Università degli Studi di Firenze

Dipartimento di Fisica

Scuola di Dottorato in Scienze

Dottorato di Ricerca in Fisica XXI ciclo

Study of $Z/\gamma^* + jets$ production with the CMS detector using matched Matrix Element-Parton Shower event generators

Piergiulio Lenzi

Relatore: **Dott. M. Meschini**

Coordinatore: **Prof. A. Cuccoli**

Settore disciplinare FIS/04

Firenze, 31 Dicembre 2008

Contents

Introduction

The characterization of $Z/\gamma^* + jets$ production at the Large Hadron Collider, with the vector boson decaying leptonically, will be one of the goals of early physics analyses of the Compact Muon Solenoid (CMS) experiment. The rather clear leptonic signature will make these events easy to identify, and the vector boson kinematics will be reconstructed with reasonable precision, even with a not perfectly calibrated detector. These events will be very useful for the calibration of the CMS calorimetric response, using the balancing of the jets and the recoiling Z boson. Kinematic observables of the Z boson, such as the transverse momentum p_T, will provide us with a detailed description of the QCD radiation pattern, because the Z boson recoils against QCD radiation. The distributions of these observables are expected to be easily reconstructed using the excellent CMS tracking system. Also, $Z/W + jets$ events represent a background for many new physics searches, such as Super-Symmetry.

For all these reasons it is extremely important to understand the differences among the Monte Carlo event generator programs that can produce such kind of events. In particular, it is crucial to understand the theoretical uncertainties, and to see how the different approximations used in these programs produce differences in the observables which can be reconstructed in experiments.

A study on different event generators for the production of $Z/\gamma^* + jets$ events at LHC is presented. These generators implement a combined use of fixed order Matrix Element calculations and all order Parton Shower calculations. We used some of the most popular programs for hadron collisions simulation: `PYTHIA` and `HERWIG`, which are probably the most widely used event generators in high energy physics; `SHERPA` and `AlpGen`, which are newer programs that implement higher order tree level corrections for a number of processes.

In this work a description of the different techniques implemented in these event generators for the simulation of $Z/\gamma^* + jets$ events is illustrated. We present three levels of comparison. First, we compare the above mentioned event generators on the ground of first order correction for the inclusive Z production. Since all the generators used are able to take into account the first order tree level correction, they should give similar results. Then we compare `AlpGen` and `SHERPA` including higher order corrections; we also present comparisons with $Z/\gamma^* + jets$ measurements performed at the Tevatron $p\bar{p}$ collider. Finally we process `AlpGen` and `SHERPA` events with the CMS detector simulation, to see how detector effects smear the distributions obtained at generator level, and to see if the differences observed at generator level are still recognizable in the reconstructed distributions.

Chapter 1

Electroweak physics and QCD at LHC

In this chapter the Standard Model of Electroweak and Strong interactions is introduced. The Standard Model [1] is a Quantum Field Theory based on a $\mathrm{SU(3)_c} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ local gauge symmetry. We shall see that the above mentioned symmetry can be satisfied only if the fermion fields are massless: this contrasts with the experimental observation of massive fermions. A mechanism known as spontaneous symmetry breaking is used in the Standard Model to provide elementary particles with mass; this mechanism requires the existence of a new, still unobserved, field known as Higgs field.

We shall see how the request for a local gauge symmetry and the spontaneous breaking of the $SU(2)_L \otimes U(1)_Y$ symmetry leads to the prediction of the existence of the weak gauge bosons Z and $W^{\pm}$ and how the fermion fields acquire mass through the interaction with the Higgs field; we will see that the unbroken $U(1)_{em}$ symmetry is responsible for the electromagnetic interaction mediated by the photon γ; we will finally include the strong interaction in this picture, through the request of a local $SU(3)_c$ gauge symmetry, that is mediated by eight colored gluons.

Within this framework we shall put into evidence the phenomenology of the production of Z and $W^{\pm}$ bosons, as their production rate at LHC will be unprecedentedly high, thus allowing very precise measurements.

Unless otherwise stated natural units $\hbar$=c=1 are used throughout this work.

1.1 Elementary particles

The Standard Model is built with six spin-$\frac{1}{2}$ particles called leptons and six spin-$\frac{1}{2}$ particles called quarks. They are classified in three generations and there is no evidence for a fourth generation so far.

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \qquad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \qquad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$$

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} s \\ c \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}.$$

Ordinary matter consists of leptons and hadrons; hadrons are classified in two categories: mesons are bound states of a quark and an antiquark, baryons are bound states of three quarks.

In the Standard Model leptons and quarks interact through three fundamental interactions mediated by spin-1 bosons:

Strong : mediated by eight gluons g.

Weak : mediated by W^+, W^-, Z.

Electromagnetic : mediated by the photon γ.

Weak and Electromagnetic interactions are actually two manifestations of the same fundamental interaction called Electroweak interaction.

1.2 Electroweak Interaction

The theory of Electroweak interaction has been formulated by S.L. Glashow [2], A. Salam [3] and S. Weinberg [4] as an $SU(2) \otimes U(1)$ local gauge theory. The local $SU(2)$ symmetry requires the absence of mass terms for the fermions in the lagrangian. We shall see how the introduction of the Higgs field and the spontaneous symmetry breaking mechanism can be used to consistently provide the elementary fermions with mass.

1.2.1 Electroweak unification

In 1957 the Madame Wu [5] and Garwin-Lederman-Weinrich [6] experiments confirmed the parity violation for weak charged current interactions. Weak charged current interactions were proven to prefer final states with left-handed particles or right handed antiparticles [7]. Already in Fermi's four fermions theory [8] of weak charged interactions two conserved currents were identified:

$$J_\mu^+ = \bar{\nu}_L \gamma_\mu \ell_L, \tag{1.1}$$

$$J_\mu^- = \bar{\ell}_L \gamma_\mu \nu_L, \tag{1.2}$$

where ν_L and ℓ_L are the components of the left handed lepton spinor L_L:

$$L_L = \frac{1}{2}(1-\gamma^5)\begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix} = \begin{pmatrix} \nu_{\ell L} \\ \ell_L \end{pmatrix}. \tag{1.3}$$

The Noether theorem states that whenever a lagrangian density is invariant for global transformations of a symmetry group a conserved current exists for each of the generators T^A of the symmetry group, with form:

$$\bar{\psi}_i \gamma_\mu T^A_{ij} \psi_j. \tag{1.4}$$

Introducing also the $SU(2)$ "up" and "down" operators

$$\tau_{\pm} = \frac{1}{2}(\tau_1 \pm i\tau_2), \tag{1.5}$$

(where τ_i with $i = 1, 2, 3$ are the Pauli matrices) we can rewrite the J_μ^+ and J_μ^- conserved currents as

$$J_\mu^{\pm} = \bar{L}_L \gamma_\mu \tau_{\pm} L_L. \tag{1.6}$$

In order to complete the $SU(2)$ invariance of the theory a third conserved current should exist with form

$$J_\mu^3 = \bar{L}_L \gamma_\mu \tau_3 L_L. \tag{1.7}$$

This current cannot be identified with the neutral current though, because the weak neutral current involves both left handed and right handed components. Also the electromagnetic current has both left and right handed components, and it does not couple with the chargeless neutrino, thus it is not easily connected to J_μ^3.

In order to save the $SU(2)$ symmetry, the existence of a new $U(1)$ symmetry is required, and a new conserved current arises. The new symmetry is known as hypercharge symmetry and is indicated with $U(1)_Y$. The corresponding conserved current is

$$J_\mu^Y = \bar{\psi} \gamma_\mu Y \psi, \tag{1.8}$$

where the hypercharge operator Y is

$$Y = 2(Q - \tau_3), \tag{1.9}$$

where Q is the electric charge, and τ_3 is the "weak isospin" quantum number, connected to the $SU(2)$ symmetry. In this way the J_μ^3 current is preserved, thus completing the $SU(2)_L$ symmetry, where the L subscript denotes the so called weak isospin quantum number. The electromagnetic current is expressed as a linear combination of J_μ^3 and J_μ^Y:

$$J_\mu^{em} = J_\mu^3 + \frac{1}{2} J_\mu^Y. \tag{1.10}$$

Eq. (1.10) represents the Electroweak unification, that is the unification of weak and electromagnetic interaction. In Sec. 1.2.2 we shall see how the weak neutral current can be expressed as an analogous linear combination. The relevant quantum numbers of electrons and quarks are summarized in Table 1.1.

1.2.2 Lepton sector

The $SU(2)_L \otimes U(1)_Y$ globally invariant electroweak interaction is expressed as follows:

$$\mathcal{L}_{\text{EW}}^{\text{lepton}} = i \sum_{k=1}^{3} \left(\bar{L}_L^k \gamma^\mu \partial_\mu L_L^k + \bar{L}_R^k \gamma^\mu \partial_\mu L_R^k \right), \tag{1.11}$$

Fermions	Quantum Numbers			
	Q_{em}	T	T_3	Y
$(\nu_\ell)_L$	0	$\frac{1}{2}$	$+\frac{1}{2}$	-1
$(\ell)_L$	-1	$\frac{1}{2}$	$-\frac{1}{2}$	-1
$(\ell)_R$	-1	0	0	-2
$(u)_L$	$+\frac{2}{3}$	$\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{3}$
$(d)_L$	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{3}$
$(u)_R$	$+\frac{2}{3}$	0	0	$+\frac{4}{3}$
$(d)_R$	$-\frac{1}{3}$	0	0	$-\frac{2}{3}$

Table 1.1: Quantum numbers of lepton (ℓ=e, μ, τ) and quark ([u, d]=[u, d], [c, s], [t, b]) helicity states: electric charge Q_{em} in unit of e, weak isospin T with third axis projection T_3 and weak hypercharge Y.

where $\tau^\mu = (1, \tau_1, \tau_2, \tau_3)$, $L_R = \ell_R$, because the right-handed neutrino is completely sterile. In order to switch from global to local $SU(2)_L \otimes U(1)_Y$ gauge invariance we have to introduce covariant derivatives ($\mathcal{D}_\mu$) in place of the ordinary derivatives. Covariant derivatives have the following form:

$$\mathcal{D}_\mu = \big(\partial_\mu + ig\vec{A}_\mu \cdot \frac{\vec{\tau}}{2} - \frac{1}{2}ig'(Y)B_\mu\big), \tag{1.12}$$

where $\vec{A}_\mu$ is a vector of three gauge fields needed to satisfy the local $SU(2)$ symmetry, while B_μ is the gauge field that guaranties the $U(1)$ local symmetry; g and g' are the coupling constants for the gauge fields; Y is a diagonal matrix with hypercharge values in its diagonal entries. Gauge fields introduced in the covariant derivatives are accompanied by the following kinetic terms, that need to be introduced in the lagrangian:

$$\mathcal{L}_{\rm EW}^{\rm boson} = -\frac{1}{4}\vec{E}_{\mu\nu} \cdot \vec{E}^{\mu\nu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{1.13}$$

where:

$$\begin{aligned} \vec{E}_{\mu\nu} &= \partial_\mu\vec{A}_\nu - \partial_\nu\vec{A}_\mu + g\big(\vec{A}_\mu \times \vec{A}_\nu\big), \\ F_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu. \end{aligned} \tag{1.14}$$

After inclusion of the covariant derivatives in Eq. (1.11) and after the following change of variables:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(A_{\mu 1} \mp iA_{\mu 2}), \\ Z_\mu &= \cos\theta_W A_{\mu 3} - \sin\theta_W B_\mu, \\ A_\mu &= \sin\theta_W A_{\mu 3} + \cos\theta_W B_\mu, \end{aligned} \tag{1.15}$$

(where $\tan\theta_W = \frac{g'}{g}$ defines the Weinberg angle) [9], two interaction terms arise, one involving vertices with net charge different from zero, that is the already mentioned

weak charged current interaction $\mathcal{L}_{\rm EW}^{\rm cc}$, and one involving vertices with zero net charge, known as weak neutral current interaction, $\mathcal{L}_{\rm EW}^{\rm nc}$:

$$\begin{aligned}\mathcal{L}_{\rm EW}^{\rm cc} &= \frac{g}{\sqrt{2}}\bar{L}_L\gamma^\mu\tau^+L_LW_\mu^+ + \frac{g}{\sqrt{2}}\bar{L}_L\gamma^\mu\tau^-L_LW_\mu^- \\ &= \frac{g}{\sqrt{2}}\big(J^{+\mu}W_\mu^+ + J^{-\mu}W_\mu^-\big),\end{aligned} \tag{1.16}$$

$$\begin{aligned}\mathcal{L}_{\rm EW}^{\rm nc} &= -i\Big(g\sin\theta_W J_\mu^3 + g'\cos\theta_W\frac{J_\mu^Y}{2}\Big)A^\mu \\ &\quad -i\Big(g\cos\theta_W J_\mu^3 - g'\sin\theta_W\frac{J_\mu^Y}{2}\Big)Z^\mu.\end{aligned} \tag{1.17}$$

The first row in Eq. (1.17) represents the electromagnetic part of the neutral current lagrangian, while the second row is the weak neutral current lagrangian. Thus, we can introduce the definition of the weak neutral conserved current $J_\mu^{\rm nc}$:

$$-i\frac{g}{\cos\theta_W}\Big(J_\mu^3 - \sin^2\theta_W J_\mu^{\rm em}\Big)Z^\mu = -i\frac{g}{\cos\theta_W}J_\mu^{\rm nc}Z^\mu, \tag{1.18}$$

thus:

$$J_\mu^{\rm nc} = J_\mu^3 - \sin^2\theta_W J_\mu^{\rm em}. \tag{1.19}$$

In conclusion, electromagnetic and weak neutral currents can be expressed as linear combinations of J_μ^3 and J_μ^Y. We can say that electromagnetism and weak neutral current are tightly bound and live in between the $SU(2)_L$ and the $U(1)_Y$ symmetries of the lagrangian.

