$\sum p_T$	0.0029	$\frac{97}{500} = 0.19$	180–200	3	0.23	0.12
	MET	0.0020	$\frac{47}{500} = 0.09$	35–45	79	52.6	4.1
1ν 1e 1γ 1j X	$\sum p_T$	0.0013	$\frac{62}{500} = 0.12$	310–365	3	0.19	0.07
	MET	0.0015	$\frac{15}{500} = 0.03$	60–65	3	0.20	0.06
1ν 2e X	$\sum p_T$	$4 \cdot 10^{-07}$	$\frac{0}{500} < 0.002$	80–100	32	7.2	2.4
	MET	$2 \cdot 10^{-4}$	$\frac{16}{1000} = 0.02$	35–105	44	21.2	2.8
1ν 2e 1j X	$\sum p_T$	0.0022	$\frac{66}{500} = 0.13$	115–160	15	4.1	2.1
	MET	0.0063	$\frac{69}{500} = 0.14$	75–105	4	0.70	0.11
1ν 2e 1γ X	$\sum p_T$	0.033	$\frac{71}{500} = 0.14$	115–125	1	0.03	0.02
	MET	0.033	$\frac{49}{500} = 0.10$	55–60	1	0.03	0.02
1ν 1e 2j X	$\sum p_T$	$9 \cdot 10^{-4}$	$\frac{62}{500} = 0.12$	140–175	198	134.7	15.2
	MET	0.0016	$\frac{75}{500} = 0.15$	30–50	453	331.5	35.7
1ν 1e 1γ 2j X	$\sum p_T$	$8 \cdot 10^{-4}$	$\frac{20}{500} = 0.04$	200–225	4	0.37	0.11
	MET	0.0049	$\frac{34}{500} = 0.07$	30–85	11	3.9	0.8
1ν 2e 2j X	$\sum p_T$	0.050	$\frac{340}{500} = 0.68$	185–255	3	0.80	0.14
	MET	0.10	$\frac{382}{500} = 0.76$	95–105	1	0.11	0.04
1ν 1e 3j X	$\sum p_T$	0.013	$\frac{396}{500} = 0.79$	215–230	13	6.0	0.9
	MET	0.014	$\frac{208}{500} = 0.42$	45–50	11	4.6	0.9
1ν 1e 4j X	$\sum p_T$	0.025	$\frac{405}{500} = 0.81$	330–380	6	2.0	0.5
	MET	0.026	$\frac{285}{500} = 0.57$	85–95	3	0.59	0.19
1ν 1e 5j X	$\sum p_T$	0.027	$\frac{220}{500} = 0.44$	340–390	2	0.22	0.12
	MET	0.035	$\frac{144}{500} = 0.29$	245–250	1	0.03	0.02
1ν 1μ 1e X	$\sum p_T$	$5 \cdot 10^{-4}$	$\frac{12}{500} = 0.024$	130–170	5	0.6	0.1
	MET	$4 \cdot 10^{-05}$	$\frac{1}{1000} = 0.001$	75–85	6	0.6	0.1
1ν 1μ 1e 1j X	$\sum p_T$	0.011	$\frac{153}{500} = 0.31$	210–295	5	1.2	0.3
	MET	$7 \cdot 10^{-4}$	$\frac{13}{500} = 0.03$	65–85	5	0.66	0.12
1ν 1μ 1e 2jX	$\sum p_T$	$7 \cdot 10^{-4}$	$\frac{18}{500} = 0.04$	250–390	5	0.62	0.16
	MET	$6 \cdot 10^{-05}$	$\frac{2}{1000} = 0.002$	65–85	5	0.36	0.09
1ν 1μ 1e 3jX	$\sum p_T$	0.095	$\frac{132}{500} = 0.26$	350–355	1	0.10	0.05
	MET	0.095	$\frac{116}{500} = 0.23$	80–85	1	0.10	0.05