1.2.3 Quark sector

Left-handed components of quarks are arranged in weak isospin doublets with $Y = \frac{1}{3}$

$$Q_L = \begin{pmatrix}U_L\\ D_L\end{pmatrix}, \tag{1.20}$$

and right-handed components are arranged into singlets

$$\begin{aligned}U_R, &\quad Y = +\tfrac{4}{3},\\ D_R, &\quad Y = -\tfrac{2}{3},\end{aligned} \tag{1.21}$$

where U and D are combinations of the mass eigenstates $u_i = u, c, t$ and $d_i = d, s, b$:

$$\begin{aligned}U_{L,R}^i &= \textstyle\sum_{j=1}^3 X(U_{L,R}^{ij})u_{L,R}^j,\\ D_{L,R}^i &= \textstyle\sum_{j=1}^3 X(D_{L,R}^{ij})d_{L,R}^j,\end{aligned} \tag{1.22}$$

where $X(U,D)_{L,R}$ are 3×3 matrices related to the Cabibbo-Kobayashi-Maskawa quark mixing matrix [10].

The expression of the quark sector electroweak lagrangian closely follows the one for leptons (Eq. (1.11)):

$$\mathcal{L}_{\mathrm{EW}}^{\mathrm{quark}} = i \sum_{i=1}^{3} \left(\bar{Q}_L^i \gamma^\mu \mathcal{D}_\mu Q_L^i + \bar{U}_R^i \gamma^\mu \mathcal{D}_\mu U_R^i + \bar{D}_R^i \gamma^\mu \mathcal{D}_\mu D_R^i \right) \tag{1.23}$$

with $\mathcal{D}_\mu$ from Eq. (1.12).

1.3 Higgs field and electroweak interaction mediators

Lagrangian densities described by Eqs. (1.11) and (1.23) do not contain mass terms for the fermions. This is in striking contrast with what is observed experimentally; a mass term in such lagrangian densities would break the $SU(2)_L \times U(1)_Y$ gauge symmetry, that is so effective in reproducing the observed electroweak currents.

This problem can be overcome enlarging the particle content of the Standard Model, introducing new fields, known as Higgs fields, organized in doublets, whose potential is invariant under $SU(2)_L \times U(1)_Y$ local gauge transformations.

The Higgs field lagrangian, in its global gauge invariant form, is written as

$$\mathcal{L}_{\mathrm{Higgs}} = (\partial_\mu \phi^\dagger)(\partial^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \tag{1.24}$$

where ϕ is a spin-$\frac{1}{2}$ spinor, μ is a complex parameter and λ is a real positive parameter. The corresponding hamiltonian density $\mathcal{H}$ is:

$$\mathcal{H} = (\partial_\mu \phi^\dagger)(\partial^\mu \phi) + V \tag{1.25}$$

where the potential V reads:

$$V = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{1.26}$$

The fundamental state of the system is obtained through the minimization of the potential V. If $\mu^2 > 0$ the fundamental state is $\phi = 0$ and it preserves all the symmetries of the lagrangian. If $\mu^2 < 0$, the derivative of the potential with respect to $\phi^\dagger \phi$ leads to the following minimum condition:

$$\mu^2 + 2\lambda \phi^\dagger \phi = 0 \tag{1.27}$$

Eq. (1.27) can be satisfied in infinite different ways because a global $SU(2) \otimes U(1)$ transformation leaves the $\phi^\dagger \phi$ product unaffected.

The lagrangian of Eq. (1.24) is not invariant under local $SU(2) \otimes U(1)$ transformations because the derivatives do not transform linearly under a transformation depending on x^μ. In order to achieve the local $SU(2) \otimes U(1)$ gauge invariance we

have to replace the ordinary derivatives with covariant derivatives ($\mathcal{D}_\mu$); the covariant derivatives are built in such a way that they transform linearly under an $SU(2) \otimes U(1)$ transformation

$$\mathcal{D}_\mu \phi \to U \mathcal{D}_\mu \phi. \tag{1.28}$$

This can be achieved through the introduction of new fields (gauge fields); in this case the covariant derivative reads:

$$\mathcal{D}_\mu = \partial_\mu + ig\vec{A}_\mu \cdot \frac{\vec{\tau}}{2} + \frac{1}{2} ig' B_\mu. \tag{1.29}$$

The Higgs lagrangian, $SU(2) \otimes U(1)$ locally invariant by construction, is:

$$\mathcal{L}_{\text{Higgs}} = (\mathcal{D}_\mu \phi)^\dagger (\mathcal{D}^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 - \frac{1}{4} \vec{E}_{\mu\nu} \cdot \vec{E}^{\mu\nu} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{1.30}$$

where $\vec{E}_{\mu\nu}$ and $F_{\mu\nu}$ are defined in Eq. (1.14); the gauge fields kinetic terms are included in Eq. (1.30) as well.

As in the globally invariant version of the lagrangian, if $\mu^2 < 0$ a degenerate fundamental state is found. By choosing a particular vacuum state we break the symmetry. We shall choose a fundamental state with form

$$\phi_0 = \begin{pmatrix} 0 \\ \eta \end{pmatrix}, \qquad \eta = \sqrt{-\frac{\mu^2}{2\lambda}}. \tag{1.31}$$

Let's apply an x dependent perturbation to the vacuum state:

$$\phi = \begin{pmatrix} 0 \\ \eta + \frac{\sigma(x)}{\sqrt{2}} \end{pmatrix}. \tag{1.32}$$

Expanding the lagrangian density in Eq. (1.30) around the ground state and using Eq. (1.15) the following expression for the lagrangian density is obtained:

$$\begin{aligned} \mathcal{L}_{\text{Higgs}} \;=\; & \frac{1}{2} \partial_\mu \sigma \partial^\mu \sigma + \mu^2 \sigma^2 + \\ & - \frac{1}{4} A_{\mu\nu} A^{\mu\nu} + \\ & - \frac{1}{4} (W^{+\dagger}_{\mu\nu} W^{\mu\nu +} + W^{-\dagger}_{\mu\nu} W^{\mu\nu -}) + \frac{g^2 \eta^2}{4} (W^{+\dagger}_{\mu} W^{\mu +} + W^{-\dagger}_{\mu} W^{\mu -}) + \\ & - \frac{1}{4} Z_{\mu\nu} Z^{\mu\nu} + \frac{g^2 \eta^2}{4 \cos^2 \theta_W} Z_\mu Z^\mu + \\ & + \textit{interaction terms}, \end{aligned} \tag{1.33}$$

where $A_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu$ and so on.

The first line in Eq. (1.33) represents the Higgs boson scalar fields, with mass $\sqrt{-2\mu^2}$; the second line represents a massless field identified with the electromagnetic field; the third line represents $W^\pm$ fields, with mass $g\eta/\sqrt{2}$; finally, the fourth line represents the Z field with mass $g\eta/(\sqrt{2}\cos\theta_W)$.

1.3.1 Fermion mass terms

We shall now see how mass terms for the fermions can arise through the interaction with the Higgs field, without breaking the $SU(2)_L \otimes U(1)_Y$ symmetry. Introducing an interaction term of the fermions with the Higgs field, in the context of spontaneous symmetry breaking, makes it possible to assign masses to the fermions. This can be done via a Yukawa coupling with coupling constant G_f with form:

$$\mathcal{L}_{\text{mass}}^{\text{fermion}} = -G_f\Big(\bar{F}_L\phi F_R + \bar{F}_R\phi^\dagger F_L\Big), \tag{1.34}$$

where F_L and F_R are fermion's helicity eigenstates.

For the leptons Eq. (1.34) reads:

$$\mathcal{L}_{i\text{mass}}^{\text{lepton}} = -G_\ell\Big(\bar{L_L}\phi L_R + \bar{L_R}\phi^\dagger L_L\Big), \tag{1.35}$$

which can be divided in two terms using Eq. (1.32) for the Higgs field and Eq. (1.3).

$$\mathcal{L}_{\text{mass}}^{\text{lepton}} = -G_\ell\eta\Big(\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L\Big) - \frac{G_\ell}{\sqrt{2}}\Big(\bar{\ell}_L\ell_R + \bar{\ell_R}\ell_L\Big)\sigma(x), \tag{1.36}$$

G_ℓ being the Yukawa coupling constant for the lepton family.

The mass term for the lepton is then:

$$m_\ell = G_\ell\eta. \tag{1.37}$$

1.4 Strong interactions

Out of the quark model proposed by Gell-Mann [11] in 1964, the idea of the "colour" quantum number was proposed by Han and Nambu [12] in 1965 to avoid the apparent paradox that the quark model seemed to require a violation of the Pauli exclusion principle to describe hadron spectroscopy. Quantum Chromo Dynamics (QCD) was then quantized as a gauge theory with $SU(3)_c$ symmetry in 1973 by Fritzsch [13], Gross and Wilczec [14], Weinberg [15].

QCD coupling constant ranges over several orders of magnitude when moving from hard, i.e. large momentum transfer processes, to soft processes. Its value grows as the momentum transfer decreases. This effect is known as asymptotic freedom, and it justifies the use of perturbation theory (perturbative QCD or pQCD) when describing hard processes. At small energies (large distances), where the value of the coupling constant becomes large, the theory behaves in a non-perturbative way; in such a regime the isolated quark or gluon cross sections vanish and are replaced by bound state dynamics. This effect is known as "confinement" and it justifies the non-observation of free quarks and gluons.

1.4.1 QCD lagrangian

QCD is an $SU(3)_c$ gauge theory whose lagrangian is written in the following form:

$$\mathcal{L}^{QCD} = \mathcal{L}_{\text{invar}} + \mathcal{L}_{\text{gauge fix}} + \mathcal{L}_{\text{ghost}} \tag{1.38}$$

$\mathcal{L}_{\text{invar}}$ is invariant under local $SU(3)_c$ transformations and reads:

$$\mathcal{L}_{\text{invar}} = \sum_f \bar{\psi}_f \big(i\gamma_\mu \mathcal{D}^\mu - m_f\big)\psi_f - \frac{1}{4} F_{\mu\nu}F^{\mu\nu} \tag{1.39}$$

where f runs over the six quark fields, $\mathcal{D}^\mu$ is the covariant derivative:

$$\mathcal{D}_\mu = \partial_\mu + igA_{\mu a}T_a \tag{1.40}$$

and

$$F_{\mu\nu} = \partial_\mu A_{\nu a} - \partial_\nu A_{\mu a} - gC_{abc}A_{\mu b}A_{\nu c} \tag{1.41}$$

where $A_{\mu a}$ are the fields of the eight colored gluons, T_a are the eight generators of $SU(3)$, C_{abc} are the structure constants that define the commutation rules of the $SU(3)$ generators.

$\mathcal{L}_{\text{gauge fix}}$ and $\mathcal{L}_{\text{ghost}}$ in Eq. (1.38) are needed for technical reasons connected to how the quantization of the QCD lagrangian is performed [16].

1.4.2 UV divergences and renormalization

As many theories QCD suffers from divergent loop integrals. This ultraviolet (UV) divergences need to be regulated so that cross sections can be calculated. We shall now see how the regularization takes place and how a theory with an arbitrary parameter (the regularization parameter) can be predictive.

When treating loop diagrams (Fig. 1.1) integrals of the following form are involved:

$$\Gamma^{un}(p) = \int \frac{d^4k}{(2\pi)^4} \frac{1}{(k^2 - M^2(p))^2} \tag{1.42}$$

where $M(p)$ is a function of the incoming momentum.

The divergent integral is replaced by a finite one by limiting the momentum in the loop to a cutoff μ. In this way the divergent integral becomes:

$$\Gamma^{un}(p) \to \Gamma^{ren}(p,\mu) = -\frac{i}{(4\pi)^2} \ln\Big(\frac{M^2(p)}{\mu^2}\Big) \tag{1.43}$$

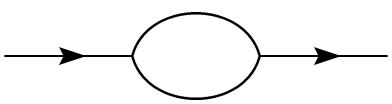


Figure 1.1: A loop diagram.

A question may arise on how predictive a theory can be if it depends on an arbitrary parameter. Suppose that we want to measure a physical observable $O(p)$, where p is a set of momenta, and we have a theoretical expression of O, which relies on a coupling $g(\mu)$ through a perturbation series with coefficients $a_i(p, \mu)$,

$$O(p) = \sum_{i=0}^{N} a_i(p, \mu) g(\mu)^{2i}, \tag{1.44}$$

where N is the highest order in perturbation theory that we were able to compute. All we have to do is to measure O, fix μ to whatever value we like and solve Eq. (1.44) to find $g(\mu)$. At this point, at the price of one measurement, we can predict O for every other value of p.

Of course, since O is a physical quantity it cannot depend on the choice of μ. As we shall see in the next section to require that O does not depend on μ defines the "running" of the strong coupling constant.

1.4.3 Running α_S: asymptotic freedom and confinement

Let's consider a physical observable O, whose value depends on a scale Q, e.g. the momentum transfer in a scattering process. As pointed above when we calculate O as a series in the coupling constant α_S we need to introduce a renormalization scale μ, to make integrals calculable. In general, O is a function of Q^2/μ^2, and of α_S itself. To require that O does not depend on our choice of μ is equivalent to the following equation:

$$\mu^2 \frac{\mathrm{d}O(Q^2/\mu^2, \alpha_S(\mu^2))}{\mathrm{d}\mu^2} = 0. \tag{1.45}$$

It can be shown [17] that solutions of Eq. (1.45) have form $O(1, \alpha_S(Q^2))$, which means that the whole scale dependence of the observable O is embedded in the running value of the coupling constant.

This means that O can be calculated at a fixed order in perturbation theory and the running of α_S predicts how O changes with the scale.

The perturbative expansion of α_S is such that

$$\frac{\mathrm{d}\alpha_S(Q^2)}{\mathrm{d}Q^2} < 0, \tag{1.46}$$

which means that the value of the coupling constant decreases at higher scales. This is one of the most peculiar characteristics of QCD with respect to QED, where exactly the opposite holds. In QED the coupling decreases with decreasing scales and increasing distances. This is intuitively interpreted as a screening of the bare electric charge due to fermion loops, and is anyway quite a little effect. On the contrary in QCD the value of the coupling constant is tiny at high energies (small distances) and becomes very big as the scale decreases.