Table 9.3: *Summary of results for inclusive electron classes*

Results from Search Algorithm for exclusive event classes of electron data							
Class	Dis	p	$\widetilde{P}$	Region [GeV]	N_{data}	N_{MC}	δN_{MC}
1ν 1e	$\sum p_T$	0.011	$\frac{115}{223} = 0.52$	125–170	9	20.5	2.7
	MET	$6 \cdot 10^{-06}$	$\frac{0}{360} < 0.003$	65–120	598	426.4	31.1
1ν 1e 1j	$\sum p_T$	0.002	$\frac{99}{500} = 0.20$	280–325	12	28.3	3.0
	MET	0.0043	$\frac{135}{400} = 0.34$	75–100	229	183.6	9.4
1ν 1e 1γ	$\sum p_T$	0.030	$\frac{401}{500} = 0.80$	85–110	27	17.2	2.2
	MET	0.015	$\frac{187}{500} = 0.37$	35–40	43	26.1	4.8
1ν 1e 1γ 1j	$\sum p_T$	0.025	$\frac{352}{500} = 0.70$	90–100	6	1.7	0.9
	MET	0.020	$\frac{158}{500} = 0.32$	60–65	2	0.12	0.17
1ν 2e	$\sum p_T$	$8 \cdot 10^{-07}$	$\frac{0}{500} < 0.002$	90–100	13	1.4	0.7
	MET	$2 \cdot 10^{-06}$	$\frac{0}{400} < 0.003$	30–95	43	14.4	3.1
1ν 2e 1j	$\sum p_T$	0.0039	$\frac{150}{1000} = 0.15$	115–160	14	3.8	2.2
	MET	0.0041	$\frac{55}{500} = 0.11$	75–95	3	0.3	0.1
1ν 2e 1γ	$\sum p_T$	0.51	$\frac{334}{500} = 0.67$	120–125	1	0.76	0.36
	MET	0.51	$\frac{336}{500} = 0.67$	55–60	1	0.76	0.36
1ν 1e 2j	$\sum p_T$	$6 \cdot 10^{-4}$	$\frac{50}{500} = 0.10$	170–175	29	12.3	2.4
	MET	0.0044	$\frac{147}{500} = 0.29$	30–50	391	293.0	32.0
1ν 1e 1γ 2j	$\sum p_T$	0.0010	$\frac{19}{500} = 0.04$	200–225	4	0.37	0.14
	MET	0.0036	$\frac{60}{1000} = 0.06$	30–85	11	3.7	0.8
1ν 2e 2j	$\sum p_T$	0.042	$\frac{306}{500} = 0.61$	185–255	3	0.7	0.1
	MET	0.10	$\frac{392}{500} = 0.78$	95–105	1	0.11	0.04
1ν 1e 3j	$\sum p_T$	0.010	$\frac{338}{500} = 0.68$	170–230	37	22.3	3.1
	MET	0.0064	$\frac{138}{500} = 0.28$	40–50	19	9.3	1.2
1ν 1e 4j	$\sum p_T$	0.053	$\frac{478}{500} = 0.96$	330–380	5	1.9	0.5
	MET	0.020	$\frac{207}{500} = 0.41$	85–90	3	0.3	0.4
1ν 1e 5j	$\sum p_T$	0.026	$\frac{205}{500} = 0.41$	340–390	2	0.2	0.1
	MET	0.035	$\frac{133}{500} = 0.27$	245–250	1	0.03	0.02
1ν 1μ 1e	$\sum p_T$	0.046	$\frac{281}{500} = 0.56$	60–80	6	2.4	0.6
	MET	0.0013	$\frac{22}{1000} = 0.02$	75–85	3	0.2	0.1
1ν 1μ 1e 2j	$\sum p_T$	0.0036	$\frac{49}{500} = 0.10$	250–390	4	0.6	0.1
	MET	$4 \cdot 10^{-4}$	$\frac{5}{500} = 0.01$	65–85	4	0.3	0.1
1ν 1μ 1e 3j	$\sum p_T$	0.095	$\frac{110}{500} = 0.22$	350–355	1	0.10	0.05
	MET	0.095	$\frac{113}{500} = 0.23$	80–85	1	0.10	0.05

Table 9.4: *Summary of results for exclusive electron classes*

Chapter 10

Conclusion

In this diploma thesis, the first Model Independent Search for New Physics during DØ Run II has been performed successfully.
In this apporach, *all* events detected with at least one electron/muon and a certain amount of Missing Transverse Energy are selected. Following the specific topology of an event, exclusive and inclusive event classes are constructed, e.g. $1\nu\,1\mu\,3jets(+X)$. This results in a "muon sample" of 16 different particle combination and an "electron sample" of 17 combinations. In a first step, the total number of events per class is compared to the Standard Model prediction using Monte Carlo simulations. In a second step, these event classes are systematically scanned for deviations from the Standard Model by examining the variable $\sum p_T$ and the MET distribution. Therefore, a special Search Algorithm has been developed, based on the work of former Model Independent analyses ([18], [19] and [20]). This algorithm identifies the region with the greatest discrepancy between data and Standard Model prediction, calculates the probablity of this deviation and tests its general significance. Statistical, as well as systematic uncertainties are implemented in this routine.

In the analysis presented here, a data set corresponding to an integrated luminosity $\mathcal{L} \approx 220\,\mathrm{pb}^{-1}$ of $p\bar{p}$-collisions is considered. These data were recorded by the DØ experiment between July 2002 and April 2004 with a center of mass energy of 1.96 TeV.
In the comparison of the total number of events per class, the general agreement between data and Standard Model prediction is good. This leads to the conclusion that the DØ-experiment is understood well and the detector simulation reproduces most of the Standard Model-processes corretly. All observed deviations could be explained with detector effects not simulated properly in the Monte Carlo simulations. Nevertheless, the deviations already observed at this level should be taken seriously in order to improve the description of the data.
What concerns the differential cross section, in the muon sample, 5 event classes with significant deviations have been found by the Search Algorithm $\{1\nu\,1\mu(+X)$, $1\nu\,1\mu\,1j(+X)$, $1\nu\,2\mu(+X)$, $1\nu\,1\mu\,2j(+X)$ and $1\nu\,1\mu\,1e\,2j(+X)\}$. In the electron sample, 4 event classes with significant deviations have been found $\{1\nu 1e(+X)$, $1\nu 2e(+X)$, $1\nu 1\mu 1e(+X)$ and $1\nu 1\mu 1e2j(+X)\}$. Hence, three classes show discrepancies in both data samples, which intensifies the disagreement further.

In another Model Independent Search analysis of DØ Run II (see [1]), significant deviations have been observed in the single lepton classes $1l + (X)$ and in the di-lepton classes $2l(+X)$, showing results consistent with the paper presented here.

The decision if this indicates New Physics or if problems with the Monte Carlo simulations are the origin cannot be made in the context of this analysis. For all deviations, hypotheses other than New Physics have been found. Further studies should investigate these deviations in both lepton samples. Only with the help of more statistics and more elaborate selection cuts, a hint of New Physics could be verified in these classes.

In addition to these results in a search for new phenomena, it has been shown that the Search Algorithm and the concept of Model Independent Search in general is of great value for a collider experiment. The broad scan of almost all data recorded by the detector provides a fast and precise test of the general description and comprehension of the data. With the help of simple selction cuts and the Search Algorithm, an overall search for New Physics can be performed. This strategy is an ideal complement to other analyses testing a specific extension of the Standard Model.

The paper presented here is meant to gather experience and provide information on the issue of Model Independent Search. The detailed description of the Search Algorithm and its assumptions could help future analyses to develop their own infrastructure. Especially for the experiments at the Large Hadron Collider, which soon will be completed, a Model Independent Search could be very useful and successful. With its general scan of the data, many things about the new detector and Standard Model processes contributing to a certain event class could be learned. And, of course, the chances of detecting New Physics will be much higher.

Appendix A

Trigger Definitions

Electron triggers for lists up to version global CMT-11.04, corresponding to runnumber < 178722, prescales ranging from $1 - 50$:

L1-requirement: One EM trigger tower having $E_T > 10$ GeV (trigger towers are coarse combinations of calorimeter towers with $\Delta\eta = \Delta\varphi = 0.2$)

EM_HI: L2: none
L3: The trigger bit set to true if one loose EM object (loose $\widehat{=}$ electromagnetic fraction $EMf > 0.9$) is found with $E_T > 30$ GeV in $|\eta| < 3$

EM_HI_SH: L2: none
L3: The trigger bit set to true if a $|\eta| < 3$ EM object is found with $E_T > 20$ GeV meeting loose criteria including a transverse shower shape requirement

EM_HI_EMFR8: L2: none
L3: The trigger bit set to true if a $|\eta| < 3$ EM object with $E_T > 40$ GeV meeting very loose criteria (EMf> 0.8) is found

EM_HI_TR: L2: none
L3: One track is found by the Global Tracker tool with $p_T > 25$ GeV