The fact that the coupling constant is small at high energies is often referred to as *asymptotic freedom*, and it justifies the success of the perturbative approach (pQCD) in describing hard processes.

On the other hand the fact that the coupling becomes so strong as the distance grows justifies the *confinement*, that is to say the fact that in our experiments we always observe bound quark states, and never free quarks or gluons; qualitatively, this happens because as the partons produced in the hard interactions go apart from each other it becomes easier to produce quark antiquark pairs, that recombine into hadrons, than to keep pulling against a growing force.

The expression of $\alpha_S(Q^2)$ follows [16]:

$$\frac{\alpha_S(Q^2)}{4\pi} = \frac{1}{\beta_1 \ln(Q^2/\Lambda^2_{QCD})} - \frac{\beta_2 \ln(\ln(Q^2/\Lambda^2_{QCD}))}{\beta_1^3 \ln^2(\ln(Q^2/\Lambda^2_{QCD}))} + O\Big(\frac{1}{\ln^3(\ln(Q^2/\Lambda^2_{QCD}))}\Big) \tag{1.47}$$

with β functions defined in [16]. Λ_{QCD} is often regarded as the parameter at which the interaction becomes strong.

Experiments usually measure α_S at a certain scale, usually the mass of the Z boson; Eq. (1.47) then predicts the value of α_S at any other scale. A recent review of α_S measurements at the Z mass for different processes is shown in Fig. 1.2.

The running of α_S as measured in experiments is shown in Fig. 1.3.

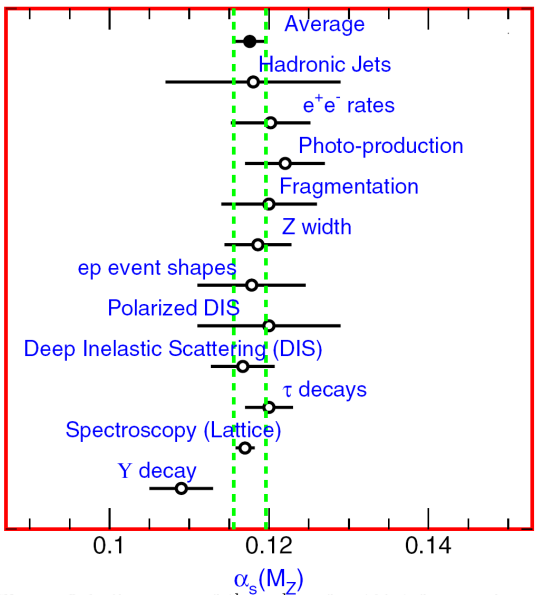


Figure 1.2: A recent summary of α_S measurements [18].

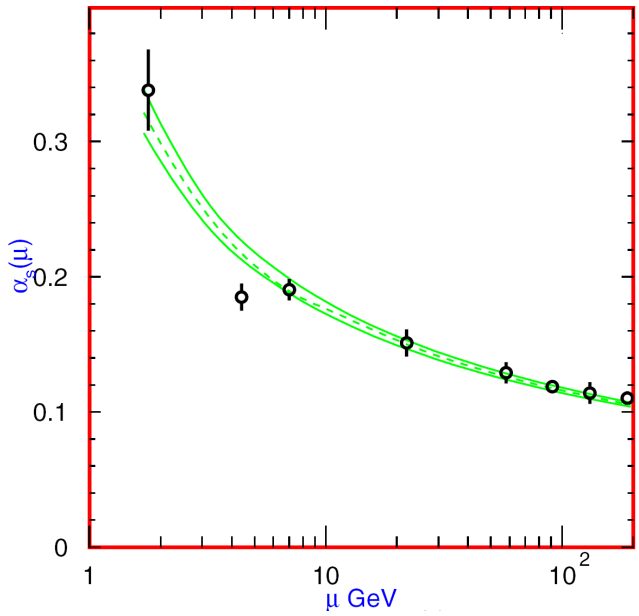


Figure 1.3: Running of α_S [18]. α_S values are plotted as a function of the scale at which they were measures. Dotted line is the best fit. $\pm 1\sigma$ solid lines are also shown.

1.4.4 Infrared and collinear safe observables

One of the most remarkable successes of pQCD is the prediction of jet inclusive cross section in e^+e^- collisions. Jets are collimated sprays of hadrons. Even if hadrons are the result of a non perturbative process involving the quarks and gluons produced in the hard process, pQCD is very good at predicting jet cross section. The reason is that quarks and gluons are produced in the "bulk" of the process, and involve a high momentum transfer. Quarks and gluons originated in the hard process then undergo the non-perturbative hadronization process, but this happens at much lower energies (and much higher distances). Hadronization happens too late to modify substantially the topology of the event. This is an example of the "factorization" properties of QCD cross section calculation that we will see in the next section.

Let's consider the e^+e^- collider case. The lowest order process contributing to the inclusive jet cross section is $e^+e^- \to q\bar{q}$, with a γ or a Z boson exchanged in the s-channel. The leading order cross section, in case of γ exchange is [17]

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta} = 3\sum_q \frac{\pi\alpha^2 Q_q^2}{2s}(1+\cos^2\theta), \tag{1.48}$$

where θ is the emission angle of the q line with respect to the direction along the beams, Q_q is the charge of the quark and s is the squared center of mass energy.

When moving to the next to leading order we have to account for possible additional gluon emission out of the quark or antiquark lines. The corresponding cross section, considering quarks and gluons on the mass shell, is conveniently expressed if we introduce energy fractions x_i

$$x_i = \frac{2E_i}{\sqrt{s}}, \qquad i = q, \bar{q}, g. \tag{1.49}$$

With this notation the cross section for emission of an additional gluon is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x_1 \mathrm{d}x_2} \propto \alpha_S \frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)}. \tag{1.50}$$

Since $1 - x_1 = x_2 E_3(1 - \cos\theta_{2,3})/\sqrt{s}$ and $1 - x_2 = x_1 E_3(1 - \cos\theta_{1,3})/\sqrt{s}$, where $1, 2$ denote q and $\bar{q}$ respectively, and 3 denotes the gluon, Eq. (1.50) is divergent when either the gluon is soft or when it is collinear with the line from which it originates.

These divergences are not of course physical, but only due to the fact that we have considered only next to leading order real emission contributions. The recipe to cure these divergences is to use a regularization procedure, introducing a cutoff: in this way divergences are replaced with large logarithms, function of the cutoff. Then, taking into account also the virtual diagrams completely cancels divergences at NLO.

Divergences like the ones encountered here always arise whenever we add a real emission line. If a quantity is free of such divergences it is called a *soft-collinear safe* quantity. To make it possible to compare observables measured from experiments with theoretical prediction it is important to define observables in a *soft-collinear safe* fashion.

An observable $O(p_1, p_N)$ function of N measured momenta p_i is soft safe if

$$O(p_1, \ldots, p_N) = O(p_1, \ldots, p_N, \epsilon), \qquad \text{where } \epsilon^2 \text{ is small}, \tag{1.51}$$

and is collinear safe if

$$O(p_1, \ldots, p_i, \ldots, p_N) = O(p_1, \ldots, p_{i1}, p_{i2}, \ldots, p_N), \qquad \text{where } p_{i1} + p_{i2} = p_i. \tag{1.52}$$

For what concerns jets, a soft-collinear safe jet measure has to be such that the number of jets in an event must remain unchanged if a soft particle is added or if the four-momentum of a particle is split into two. An example of such a measure for e^+e^- collisions is the JADE [19] jet measure. When running on an N particle final state the JADE algorithm identifies two particles i, j as two jets if their invariant mass M_{ij} is such that:

$$\frac{M_{ij}^2}{s} > y_{cut} \tag{1.53}$$

where $\sqrt{s}$ is the center of mass energy, and y_{cut} is a cut chosen according to the user needs. If two particles do not satisfy Eq. (1.53) they are recombined into a single

object summing their four-momenta, and the algorithm is repeated with the new object.

In case of massless particles Eq. (1.53) reduces to

$$\frac{M_{ij}^2}{s} = \frac{2E_iE_j(1-\cos\theta_{ij})}{s} > y_{cut} \tag{1.54}$$

where E_i and E_j are the energies of particle i and j respectively, and θ_{ij} is the angle between the two. From Eq. (1.54) it is clear that the regions of soft divergences ($E_{i,j} \to 0$) and the regions of collinear divergences ($\theta_{ij} \to 0$) are cut away.

A variant of the JADE algorithm that is often used in hadron collision, and that shares with JADE the soft and collinear safety is the longitudinally invariant k_t algorithm. In this algorithm, originally proposed in [20], two distances are calculated for a set of N particles:

$$d_{iB} \equiv p_{T,i}^2 \tag{1.55}$$

and

$$d_{ij} \equiv min\{p_{T,i}^2, p_{T,j}^2\}\left(\frac{\Delta R^2}{D^2}\right) \tag{1.56}$$

where $p_{T,i}$ is the transverse momentum of particle i, $\Delta R^2 = \Delta y^2 + \Delta\phi^2$ (y is the particle rapidity, ϕ the azimuthal angle) and D is a parameter, usually of the order of 1.

A cut value d_{cut} is chosen and particles with $d_{ij} < d_{cut}$ are recombined summing their four-momenta (also other recombination techniques can be used). Particles with $d_{iB} < d_{cut}$ belong to the so called beam-jets, and are removed from the list. The procedure is iterated until all the d_{iB} and d_{ij} are above d_{cut}. This version of the k_t algorithm is often referred to as the *exclusive* k_t algorithm.

An *inclusive* variant [21] exists in which no d_{cut} is present. In the *inclusive* k_t algorithm the values of d_{iB} and d_{ij} are ordered from the minimum to the maximum and a d_{min} is found. If d_{min} if of type d_{iB} then i is removed from the list and assigned to the list of jets. If d_{min} is of type d_{ij} then i and j are recombined in a pseudo-jet. The procedure is repeated until no pseudo-jets remain. In this way it is guaranteed that final jets will be separated in angle by at least D (defined in Eq. (1.56)).

It is worth noticing that the inclusive $k_\perp$ algorithm as presented above is soft unsafe, because a soft particle well separated from anything else would give rise to an additional jet. When running the $k_\perp$ algorithm in inclusive mode a minimum p_T cut for jets needs to be applied to recover soft safety.

The $k_\perp$ algorithm is not the only way to build jets. In many studies and experiments cone algorithms were extensively used.

While in $k_\perp$ algorithm jets are built in an iterative way, progressively recombining particles, cone algorithms are not based on progressive recombination techniques; the basic idea of most of cone algorithms is to identify "seed" directions, along high p_T particles and draw cones with fixed radius around these directions.

The position of each cone, and thus its particle content, is then adjusted according to the particle content until a "stable cone" is found, that is a cone whose position is stable. Stable cones are then turned into jets: this is done with different techniques in the various cone algorithm available; what changes is mainly the way in which overlapping cones are treated.

Most cone jet clustering algorithms (like for example JetClu, MidPoint [21]) suffer from infrared or collinear unsafety due to the use of seeds to start the search of stable cones.

In a nutshell, the reason why the use of seeds makes these algorithms unsafe, is that the algorithms always starts from the hardest seed and then proceeds towards the softest. If a particle used as seed is replaced with a collinear pair it may not be a seed anymore. It can be shown that this leads to a different sets of final jets, thus confirming the collinear unsafety of the algorithm (details can be found in [22]).

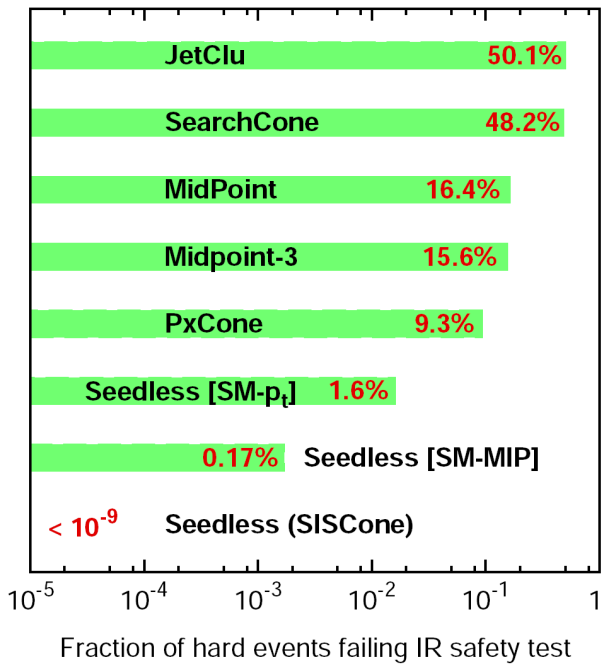


Figure 1.4: Failure rate in an infrared safety test for various cone algorithms.

Only seedless implementation of cone algorithms, like SISCone [22] proved to be infrared and collinear safe. Fig. 1.4 shows the failure rate of several cone algorithms in a infrared safety test in which soft particles were added randomly to the initial set of particles.

1.4.5 Factorization theorems: Parton density functions

The cross section for a $2 \to N$ process at an hadronic collider is conveniently expressed as

$$d\sigma_{pp\to N} = \int_0^1 dx_1 \int_0^1 dx_2 f_1(x_1, \mu_F^2) f_2(x_2, \mu_F^2) d\hat{\sigma}_{pp\to N}(\mu_F^2) \tag{1.57}$$

In this expression $\hat{\sigma}$ is the parton level cross section, x_1 and x_2 are the momentum fraction of the proton momentum carried by the two colliding partons; $f_{1,2}$ are the parton density functions (PDFs), that describe the probability that a parton carries momentum fraction $x_{1,2}$; μ_F is the factorization scale, that is the scale at which the separation between the hard perturbative interaction and the long distance, non-perturbative, evolution of the produced partons takes place.

The PDFs for quarks and gluons at a scale μ^2=10 GeV^2 scale as calculated by the MRST [23] collaboration in 2004 are shown in Fig. 1.5.

PDFs evolution with scale is governed by the DGLAP equation [24], as long as $\alpha_S(Q)$ remains in the perturbative validity region. DGLAP equation allows global fits of a variety of data taken from different experiments, at different scales. Two collaborations are the main provider of global PDFs fits, CTEQ [25] and MRST [23].

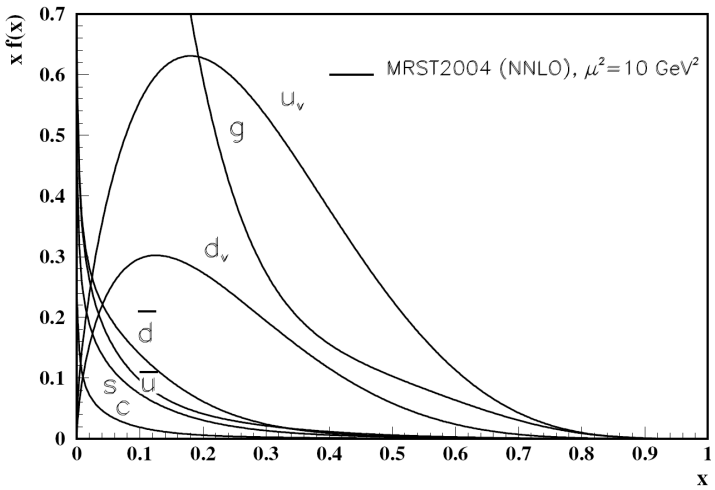


Figure 1.5: Distribution of $xf(x)$ as a function of the momentum fraction x at μ^2=10 GeV^2 for different partons [18].

Chapter 2

The CMS experiment at LHC

The Large Hadron Collider (LHC) [26] is the most powerful hadron collider running in the next two decades. Its installation has finished in July 2008. LHC will investigate processes with really tiny cross sections, down to the femtobarn. The main reason that drove the choice of a hadron collider instead of an electron collider like LEP [27–29] was the need to build a machine capable of reaching center of mass energies much higher than LEP, to be housed in the LEP tunnel. This can be achieved with a hadron machine thanks to the lower amount of synchrotron radiation emitted by circulating hadrons. Hadron colliders, despite the production of many low energy particles in a complex environment if compared to electron-positron events, provide access to a wider energy spectrum, which in addition can be explored simultaneously.

In this chapter I will briefly review the main characteristics of the LHC and I will describe the LHC experiment I am involved in, the Compact Muon Solenoid (CMS) [30].

2.1 The Large Hadron Collider

The LHC accelerator has been installed in the underground tunnel which housed the LEP electron-positron collider until 2000. A schematic description of the LHC accelerator complex and its services is shown in Fig. 2.1. LHC will have two counter circulating proton beams, accelerated at 7 TeV in a 27 km ring, resulting in a total center of mass energy of 14 TeV. The two beams will collide in four interaction points; four experiments are built around the interaction points. Two general purpose experiments, called ATLAS [31] and CMS [30], will do general Standard Model measurements and will seek new physics; one experiment called LHCb [32] is dedicated to the B meson physics and it will carry out precise measurements of CP violation; one experiment called ALICE [33] will investigate heavy ion physics (Pb ions will be accelerated in a later phase of LHC operation).

LHC can be regarded as a discovery machine with an extremely wide energy dynamic range, being able to investigate mass scales from order of few GeV, as

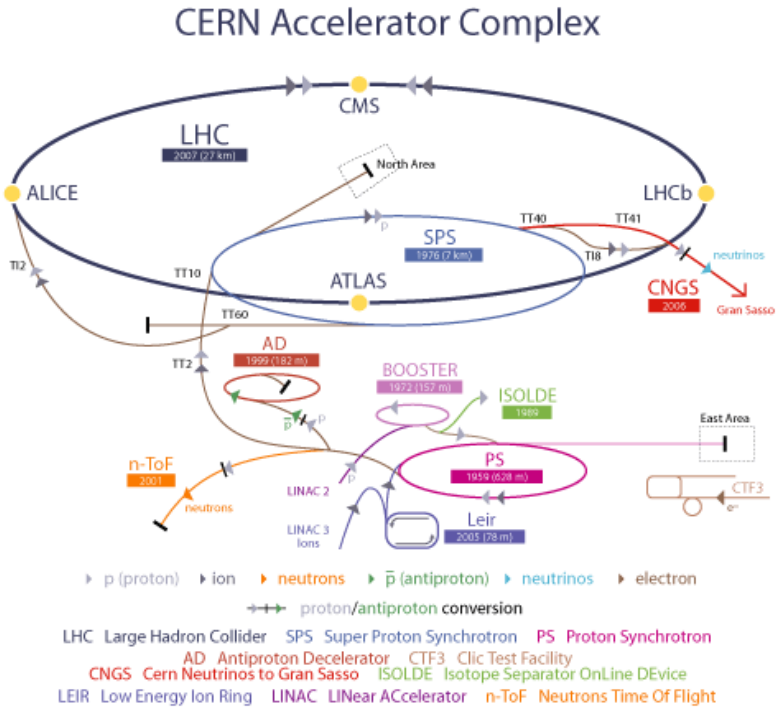


Figure 2.1: The LHC accelerator complex. The location of the experiments along the ring is indicated.

in the case of B-meson physics, up to a few TeV, for the discovery of new vector bosons or quark compositeness.

In order to extend the LHC capability to explore new physics rare processes an enormous effort has been made to raise the proton momentum as much as possible. In particular, a very sophisticated magnet system is needed to keep such high momentum protons in the machine orbit. The formula that connects the bending radius with the particle's momentum and the magnetic field is:

$$B[\mathrm{T}] = \frac{\mathrm{p}[\,\mathrm{GeV}]}{0.3\rho[\mathrm{m}]} \tag{2.1}$$

where B is the magnetic field in Tesla, p the momentum in GeV, ρ the orbit radius in metres. For a circumference of about 27 km, the magnetic field needed for 7 TeV protons is about 5.4 T. Actually, since LHC is made of curved and rectilinear sections, the bending magnetic superconductor dipoles need to produce an 8.3 T magnetic field. This value is close to the technological edge for superconducting magnets nowadays.

Since the beam energy is limited by the bending power of the magnetic system and by the circumference of the machine, another handle to raise the rate of interesting and rare events is the luminosity $\mathcal{L}$. The event rate n for a process with cross section σ is

$$n = \mathcal{L}\sigma \tag{2.2}$$

The luminosity is connected to the beam properties with the following approximated formula [18]:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{2.3}$$

where n_1 and n_2 are the number of particles in beam 1 and 2 respectively, f is the collision frequency, σ_x and σ_y are transverse dimensions of the beams. The proton bunches at LHC will collide at a frequency of about 40 MHz, corresponding to a spatial separation between bunches of about 7.5 m. The frequency cannot be raised further, because of the limiting requirement of avoiding further collisions on the side of each interaction region. The transverse dimensions of the beam can be squeezed down to 15 μm.

In the startup period (the first six months of operation approximately) the luminosity will be $2 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The luminosity will then be raised gradually to $1 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. In this condition an integrated luminosity of $6\,\mathrm{fb}^{-1}$ should be collected during 2009. The luminosity will be gradually raised, and it will reach about $1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ in 2012. The total integrated luminosity in 2012 should be about 100 fb^{-1}. In 2016 a luminosity of $3 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ is foreseen. The integrated luminosity at this point is expected to be about 650 fb^{-1}.

The need for such a high luminosity has driven the choice of a proton-proton collider, instead of a proton-antiproton. In fact, even if a proton-antiproton machine has the advantage that both beams can be kept in the same beam-pipe, to produce the number of antiprotons needed to reach the desired luminosity is an unfeasible task. Table 2.1 describes the main design characteristics of LHC.

In the hard proton proton collision, with high transferred momentum, the center of mass energy $\sqrt{\hat{s}}$ is connected to the total center of mass energy $\sqrt{s}$ as:

$$\sqrt{\hat{s}} = \sqrt{x_1 x_2 s} \tag{2.4}$$

where x_1 and x_2 are the energy fractions of the two partons participating in the hard scattering.

The center of mass of the two hardly interacting partons is not motionless in the experiment frame, but rather it is on average boosted along the direction defined by the colliding beams. For this reason boost invariant observables ave very important to characterize the event. One of such observables is the transverse momentum p_T, defined as the projection of the momentum vector on a plane perpendicular to the beam axis.

Another useful observable is the rapidity y

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} = \tanh^{-1}\left(\frac{p_z}{E}\right) \tag{2.5}$$

where E is the particle's energy, p_z the projection of particle's momentum along the beam direction. Under a boost along z with speed β, y undergoes the following transformation: $y \to y - \tanh^{-1}\beta$, hence rapidity differences are invariant, thus the shape of the rapidity distribution $\mathrm{d}N/\mathrm{d}y$ is invariant.

Circumference	26.659 km	
Maximum Dipole field	8.33 T	
Magnetic Temperature	1.9 K	
	p − p	$^{82}_{208}$Pb-$^{82}_{208}$Pb
Beam energy at injection	450 GeV	73.8 TeV
Beam energy at collision	7 TeV	574 TeV(2.76 A TeV)
Maximum Luminosity	1×10^{34} cm^{-2}s^{-1}	2×10^{27} cm^{-2}s^{-1}
Number of Bunches	2 808	608
Bunch spacing	7.48 cm	5.3 cm
Bunch separation	24.95 ns	124.75 ns
Number of particles per bunch	1.1×10^{11}	8×10^{7}
Total crossing angle	300 μrad	$<$100 μrad
Bunch Length (r.m.s.)	7.5 cm	7.5 cm
Transverse beam size at Impact Point	15 μm	15 μm
Luminosity lifetime	10 h	4.2 h
Filling time per ring	4.3 min	9.8 min
Energy loss per turn	7 keV	
Total radiated power per beam	3.8 kW	
Stored energy per beam	350 MJ	

Table 2.1: Technical parameters of LHC for p-p and Pb-Pb collisions.

In the ultrarelativistic approximation the rapidity y is the same as the pseudo-rapidity η defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.6}$$

It's often useful to refer to pseudorapidity as it depends only on the direction of the three-vector.

2.2 The CMS detector

The Compact Muon Solenoid experiment (CMS) [30] is a general purpose LHC experiment. Its main feature is the 4T superconducting solenoidal magnet; such a strong magnetic field permits a compact design of the apparatus. The main design priorities of CMS were a redundant muon tracking system, a good electromagnetic calorimeter and a high quality inner tracking system.

The structure of CMS is typical for general purpose collider detectors. It consists of several cylindrical detecting layers, coaxial with the beam direction (*barrel* region), closed at both ends with disks (*endcap* region).

Figs. 2.2 and 2.3 show two schematic views of the CMS detector, that has a full length of 21.6 m, a diameter of 15 m, and a total weight of 14500 tons.

The coordinate frame used in CMS is a right-handed tern, with the x axis pointing towards the LHC centre, y axis directed upward along the vertical and z axis along the beam direction with the direction required to complete the right-handed tern. The cylindrical symmetry of CMS design and the invariant description of proton-proton physics suggest the use of a pseudo-angular reference frame, given by the triplet (r, ϕ, η), where r is the distance from the z axis, ϕ is the azimuthal angle, measured starting from the x axis positive direction, η is defined in Eq. (2.6) (where θ is the polar angle).

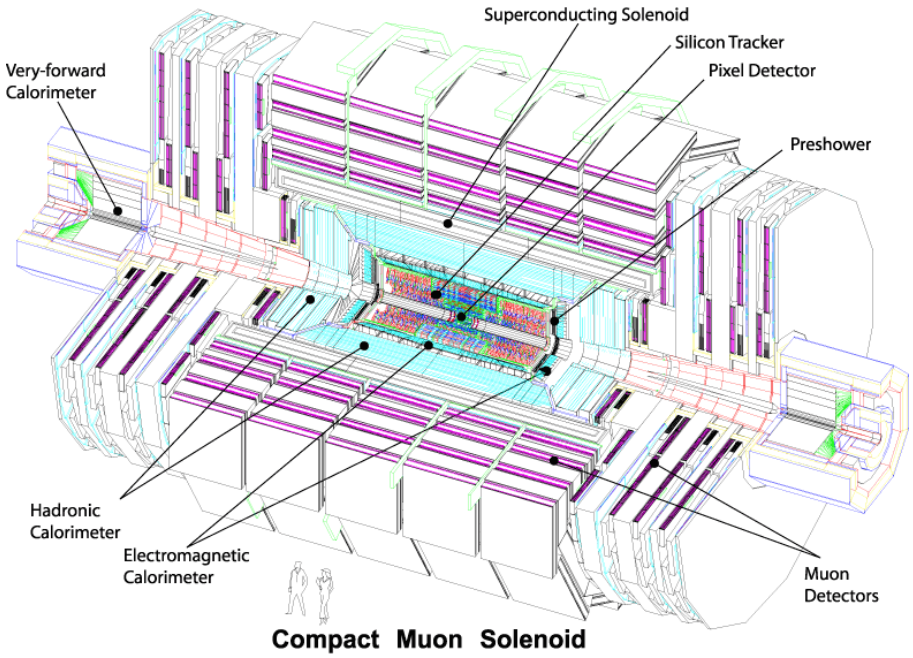


Figure 2.2: A view of the CMS detector with its subdetectors labeled.

CMS is made up of four main subdetectors:

- Silicon Tracker: it is made of a Silicon Pixel vertex detector and a surrounding Silicon Microstrip detector, with a total active area of about 215 m^2. It is used to reconstruct charged particle tracks and vertices.

- ECAL: it is an electromagnetic calorimeter to precisely measure electrons and photons.