EM_HI_SH_TR: L2: none
L3: One track is found by the Global Tracker tool with $p_T > 12$ GeV and one loose calorimeter EM object satisfying transverse shower shape requirements and $E_T > 12$ GeV

Electron-triggers for lists starting with version global CMT-12.10, corresponding to 178721 <runnumber< 192159. These triggers have different L1-definitions but all share the same prescale 1:

L1-requirement: E1: Require one calorimeter EM object with $E_T > 11$ GeV and not calorimeter unsuppressed readout (to reduce data size)
E2: Require two calorimeter EM objects with $E_T > 6$ GeV and not calorimeter unsuppressed readout
E3: Require two calorimeter EM towers with $E_T > 3$ GeV. One of the towers must have $E_T > 9$ GeV. Also, not calorimeter unsuppressed readout

E1_SHT20: L2: none
L3: Requires an EM object satisfying tight shower shape requirements with $E_T > 20$ GeV meeting loose criteria ($EMf > 0.9$) in $|\eta| < 3.6$

E2_SHT20: L2: none
L3: Requires an EM object satisfying tight shower shape requirements with $E_T > 20$ GeV meeting loose criteria ($EMf > 0.9$) in $|\eta| < 3.6$

E3_SHT20: L2: none
L3: Requires an EM object satisfying tight shower shape requirements with $E_T > 20$ GeV meeting loose criteria ($EMf > 0.9$) in $|\eta| < 3.6$

E1_SH30: L2: none
L3: One EM object meeting loose criteria ($EMf > 0.9$) with loose shower requirements with $E_T > 30$ GeV in $|\eta| < 3.6$

E2_SH30: L2: none
L3: One EM object meeting loose criteria ($EMf > 0.9$) with loose shower requirements with $E_T > 30$ GeV in $|\eta| < 3.6$

E3_SH30: L2: none
L3: One EM object meeting loose criteria ($EMf > 0.9$) with loose shower requirements with $E_T > 30$ GeV in $|\eta| < 3.6$

Muon-triggers for lists up to version global CMT-10.03, corresponding to runnumber$<$ 173102, prescales ranging from $1-361$:

L1-requirement: Wide-region (CFT) muon scintillator trigger with tight condition (one hit before and one after toroid) also requiring the minimum-bias condition (luminosity monitors above threshold in coincidence)

MU_W_L2M5_TRK10: L2: Medium quality muon candidate (one scintillator hit and two wire hits) with $p_T > 5$ GeV
L3: The trigger bit set to true if one track is found by the Global Tracker tool with $p_T > 10$ GeV

Muon-triggers for lists starting with version global CMT-10.30, corresponding to 173352 $<$runnumber$<$ 192159, prescales ranging from $1-2$:

L1-requirement: Wide-region (CFT) muon with tight scintillator (one hit before and one after toroid) and loose wire requirements (at least two hits)

MUW_W_L2M3_TRK10: L2: Medium quality muon candidate (one scintillator hit and two wire hits) with $p_T > 3$ GeV
L3: The trigger bit set to true if one track is found by the Global Tracker tool with $p_T > 10$ GeV

Bibliography

[1] Oliver Kraff. Search for new physics in proton-antiproton collisions in final states with leptons and jets. RWTH-Aachen, 2004.

[2] M. Peskin and D. Schroeder. *An Introduction to Quantum Field Theory.* 1995.

[3] `http://www.science-park.info/particle/force_unification.gif`.

[4] S. Chekanov et al. A zeus next-to-leading-order qcd analysis of data on deep inelastic scattering. *Phys. Rev.*, D67:012007, 2003.

[5] `http://cmsinfo.cern.ch/Welcome.html/CMSdocuments/CMSdocuments.html`.

[6] DØ-Presentations. `http://www-d0.fnal.gov/Run2Physics/displays/presentations/`.

[7] Run II Handbook. `http://www-bd.fnal.gov/runII/index.html`.

[8] Accelerator Division. `http://www-bd.fnal.gov/`.

[9] Visual Media Service: Photo Database. `http://www-visualmedia.fnal.gov/VMS_Site/active.html`.

[10] S. Abachi et al. The DØ detector. *Nucl. Instrum. Meth.*, A338:185–253, 1994.

[11] S. Abachi et al. The DØ upgrade: The detector and its physics. FERMILAB-PUB-96-357-E.