- HCAL: it is a hadronic calorimeter for jet direction and energy measurement.

- Muon System: it is a composite tracking system for muons. It consists of Cathode Strip Chambers (CSC) in the barrel region and Drift Tube (DT) in the endcaps. A complementary system of Resistive Plate Chambers (RPC) is used both in the barrel and in the endcaps.

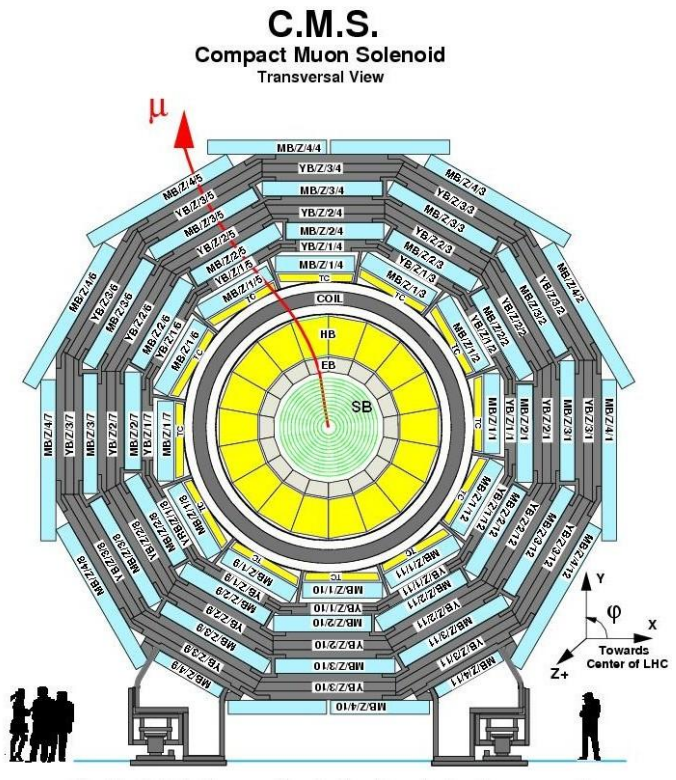


Figure 2.3: A transverse view of the CMS detector.

The Silicon Tracker, ECAL and HCAL are located inside the magnetic coil. Muon Chambers are located in the magnet return yoke. In the following sections a brief description of each component is given.

2.2.1 Magnet

CMS magnet [34] is the biggest superconducting magnet ever built. It is able to generate a 4T magnetic field in a huge volume. The magnet is made of five modules, 6 m diameter, 2.5 m length, 50 tons weight each. Spires in each module are made of a Niobium-Titanium compound, which has superconducting properties, and they are kept at 4K during operation with a liquid helium cooling system. A current of 20 kA flows in the spires during operation.

A 12000 tons weight iron yoke is built around the magnet to bridle the magnetic field lines. The yoke consists of a barrel region, made of five rings, and two endcaps consisting of three disks each.

Thanks to the intense strength of the field, a precise measurement of charged particles momenta is possible. Besides, the field in the return yoke, where muon chambers are located, permits an independent measure of muons momentum.

2.2.2 Tracker

The Silicon Tracker [35, 36] is the CMS innermost detector. It consists of a Silicon Pixel detector and a surrounding Silicon Microstrip detector.

It covers the region $|\eta| < 2.4$, $r < 120$ cm. Its goal is to provide a precise momentum estimate for charged particles, and to allow a precise determination of the position of secondary vertices. LHC events will be very complex, and track reconstruction comes as a complex pattern recognition problem. In order to ease pattern recognition two requirements are fundamental:

- low detector occupancy,
- large hit redundancy.

The low hit occupancy is achieved with a highly granular detector, while the redundancy is achieved with a large number of detecting layers.

The pixel detector is made of three barrel layers and two endcap disks per side (Fig. 2.4). The overall number of readout channels is about 60 millions and it covers the region with 4.4 cm $< r <$ 10.2 cm, $|z| <$ 47 cm. The high granularity of the detector permits an efficient separation of different track segments.

The Silicon Strip Tracker (SST) is made of ten barrel layers and twelve endcap disks on each side. It has about 10 millions readout channels. The main components of the SST are shown in Fig. 2.5. As indicated in Fig. 2.5 some of the layers are equipped with *single sided* detectors, some with *double sided* detectors. Single sided detectors can provide the particle's impact point position in the direction perpendicular to the strips. Double sided detectors can provide both coordinates on the detector surface, as they are made with two single sided detectors glued back-to-back with an angle of 100 mrad between the strips directions. Inner layers

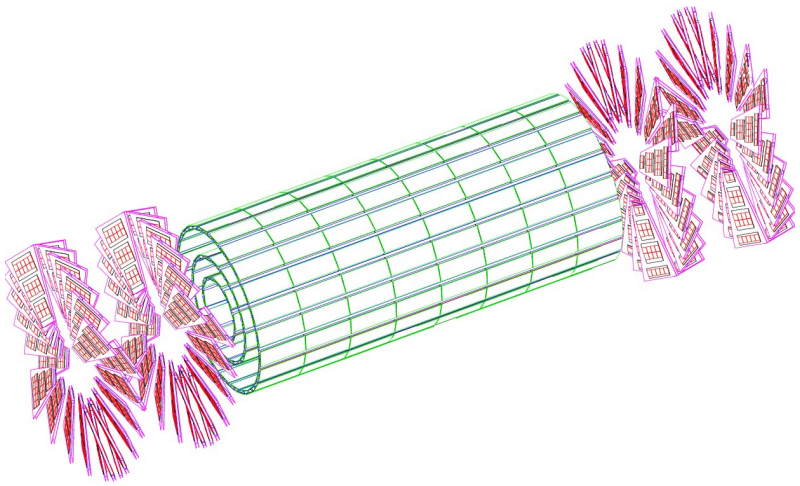


Figure 2.4: A schematic view of the pixel vertex detector.

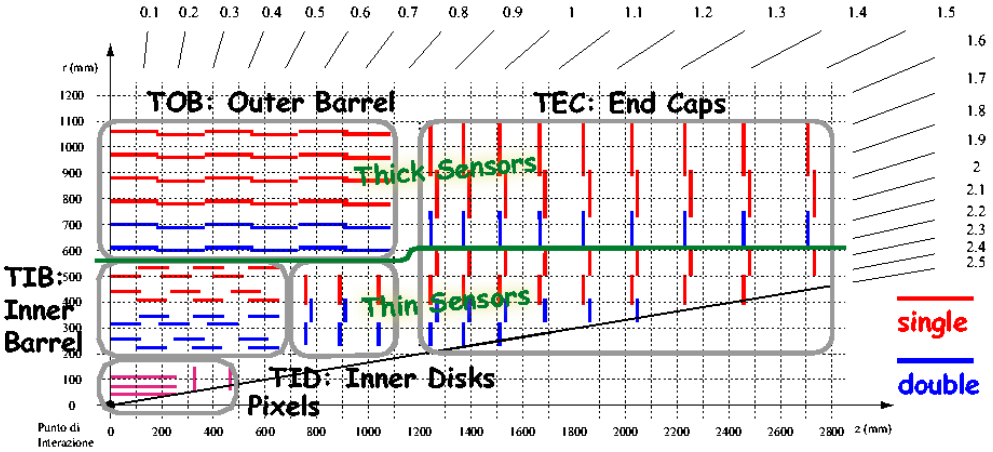


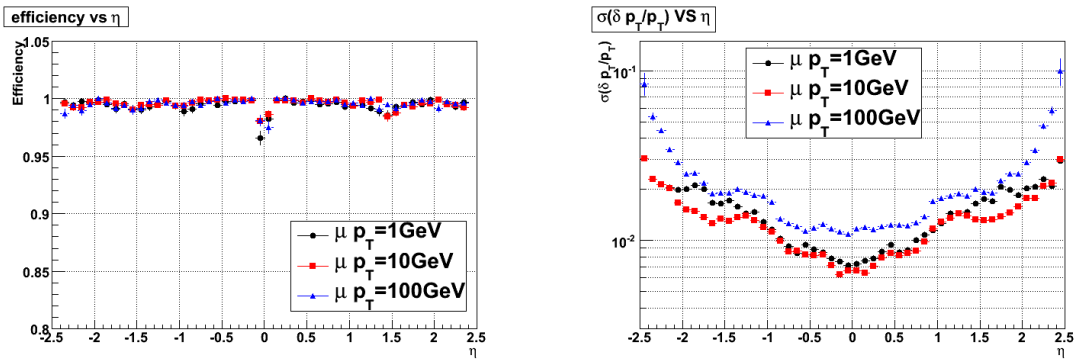
Figure 2.5: An $r-z$ schematic view of a sector of the Silicon Strip Tracker. The location of single sided and double sided detectors is put into evidence.

are equipped with 300 μm thick sensors, while outer layers are equipped with 500 μm thick sensors.

The high flux of radiation through the tracker sensors causes damages. Pixel and microstrip detectors and readout electronics are radiation hard. Nevertheless, the pixel detector, which is exposed to the highest flux per unit area, will need to be replaced at least once during LHC lifetime. In order to limit the effect of radiation damage on sensor performances the tracker is operated at low temperature (-10°C).

The CMS tracker has to fulfill the following requirements:

- Isolated lepton reconstruction efficiency close to 100% within $|\eta| < 2$. Fig. 2.6(a) shows the reconstruction efficiency for single muon events.
- Transverse momentum resolution better than 4% within $|\eta| < 2$. Fig. 2.6 (b)



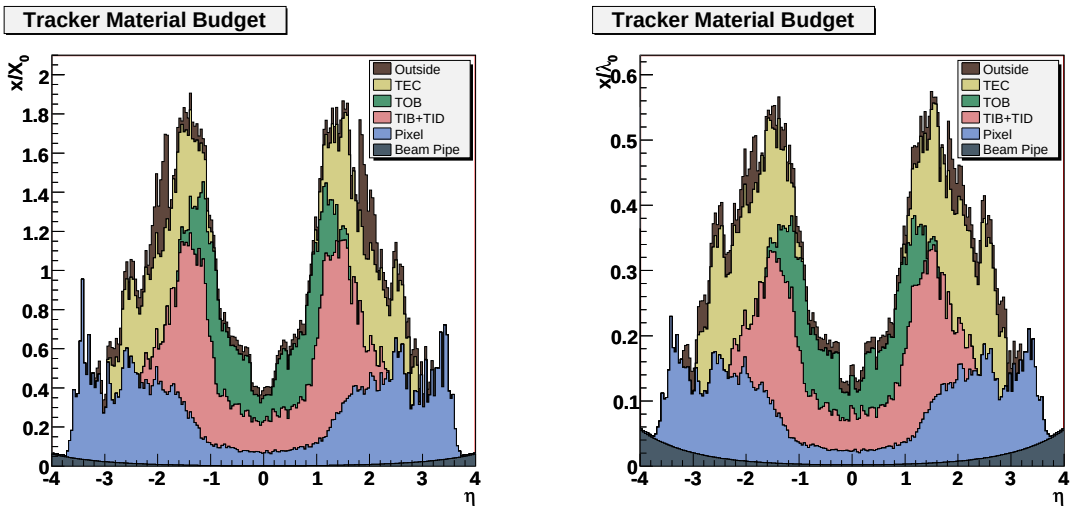
(a) (b)

Figure 2.6: (a) Reconstruction efficiency and (b) momentum resolution for single muons with p_T=1, 10, 100 GeV.

shows the momentum resolution for single muons.

- Tagging and reconstruction for b jets.

The material budget in the tracker has to be as limited as possible, as the electron energy loss due to bremsstrahlung and nuclear interactions of hadrons need to be kept as low as possible. This is needed not to spoil tracking performances and to keep the number of photons that get converted into an e^+e^- pair through interaction with the material as low as possible. The tracker depth in terms of radiation length X/X_0 [1] and in terms of interaction length λ/λ_0 [2] as obtained from the full simulation of the tracker is shown in Fig. 2.7 as a function of η. The material budget is higher in the region $1 < |\eta| < 2$ due to the presence of cables and services in this region.



(a) (b)

Figure 2.7: (a) Radiation length and (b) interaction length of the tracker as a function of η. Contributions from different components are put into evidence.

2.2.3 Electromagnetic calorimeter

The CMS electromagnetic calorimeter [37,38] is a highly segmented calorimeter, with excellent energy resolution, whose design was prompted by the possibility to observe Higgs decay into two photons. Since the intrinsic Higgs width in the region $m_H < 140\,\text{GeV}$ is of the order of 100 MeV, the width of the reconstructed $\gamma\gamma$

[1] X_0 is the distance over which a high energy electron reduces its energy to a fraction $1/e$ of the initial energy.

[2] λ_0 is the mean free path of a hadron before having an interaction when traversing a material.

invariant mass spectrum will be dominated by experimental resolution. Thus, an electromagnetic calorimeter with resolution of order 1% is needed.

ECAL is made of lead tungstate ($PbWO_4$) crystals. Lead tungstate is a radiation resistant scintillating material; radiation robustness was a key design requirement, because the absorbed dose per hour in high luminosity condition will range from 0.18 Gy/h at η=0, to 15 Gy/h at η=3.

Because of its high density (8.28 g/ cm^2), lead tungstate has a short radiation length X_0=8.9 mm. Such a short radiation length permitted a very compact design that made it possible to fit the calorimeter inside the magnetic coil, thus limiting the non-sensitive material traversed by electrons and photons. Another advantage of lead tungstate is the small Molière radius[3] (2.2 cm) and the fast scintillation decay time τ=10 ns that permits the collection of about 80% of the produced light in the 25 ns interval between two bunch crossings.

The main drawbacks of $PbWO_4$ are the low light yield (100 photons/ MeV) and the strong dependency of the response on the operating temperature, that makes it necessary to operate the crystals at stabilized temperature (18°C).

As shown in Fig. 2.8, ECAL is subdivided into a barrel region covering $|\eta| < 1.48$, and two endcap regions covering $1.48 < |\eta| < 3.0$.