[12] The Top Physics Working Group of the DØ Collaboration. DØ top analysis and data sample for the winter conferences 2004. DØ-Note 4419.

[13] Detector Drawings. `http://www-d0.fnal.gov/Run2Physics/WWW/drawings.htm`.

[14] Gregory Arthur Davis. *First Measurement of the Differential Inclusive Cross Section for Jet Production at DØ Run II.* PhD thesis, University of Rochester, 2004.

[15] Bing Zhou, Quichun Xu, and James Degenhardt. Parameterized monte carlo simulation muon package. DØ-Note 4065.

[16] The updated DØ luminosity determination. DØ-Note 4328.

[17] Trigger Meister's Webpage. `http://www-d0.fnal.gov/trigger_meister/private/www/tl_desc/global.html`.

[18] B. Abbott et al. A quasi-model-independent search for new physics at large transverse momentum. *Phys. Rev.*, D64:012004, 2001.

[19] A. Aktas et al. A general search for new phenomena in e p scattering at HERA. *Phys. Lett.*, B602:14–30, 2004.

[20] Thomas Hebbeker. A global comparison between L3 data and standard model monte carlo - a first attempt. L3-Note 2305.

[21] DØ Farm Production Page. `http://www-d0.fnal.gov/computing/production/talks/farm-report-alld0-2001-10-13.pdf`.

[22] Common Sample Group. `http://www-d0.fnal.gov/Run2Physics/cs/index.html`.

[23] M. Klute, M. Phaf, and D. Whiteson. Topanalyze - a framework analyze package for top group analyses. DØ-Note 4122.

[24] R. Brun et al. *ROOT-User Guide v3.1b*. June 2001.

[25] Offline Quality Data Base. `http://d0db.fnal.gov/qualitygrabber/qualQueries.html`.

[26] JET/MET Certification Results. `http://www-d0.fnal.gov/computing/algorithms/calgo/jet_met/runsel.html`.

[27] Bad Ring of-Fire luminosity blocks. `http://d0-france.in2p3.fr/WORKING_GROUPS/DQ/RF/lum_blocks_mht30_bad.dat`.

[28] H. Schellmann and M. Verzocchi. `http://www-d0.fnal.gov/phys_id/luminosity/data_access/lm_access/doc/runrange_luminosity.html`.

[29] EM ID Certification Results. `http://www-d0.fnal.gov/phys_id/emid/d0_private/certification/main_v5_0.html`.

[30] Georg Steinbrueck. *Measurement of the Angular Distribution of Electrons from W Boson Decays at DØ*. PhD thesis, University of Oklahoma Graduate College, 1999.

[31] D. Chapin and H. Fox. Measurement of $Z \to e^+e^-$ and $W \to e^\pm\nu$ production cross section with $|\eta| < 2.3$. DØ-Note 4403.

[32] S. Lager, G. Watts, et al. Primary vertex certification in p14. DØ-Note 4320.

[33] Carsten Magass. Bestimmung des inklusiven Wirkungsquerschnitts für Elektronproduktion bei Proton-Antiprotin-Kollisionen. RWTH-Aachen, 2003.

[34] C. Clément, F. Deliot, et al. Muonid certification for p14. DØ-Note 4350.

[35] F. Deliot and G. Hesketh. Measurement of $\sigma(p\bar{p} \to WX) \times BR(W \to \mu\nu)$ at $\sqrt{s} =$ 1.96 TeV. DØ-Note XXXX.

[36] Thomas Stapelberg. Messung des inklusiven Wirkungsquerschnitts für Myonproduktion am Tevatron Beschleuniger. RWTH-Aachen, 2003.

[37] Jerry Blazey et al. Run II jet physics. DØ-Note 3750.

[38] Jet ID Certification Results. `http://www-d0.fnal.gov/computing/algorithms/calgo/jet/jetID_p14.html`.

[39] Jet Energy Scale Group. http://www-d0.fnal.gov/phys_id/jes/d0_private/certified/certified.html.

[40] L. Sawyer and A. Stone. Missing ET reconstruction: Variables and methods. DØ-Note 3957.

[41] Torbjorn Sjostrand, Leif Lonnblad, and Stephen Mrenna. Pythia 6.2: Physics and manual. 2001.

[42] http://www-subatech.in2p3.fr/~photons/subatech/physics/collisionneurs/node58.html.