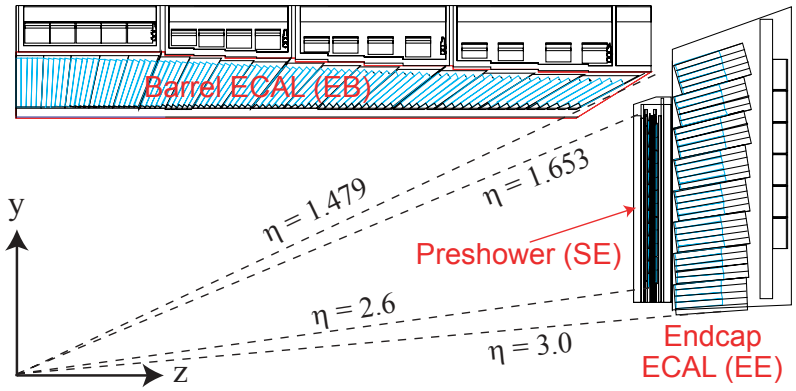


Figure 2.8: A schematic representation of a quadrant of ECAL.

Crystals in the barrel region are tapered shaped, with a 2.2 cm×2.2 cm front face and 23 cm length, and they are positioned at a radius of 1.24 m. The $\Delta\eta \times \Delta\phi$ granularity in the barrel is 0.0175×0.0175. The depth in radiation lengths in the barrel region is about 26 X_0. Crystal with a 3 cm×3 cm front face, 22 cm long (24.7 X_0) are used in the endcaps. The $\Delta\eta \times \Delta\phi$ granularity in the endcaps is 0.05×0.05.

The reduced depth in radiation length and the larger granularity in the endcaps with respect to the barrel are partially compensated with a preshower detector placed in front of the endcaps. Each preshower is made of two lead radiators and

[3]The Molière radius characterizes the transverse dimension of the electromagnetic shower evolving in a calorimeter.

two planes of silicon microstrip detectors. The preshower will improve π^0 rejection power in the forward region.

Barrel region crystals are read with avalanche photodiodes, APD. ECAL APDs are able to operate in the magnetic field and can cope with the low light yield of the crystals. Each crystal is equipped with two APDs that produce a total of about 4000 photoelectrons per GeV of deposited energy. In the endcaps, because of the higher irradiation level, APDs would suffer from high leakage current, thus the forward crystals are readout with vacuum photodiodes (VPD) [39].

In the range $25\,\text{GeV} < E < 500\,\text{GeV}$, the energy resolution σ_E is

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2, \tag{2.7}$$

where:

- a is $2.7\%\,\text{GeV}^{1/2}$ in the barrel and $5.7\%\,\text{GeV}^{1/2}$ in the endcaps. It is a stochastic term, and it is determined by the photoelectrons statistic;
- b is 155 MeV in the barrel and 200 MeV in the endcaps. It is determined by electronic noise and pileup;
- c is 0.55% both in the barrel and in the endcaps. It is related to the longitudinal shower evolution containment, the uniformity of the light collection in the crystals and the precision of the inter-calibration between crystals.

2.2.4 Hadron calorimeter

The CMS hadron calorimeter (HCAL) [40] is used together with ECAL to measure energy and direction of jets, the transverse energy E_T and the imbalance in transverse energy E_T^{miss}. It provides good segmentation, moderate energy resolution and angular coverage up to $|\eta| < 5$.

HCAL is made of four subdetectors (Fig. 2.9):

- the Barrel Hadronic Calorimeter (HB) is placed inside the magnetic coil and it covers the central pseudorapidity region, up to $|\eta| = 1.3$;
- the Endcap Hadronic Calorimeter (HE) is inside the magnetic coil as well and it is made of two endcaps extending the angular coverage up to $|\eta| = 3$;
- the Outer Hadronic Calorimeter (HO, or Tail Catcher) is placed in the barrel region, outside the magnetic coil and is needed to enhance the depth of the calorimeter in terms of λ_I;
- the Forward Hadronic Calorimeter (HF) consists of two units placed outside the magnetic coil, at ± 11 m from the interaction point along the beams direction. It extends the pseudorapidity coverage up to $|\eta| = 5$.

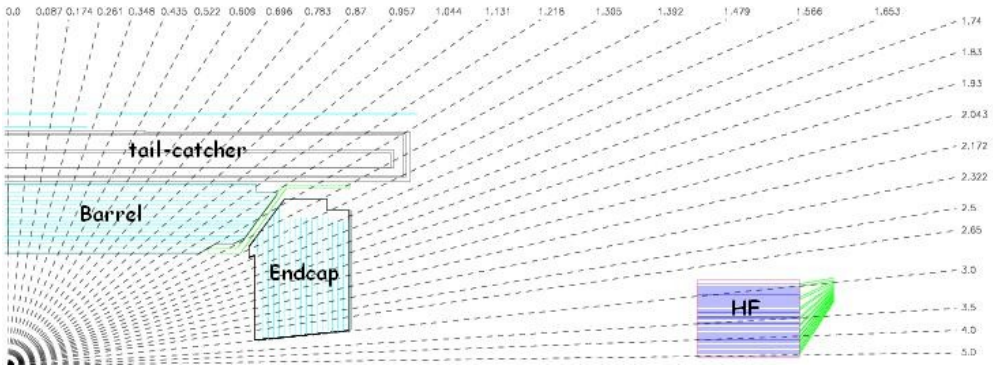


Figure 2.9: A schematic representation of a quadrant of HCAL.

HB and HE are made with layers of 5 cm thick brass absorber interleaved with 3.7 mm thick plastic scintillators. The signal is readout through wavelength-shift fibres. The $\Delta\eta \times \Delta\phi$ granularity is 0.087×0.087.

HB has an energy resolution for single pions of approximately $120\%/\sqrt{E}$. The minimum depth is about $5.8\lambda_I$. In order to increase the calorimeter depth in the barrel region a tail catcher (HO) has been added outside the magnetic coil. HO is made of two scintillator layers, with the same granularity as HB; the total depth in the central region is thus extended to about $11.8\lambda_I$, with an improvement in both linearity and energy resolution. HE has a minimum depth of $10\lambda_I$.

The two HFs are made of steel absorbers with embedded radiation hard quartz fibers. It provides fast Čerenkov light that is collected with photomultipliers. The granularity is $\Delta\eta \times \Delta\phi$=0.17×0.1745.

2.2.5 Muon system

The CMS muon system [41] is dedicated to the identification and measure of high p_T muons, in combination with the tracker. The system is placed outside the magnetic coil, embedded in the return yoke, to fully exploit the 1.8T return flux.

The system consists of three independent subsystems (Fig. 2.10):

- Drift Tubes (DT) are placed in the barrel region, where the occupancy is relatively low (< 10 Hz/m^2).
- Cathode Strip Chambers (CSC) are in the endcaps, where the occupancy is higher (> 100 Hz/m^2).
- Resistive Plate Chambers (RPC) are both in the barrel and in the endcaps.

The Drift Tube system is made of chambers consisting of twelve layers of drift tubes each, packed in three independent substructures called super-layers. In each super-layer two chambers have anode wires parallel to the beam axis, two have perpendicular wires. Thus, each super-layer can provide two measurements of the

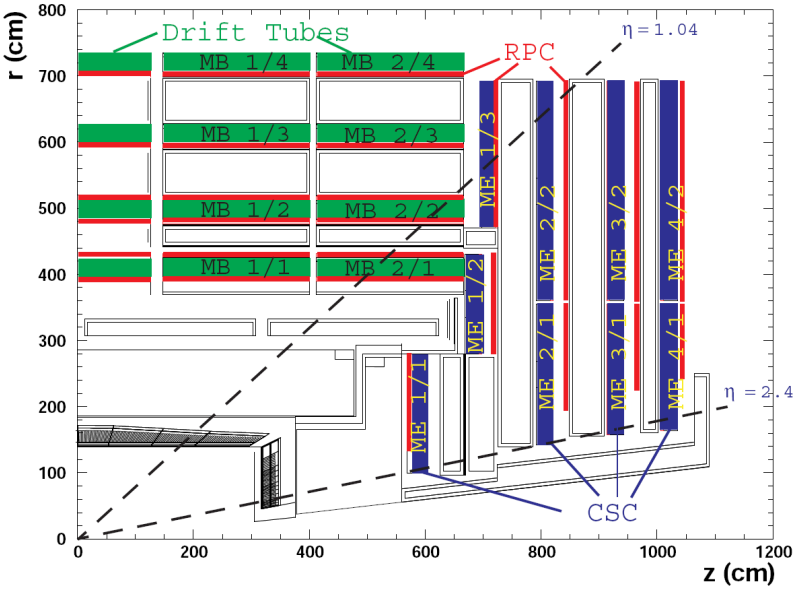


Figure 2.10: A schematic view of a quadrant of the CMS muon system.

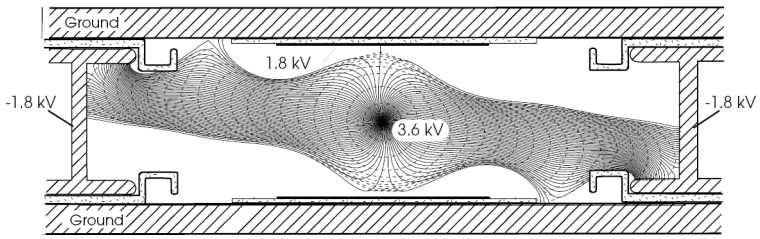


Figure 2.11: A schematic representation of a drift tube chamber. Drift lines in presence of magnetic field are also shown.

$r-\phi$ coordinate and two measures of the z coordinate of the track hit positions. Each chamber (Fig. 2.11) is made of two parallel aluminum plates with "I" shaped spacer cathodes, isolated from the aluminum plates with polycarbonate plastic. Chambers are filled with a gas mixture of Ar(85%) and CO_2(15%). The position resolution is about 100 μm in both $r\phi$ and rz.

Cathode Strip Chambers are multi-wire proportional chambers with segmented cathodes (Fig. 2.12). Each chamber can provide both hit position coordinates. Chambers are filled with a gas mixture of Ar(40%), CO_2(50%), CF_4(10%). The chamber spatial resolution is about 80-85 μm.

Resistive Plate Chambers are made of parallel bakelite planes, with a bulk resistivity of $10^{10} \div 10^{11}$ Ωcm. The gap between the plates if filled with a mixture of $C_2H_2F_4$ (94.5%) and i-C_4H_{10}. They operate in avalanche mode. Those chambers have limited spatial resolution, but they have excellent timing performances; they are mainly used for bunch crossing identification.

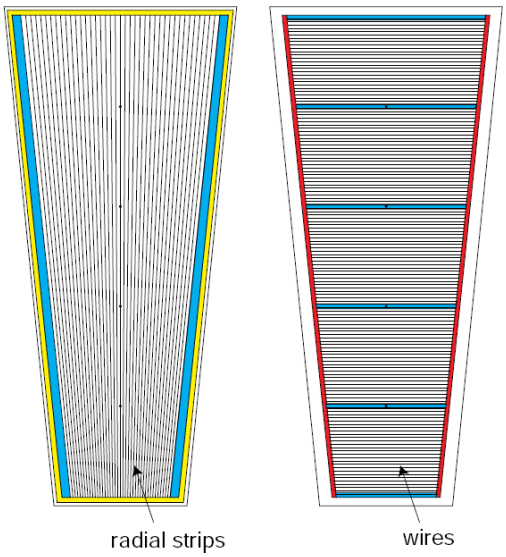


Figure 2.12: A schematic representation of CSC cathode panel (left) and anode panel (right).

2.3 Trigger system

LHC will produce interactions at 40 MHz frequency, but only a small fraction of these events can, and is worth to, be written on disk. On the one hand the speed at which data can be written to mass storage is limited; on the other hand the vast majority of events produced is not interesting, because it involves low transferred momentum interactions (minimum bias events). Thus, a trigger system is needed to save interesting events at the highest possible rate. The expected rate of events written to disk is foreseen to be 100 Hz.

CMS has chosen a two-level trigger system, consisting of a Level-1 Trigger (L1) [42] and a High Level Trigger (HLT) [43]. Level-1 Trigger runs on dedicated processors, and accesses coarse level granularity information from calorimetry and muon system. A Level-1 Trigger decision has to be taken for each bunch crossing within 3.2 μs. Level-1 Trigger task is to reduce the data flow from 40 MHz to 100 kHz.

The High Level Trigger is responsible for reducing the L1 output rate down to the target of 100 Hz. HLT code runs on a farm of commercial processors and can access the full granularity information of all the subdetectors.

We will now review the main characteristics of the CMS trigger system.

2.3.1 Level-1 Trigger

The Level-1 trigger is responsible for the identification of electrons, muons, photons, jets and missing transverse energy. It has to have a high and carefully understood efficiency. Its output rate and speed are limited by the readout electronics

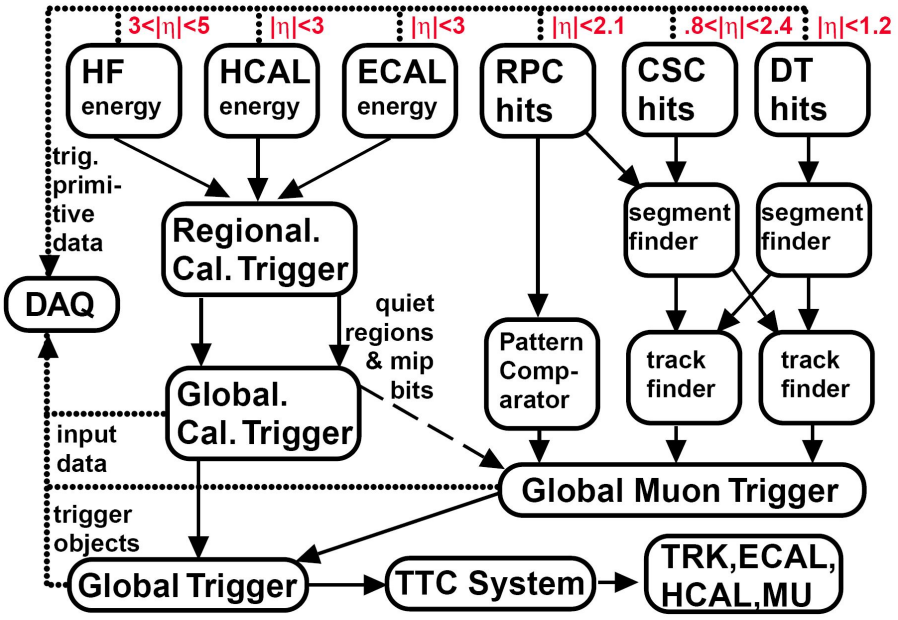


Figure 2.13: Level-1 trigger components.

and by the performances of the Data Acquisition (DAQ) system.

It consists of three main subsystems:

- L1 Calorimeter Trigger;
- L1 Muon Trigger;
- L1 Global Trigger.

The L1 Global Trigger is responsible for combining the output of L1 Calorimeter Trigger and L1 Muon Trigger and for making the decision. L1 Muon Trigger is actually a composed system itself: information from RPC, CSC and DT specific triggers are combined in the so called L1 Global Muon Trigger. The organization of CMS Level-1 Trigger is schematically summarized in Fig. 2.13.

L1 Calorimeter Trigger

The input for L1 Calorimeter Trigger is calorimeter towers, that is clusters of signals collected both from ECAL and HCAL. Towers are calculated by calorimeter high level readout circuits, called Trigger Primitive Generators.

The Regional Calorimeter Trigger finds out electron, photon, τ and jet candidates along with their transverse energy and sends them to the Global Calorimeter Trigger.

The Global Calorimeter Trigger sorts the candidates according to their transverse energy and sends the first four to the L1 Global Trigger.

L1 Muon Trigger

The RPC trigger electronics builds Track Segments, gives an estimate of the p_T and sends these segments to the Global Muon Trigger. It also provides the CSC logic unit with information to solve hit position ambiguities in case two or more muon tracks cross the same CSC chamber.

The CSC trigger builds Local Charged Tracks (LCT), that are track segments made out of the cathode strips only. A p_T value and a quality flag are assigned to the LCTs. The best three LCTs in each sector of nine CSC chambers are passed to the CSC Track Finder, that uses the full CSC information to build tracks, assign them a p_T and a quality flag and sends them to the Global Muon Trigger.

DTs are equipped with Track Identifier electronics, which is able to find groups of aligned hits in the four chambers of a super-layer. Those Track Segments are sent to the DT Track Correlator that tries to combine segments from two super-layers, measuring the ϕ coordinate. The best two segments are sent to the DT Track Finder that builds tracks and sends them to the Global Muon Trigger.

The Global Muon Trigger sorts the RPC, CSC and DT muon tracks and tries to combine them. The final set of muons is sorted according to the quality, and the best four tracks are passed to the L1 Global Trigger.

L1 Global Trigger

The L1 Global Trigger is responsible for collecting objects created from the Calorimeter and Muon Triggers and for making a decision whether to retain the event or not. If the event is accepted the decision in sent to the Timing Trigger and Control System, that commands the readout of the remaining subsystems.

In order to take the decision, the L1 Global Trigger sorts the ranked objects produced by calorimetry and muon system and checks if at least one of the thresholds in the Level-1 Trigger table is passed.

Since there are large uncertainties in the cross section of many processes, the Level-1 trigger thresholds for the initial low luminosity data taking have been designed for an output rate of 16 kHz, instead of the planned 50 kHz, that is the design limit for low luminosity. The L1 trigger table is reported in Table 2.2.

2.3.2 High Level Trigger

The High Level Trigger is designed to reduce the Level-1 output rate to the goal of 100 events/s that are definitely going to be written to mass storage. HLT code runs on commercial processors and performs reconstruction using the information from all subdetectors. Data read from subdetectors are assembled by a builder unit and then assigned to a switching network that dispatches events to the processor farm. The CMS switching network has a bandwidth of 1Tbit/s.

This simple design ensures maximum flexibility to the system, the only limitation being the total bandwidth and the number of processors. The system can be easily

Table 2.2: Level-1 Trigger table at low (high) luminosity. Thresholds correspond to values with 95% efficiency [43].

Trigger	**Threshold** (**GeV**/c^2 **or GeV**/c)	**Rate** **(kHz)**	**Cumulative Rate** (**kHz)**
Inclusive isolated e/γ	29 (34)	3.3 (6.5)	3.3 (6.5)
Di-e/di-γ	17 (19)	1.3 (3.3)	4.3 (9.4)
Inclusive isolated μ	14 (20)	2.7 (6.2)	7.0 (15.6)
Di-μ	3 (5)	0.9 (1.7)	7.9 (17.3)
Single τ-jet	86 (101)	2.2 (5.3)	10.1 (22.6)
Two τ-jet	59 (67)	1.0 (3.6)	10.9 (25.0)
1-jet, 3-jets, 4-jets	177,86,70 (250,110,95)	3.0 (3.0)	12.5 (26.7)
Jet $\otimes$ E_T^{miss}	86 $\otimes$ 46 (113 $\otimes$ 70)	2.3 (4.5)	14.3 (30.4)
e $\otimes$ jet	21 $\otimes$ 45 (25 $\otimes$ 52)	0.8 (1.3)	15.1 (31.7)
μ $\otimes$ jet	- (15 $\otimes$ 40)	- (0.8)	15.1 (32.5)
Minimum bias		0.9 (1.0)	16.0 (33.5)
Total			**16.0 (33.5)**

upgraded adding new processors or replacing the existing ones with faster ones as they become available. Since the algorithm implementation is fully software, improvements in the algorithms can be easily implemented and do not require any hardware intervention.

Event by event, the HLT code is run on a single processor, and the time available to make a decision is about 300 ms. The real time nature of this selection imposes several constraints on the resources an algorithm can use. The reliability of HLT algorithms is of capital importance, because events not selected by HLT are lost.

In order to efficiently process events the HLT code has to be able to reject not interesting events as soon as possible; computationally expensive algorithms must be run only on good candidates for interesting events. In order to meet this requirement the HLT code is organized in a virtually layered structure:

- Level 2: uses only muon and calorimetry information;
- Level 2.5: uses also the pixel information;
- Level 3: makes use of the full information from all the tracking detectors.

Each step reduces the number of events to be processed in the next step. The most computationally expensive tasks are executed in the Level 3; time consuming algorithms such as track reconstruction are only executed in the region of interest. Besides, since the ultimate precision is not required at HLT, track reconstruction is performed on a limited set of hits, and is stopped once the required resolution is achieved.

Table 2.3 summarizes the HLT requirements at low luminosity to match the cumulative rate of 100 Hz.

Table 2.4 shows the expected efficiency for the benchmark physics channels.

Table 2.3: High-Level Trigger thresholds at low luminosity for various channels [43]. The CPU time refer to a 1 GHz Intel Pentium III CPU.

Trigger	**HLT Threshold** (**GeV**/c^2 **or GeV**/c)	**HLT Rate** **(Hz)**	**CPU time** **(m s)**
1e ,2e	29, 17	33, 1	160
1μ, 2μ	19, 7	25, 4	710
1τ, 2τ	86, 59	3, 1	130
Jet $\otimes$ E_T^{miss}	180 $\otimes$ 123	5	50
e $\otimes$ jet	19 $\otimes$ 45	2	165
Inclusive b jets	237	5	300

2.4 CMS simulation and reconstruction software

2.4.1 Framework implementation

The CMS simulation and reconstruction software, CMSSW [44], is a C++ [45] framework that can be configured via Python [46] scripts.

CMS Event Data Model (EDM) is based on the concept of `Event`. An `Event` is a C++ class that contains the information about a physics event, both raw level data and reconstructed quantities. Reconstruction algorithms can access information from the `Event` and put reconstructed quantities in the event. `Events` can be read from and written to ROOT [47] files.

CMSSW can be run feeding the desired Python configuration script into the executable `cmsRun`. The configuration file contains the modules, i.e. the algorithms, that the user wants to run and it specifies the order in which they need to be run. The executable reads in the configuration file and, using a plugin manager, finds out in which libraries the modules to be run are defined and loads them.

Table 2.4: Performance of HLT selection at low luminosity after applying the cuts listed in Table 2.3.

Channel	**Efficiency**
H(115 GeV/c^2)$\rightarrow \gamma\gamma$	77%
H(160 GeV/c^2)$\rightarrow$WW*	99%
H$\rightarrow$ZZ$\rightarrow 4\mu$	99%
A/H(200 GeV/c^2)$\rightarrow 2\tau$	45%
SUSY (0.5 TeV/c^2 s-particles)	60%
W$\rightarrow$eν_e	67%
W$\rightarrow \mu\nu_\mu$	69%
$t\bar{t} \rightarrow \mu$+X	72%

Six types of modules can be implemented in CMSSW and dynamically loaded via the plugin mechanism:

- `Source`: these modules are used either to load events from a ROOT file or to produce events running an event generator.
- `EDProducer`: these modules read in the events provided by a `Source`, apply an algorithm to the data contained in the event and produce other data to be put in the `Event`. All the reconstruction algorithms are implemented as `EDProducers`.
- `EDFilter`: they work exactly as an `EDProducer`, but they are able to return a boolean value after the event has been processed. This boolean value can be used to decide whether to continue the reconstruction or to stop.
- `EDAnalyzer`: these modules are used to analyze and characterize events. They cannot put additional data in the `Event`, but can access the information stored in the `Event` and, i.e., produce analysis histograms.
- `EDLooper`: are used for particular tasks, such are track based alignment, in which there is a need to loop more than once on a set of events.
- `OutputModule`: these modules are used to write events to file after all the other modules were executed.

Often modules need auxiliary information that is not stored in the event; these information is stored in the `EventSetup` object.

2.4.2 Event Generation and Reconstruction

Event generation in CMSSW can be done with many event generator programs. Those programs can be run from within the framework, using dedicated interface libraries. The configuration of the event generators is performed feeding `cmsRun` with the appropriate configuration file containing the flags to be set in the event generator. The event generator is responsible for filling the HepMC [48] record with all the information about the currently generated event. The HepMC record is then captured by the CMSSW framework and stored in the `Event`.

After the event has been generated the simulation of detector follows. The first step in the simulation of instrumental effects is the smearing of the vertex position. The event primary vertex, that is placed by the event generator at the origin of CMS coordinate system, is smeared according to the expected pp impaction point position distribution per bunch crossings. The next step is the simulation of the interaction of particles with the detector. The description of these interactions is achieved using GEANT4 [49]. Once energy deposits and multiple scattering effects in the CMS subdetectors are simulated, the simulation of the signals produced by the subdetectors follows. This step is knows as "digitization".

The chain described so far is often referred to as "full simulation" chain. The most time consuming step of the full simulation is actually the simulation of detector effects using GEANT; the time needed to full-simulate an event with GEANT can amount to several minutes.

For this reason a "fast simulation" of the detector effects has been set up. In the fast simulation the GEANT step and the digitization step are skipped and detector level quantities, such as the hit positions in the tracker and the energy deposits in the calorimeters, are described using parametrized functions that aim at reproducing the full simulation result.

Starting from the simulated signals in each subdetector (or from the low level reconstructed quantities produced by the fast simulation), the reconstruction of the event follows. With this approach exactly the same algorithms that will be used on real data can be run on simulated samples.

Chapter 3

Proton-proton phenomenology: Monte Carlo method

The structure of events produced at high energy colliders is extremely complex, and numeric simulations are necessary to effectively simulate realistic events. Monte Carlo Event generators are complex computer programs that subdivide the problem of producing realistic events into a sequence of tasks that can be handled separately with the help of both analytic and numeric computation.

Different event generators exist that implement computations at different levels of precision and with different techniques. Typically, the highest precision calculations, that take into account several orders in perturbation theory, are only available for a limited number of processes, thus making it hard to derive predictions on inclusive quantities; on the other hand these quantities can often be described with reasonable precision with programs that implement lower order calculations.

In this chapter we shall review the main aspects of the computations that lead to the generation of realistic events; we will describe how the different steps are implemented in some of the most popular generators, and we will highlight the differences. To understand the differences among the event generators "on the market" is a fundamental prerequisite to interpret experimental measurements. In particular we will describe in detail the computation techniques used in `PYTHIA` [50], `AlpGen` [51] and `SHERPA` [52] event generators, that were used in this work to produce $Z/\gamma^* + jets$ events.

3.1 Event Generator components

A schematic representation of the different components (and calculation steps) that are implemented in event generators is shown in Fig. 3.1.

The production of hadron-hadron collision events is the result of the following chain of calculations:

- The first step is the calculation of cross sections for the selected processes.