[43] DØ Run II Simulation. http://www-d0.fnal.gov/computing/MonteCarlo/MonteCarlo.html.

[44] http://www-d0.fnal.gov/computing/MonteCarlo/d0gstar_pid_2.gif.

[45] NP MC Requests/Status. http://www-d0.fnal.gov/Run2Physics/np/.

[46] private correspondence with Ulla Blumenschein. University of Freiburg.

[47] H. Plothow-Besch: PDFLIB. http://www-d0.fnal.gov/computing/MonteCarlo/generator_tools/pdflib.html.

[48] Andreas Vogt et al. http://www.ippp.dur.ac.uk/ avogt/.

[49] Thomas Nunnemann. NNLO cross-sections for Drell-Yan, Z and W production using modern parton distribution functions. DØ-Note 4476.

[50] J. M. Campbell and R. K. Ellis. An update on vector boson pair production at hadron colliders. *Phys. Rev.*, D60:113006, 1999.

[51] Supriya Jain. Scale and over-smearing for hight-p_T MC electrons. DØ-Note 4402.

[52] Muon smearing. http://www-d0.fnal.gov/Run2Physics/top/d0_private/wg/production/analysis_summer04/muon_smearing.txt.

[53] Robert Kehoe. Oversmearing of missing transverse energy in $Z \rightarrow ee + X$ Monte Carlo events. DØ-Note 4551.

[54] A. Alton et al. Measurement of Wγ events in DØ Run II data. DØ-Note XXXX.

[55] D0_MESS: Generator level event selector. http://www-d0.fnal.gov/computing/MonteCarlo/generator_tools/d0_mess.html.

[56] S. Anderson et al. Measurement of the $t\bar{t}$ production cross-section at $\sqrt{s} = 1.96$ TeV in dilepton final sates. DØ-Note XXXX.

[57] J. Campbell and J. Huston. PDF error discussion: the sequel to the sequel. http://www-d0.fnal.gov/computing/MonteCarlo/generator_tools/pdf_error4.pdf.

Acknowledgements

So, the end is near and only the credits are missing...

These pages represent one year of work, of new experiences and of numerous challenges. These challenges could only be coped with the help of many people who contributed a lot to this stack of paper:

First I would like to thank Oliver Kraff for the extraordinary team work which pushed this analysis way forward. I really enjoyed our common way of solving problems together.
In addition to this, I owe countless thanks to Carsten Magass, Daniela Käfer and Christian Autermann for having time for so many computer problems and fundamental questions of mine. Your help and advice makes the superb working atmosphere in this analysis group possible. Then I want to thank Thomas Stapelberg, Martin Wegner, Jan Coenen, Dr. Steffen Kappler, Matthias Kirsch and Dr. Arndt Meyer for the fruitful discussions during our meetings or at other occasions.
Especially I would like to thank the best office associate in the world, Lotte Wilke, for bringing color into the grey everyday life and for alwyas being there. I will miss all the fun we had very much.

Finally, I would like to thank Prof. Dr. T. Hebbeker for his remarkable supervision. The weekly meetings and his honest interest should not be taken for granted. Besides this, the opportunity to take part at an experiment taking real data and actually visiting it has been a nice experience.
In the same manner I would like to thank Prof. Dr. M. Erdmann who was instantaneously interested in the topic of this diploma thesis and agreed to be my second referee.

As there also is a life next to work, I would like to express honest gratitude to my parents and to my brother Christof as all of you alwyas supported me, financially and emotionally, and believed in the decisions I had to make. Then I want to thank Helga for sacrificing their holidays to check my English spelling.
For more than just support, I would like to thank my friends which I have known for almost half of my life: Kevin, Marcel, Tim, Mascha, Jennifer and Barbara. Any good time I had is connected to you, and I am very grateful that you showed so much sympathy if my time schedule was too narrow to include you.

Then I want to thank my friends from Aachen for all the good times we spent together: Philip, Saskia, Patrik, Maria, JarJar, Thomas, Nico, Sebastian, Benno, Henning and the nightlife crew Julia+Verena. It is hard to imagine that university is over now!

I owe special thanks to “the sweetest thing”, Mira, for supporting me in the final stage of this work and for transforming a stressful time into a wonderful experience.

At this point I have to apologize to all the people I forgot to mention in these few lines. To my excuse, it is three o’ clock in the morning on Christmas Eve. So Merry Christmas to everybody and good night...