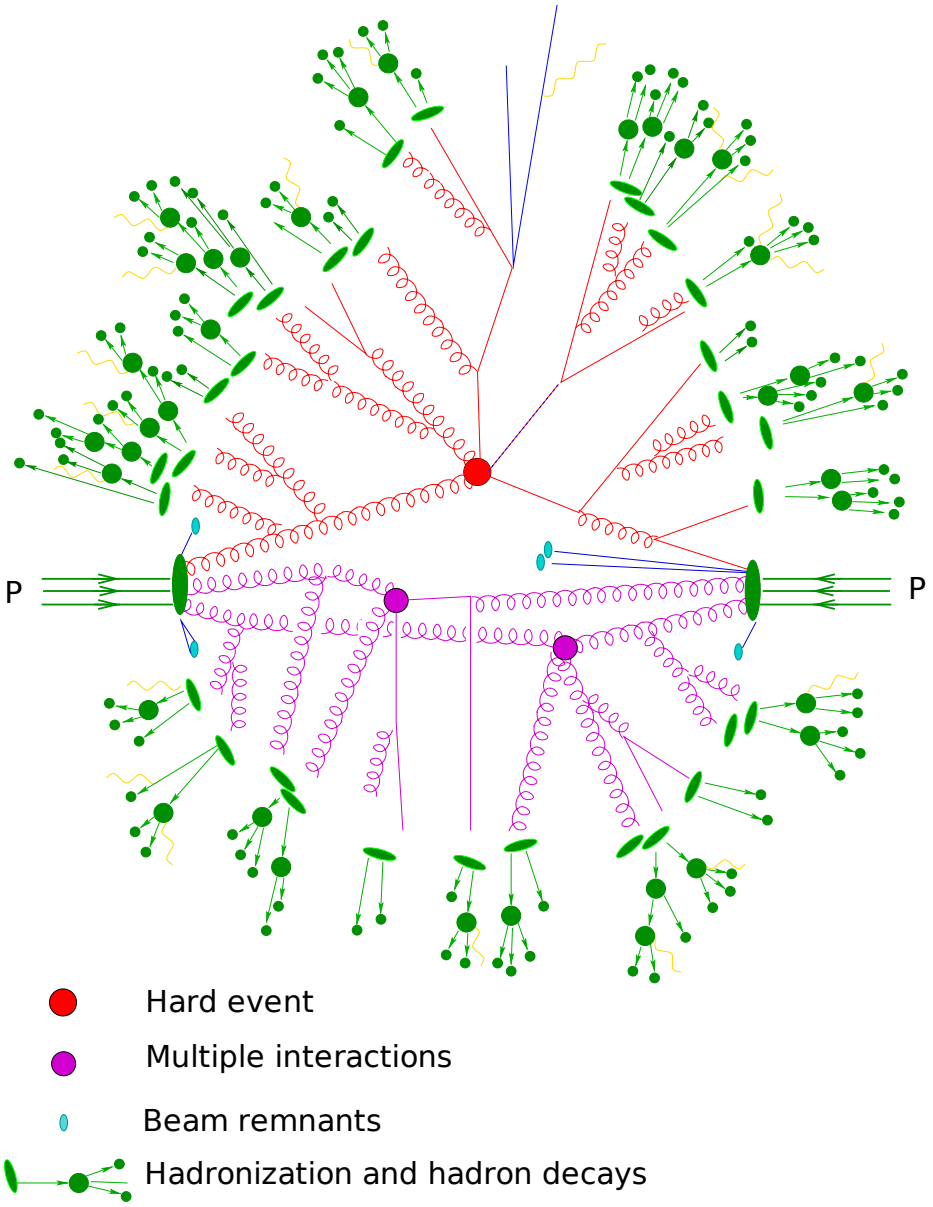


Figure 3.1: A schematic representation of the generation of an event in a typical event generator [52]. Partons from the two incoming hadrons participate in the hard scattering and in softer multiple interactions. Hadron remnants are treated. Quarks and gluons are turned into hadrons by hadronization and then hadrons decay.

Cross sections are calculated for a pair of incoming partons (quarks and gluons) extracted from the colliding hadrons.

- The event production starts with two colliding hadrons with given momenta. One parton out of each hadron is selected to enter the scattering process we are interested in. This step is often referred to as hard scattering generation. Final state partons and leptons are produced according to the calculated differential cross sections.

- Resonances produced in the hard event are decayed.

- When two partons take part in the hard event, accelerated colour charges are present, thus bremsstrahlung can occur. This effect is called Initial State Radiation (ISR) and is simulated with the so called Initial State Parton Showers. To simulate ISR knowledge of the parton density function is needed.

- Also the final state partons can produce further radiation, called Final State Radiation (FSR). Such radiation is simulated by the Final State Parton Showers.

- In addition to the partons taking part in the hard interaction, several other parton pairs can interact during a hadron-hadron collision, giving rise to interactions with smaller transferred momentum. These Multiple Parton Interactions (MPI) contribute to the so called underlying structure of the event. Such interactions need to be simulated too if we want to produce realistic events, and ISR and FSR need to be simulated for these collisions too.

- Leftovers of the interacting hadrons needs to be simulated to balance the colour charge and for momentum conservation. The beam remnant handling is thus another step in the event generation.

- The calculations described so far are carried out in the perturbative regime, but, as the produced partons move apart from each other, the coupling constant gets stronger and stronger and confinement effects take place. When the coupling constant is strong enough quark-antiquark pairs are produced from the vacuum and the partons turn into hadrons. This generation step is referred to as hadronization.

- Finally, the event generator takes care of decaying τ leptons and B-hadrons; in general particles with very short lifetime are decayed by the generator itself. Those that live enough to reach the detector are left undecayed.

Many generators exist nowadays that can make calculations up to several partons in the final state. Among the general purpose ones there are many tree-level generators: `PYTHIA` [50], `HERWIG` [53], `AlpGen` [51], `SHERPA` [52], `MADGRAPH/MADEVENT` [54, 55] are able to make matrix element calculations for a number of processes, and

to match the matrix element outcome with parton showers. A few generators which can perform the full Next-to-Leading Order (NLO) calculation with all virtual corrections included are also available for a limited number of processes; an example is the program called MC@NLO [56].

3.2 Matrix element calculations

The first step in the generation of an event is the calculation of the hard processes cross sections. General purpose event generators can perform such calculations for a vast variety of processes. Nevertheless it is often useful to interface such generators with dedicated hard process libraries in order to produce particular events, such as Super-symmetry (SUSY) processes for example.

The state-of-the-art in the field of matrix element (ME) calculation is NLO, with all the virtual loop corrections included. Loop calculations are complex and they are available for a limited number of processes; for this reason tree-level matrix element calculations still play an important role in the simulation of events produced at hadron colliders; tree-level cross section calculations can be performed up to several (eight or even more) partons in the final state.

The main problems with tree-level matrix elements are the soft and collinear divergences (Sec. 1.4.4). Since at tree-level the loop corrections that would cancel these divergences are omitted, the phase space has to be carefully tailored to avoid the problematic regions. This means that the matrix element cross section calculations are performed away from soft and collinear divergences.

In order to produce realistic events, phase-space regions omitted from the matrix element calculations need to be recovered, with care to avoid divergences. This is done in a quite effective way by using parton shower calculations (Sec. 3.3).

In the following we will briefly describe various techniques to derive tree-level matrix elements as they are implemented in three different generators: PYTHIA, AlpGen and SHERPA.

3.2.1 Hard-coded differential cross sections

In many event generators, like PYTHIA, differential cross sections are hard-coded. In PYTHIA many $2 \to 2$ processes and resonance production $2 \to 1$ processes are implemented.

With this approach, however, it is difficult to go much further than $2 \to 2$ processes, even if one considers tree-level only. As the number of final state partons grows the number of processes grows too and the hard-coding of many extremely complicated cross section expressions becomes difficult.

If tree-level matrix elements with a large number of final state partons are needed, automated calculation procedures need to be used. Two examples are given in Secs. 3.2.2 and 3.2.3.

3.2.2 ALPHA algorithm

The ALPHA algorithm [57] used in `AlpGen` is a numeric method to calculate tree-level scattering amplitudes without Feynman diagrams. Scattering amplitudes are connected with the Z-functional of the Lagrangian, that is the Legendre transform of the Lagrangian as long as tree-level is concerned. The algorithm consists in a numeric prescription to compute the Lagrangian Legendre transform, and thus the scattering amplitudes in a straightforward way.

This algorithm is particularly fast and it allows for the computation of the matrix element up to several partons in the final state, without the need to implement the Feynman diagrams in the code. In this way both speed and generality of the algorithm for a great variety of processes are addressed.

3.2.3 Helicity amplitudes

The helicity amplitudes method [58, 59] used in `SHERPA` makes use of the Feynman diagrams, but amplitudes are calculated in a numeric fashion. While in an analytical approach amplitudes are calculated for each Feynman diagram, the basic idea of the helicity amplitudes method is that, given a set of particles momenta and helicities, for a given Feynman diagram, the amplitude is nothing else but a complex number. Such numbers can be calculated sampling the phase space and then they are squared and summed.

This approach relies on the decomposition of the Feynman amplitudes into building blocks consisting of scalar products of four-momenta and spinor products of the form $u(p_i, \lambda_i)\bar{u}(p_j, \lambda_j)$, where p_i and λ_i are momentum and helicity respectively. Such products can be easily handled numerically.

In order to integrate the amplitudes sophisticated integration techniques are needed to efficiently sample the full phase-space.

3.3 Parton showers

When treating $2 \to n$ processes, tree-level matrix elements suffer from divergences in the soft and collinear regions. The splittings that suffer from these divergences are $q \to qg$, $\bar{q} \to \bar{q}g$, $g \to gg$: the two first processes have a QED counterpart, while the third comes from the non-abelian nature of QCD. The splitting $g \to q\bar{q}$ does not suffer from the soft divergence.

The tree-level divergences would be removed including also virtual corrections; such calculations, however, are extremely complex and are available only for a limited set of processes. Parton showers offer an alternative way both to handle the complexity of several successive branchings and to remove soft and collinear divergences in a physical way.

Far before the hard scattering process and far after that, partons are on the mass shell; however, by the uncertainty principle, as the colliding partons approach

the hard scattering (i.e. closer and closer in time to the hard scattering) they can go more and more off-shell; thus, as the partons approach the hard scattering they can emit harder and harder gluons. The virtuality of the emitting partons in such conditions is space-like[1]. This behavior is modeled in the Initial State parton showers.

As partons arising from the hard event move apart, the allowed virtuality decreases, so softer and softer gluons can be emitted. The emitting parton virtuality in these circumstances is time-like[1]. This effect is modeled by Final State parton showers.

A description of Initial and Final State Showers follows in Secs. 3.3.1 and 3.3.2, starting from the Final State Shower because they are simpler.

3.3.1 Final State showers

In the vast majority of event generators both final and initial state parton showers are modeled as a series of $a \to bc$ splittings. The evolution of the shower is described with two parameters: the energy fraction z carried by one of the two emerging partons, $z = E_b/E_a$, and an ordering variable t. Different choices are possible for the ordering variable. A common choice, used both in `PYTHIA` and in the shower module of `SHERPA`, is the virtuality Q^2 of the parton that is going to split, $Q^2 = p_a^2$. Saying that Q^2 is the ordering variable of the shower means that the final state radiation emission is strictly ordered with decreasing Q^2. Other choices are possible however. If the transverse momentum is used as an ordering variable,

[1]To see why an initial state branching involves space-like virtuality and a final state branching involves time-like virtuality let's consider a generic $a \to bc$ splitting. Let's suppose a is moving along the z axis of some reference frame. It's convenient to introduce light-cone momenta $p_\pm$ defined as:

$$p_\pm = E \pm p_z. \tag{3.1}$$

For a particle with virtuality m:

$$p_+ p_- = m^2 + p_x^2 + p_y^2 = m^2 + p_\perp^2. \tag{3.2}$$

If we define f as $p_+^b = f p_+^a$, then, conservation of p_+ leads to $p_+^c = (1-f)p_+^a$; also, $p_\perp^b = p_\perp^c = p_\perp$. Recalling Eq. (3.2), conservation of p_- is then written as:

$$\frac{m_a^2}{p_+^a} = \frac{m_b^2 + p_\perp^2}{f p_+^a} + \frac{m_c^2 + p_\perp^2}{(1-f)p_+^a}, \tag{3.3}$$

thus:

$$m_a^2 = \frac{m_b^2}{f} + \frac{m_c^2}{1-f} + \frac{p_\perp^2}{f(1-f)}. \tag{3.4}$$

In an initial state branching, if we assume that a is on shell and that c does not branch any further, which means that it is on shell too, from Eq. (3.4) we see that $m_b^2 = -p_\perp^2/(1-f) < 0$, which means that b is space like. On the other hand, in a final state branching, assuming that neither b nor c is going to branch any further, i.e. they are both on shell, we have $m_a^2 = p_\perp^2/f(1-f) > 0$, which means that a is time-like.

the Parton Shower is said to be p_T-ordered. Another possible choice is the energy weighted opening angle of each emission, $E^2\theta^2$; angular ordering is used in HERWIG.

Let's now consider in detail a virtuality ordered final state shower. Let us consider a $q\bar{q}g$ final state (Fig. 3.2). If we let $x_i = 2E_i/\sqrt{s}$ (i=q, $\bar{q}$, g) be the energy

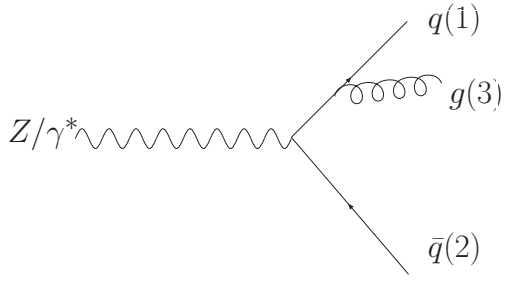


Figure 3.2: A $q\bar{q}g$ final state.

fractions of the final state particles we see that the following holds:

$$1 - x_2 = \frac{m_{13}^2}{s}, \tag{3.5}$$

where m_{13} is the invariant mass of the qg pair, and thus the virtuality Q^2 of the intermediate fermion line. The limit $x_2 \to 1$ corresponds to the collinear emission limit ($Q^2 = m_{13}^2 \to 0$) as long as massless final state particles are concerned (a massless particle can split and remain massless only if the emission is soft or collinear). In this limit x_1 corresponds to the energy fraction $z = E_1/E_{13}$, thus $x_1 \approx z$, $x_3 \approx 1 - z$.

The splitting probability $\mathrm{d}\mathcal{P}_{a\to bc}$ in the limit of collinear emission is conveniently expressed in terms of z and $t = \ln(Q^2/\Lambda^2)$ as:

$$\mathrm{d}\mathcal{P}_{a\to bc}(t,z) = \sum_{bc} \frac{\alpha_{abc}}{2\pi} P_{a\to bc}(z)\mathrm{d}t\mathrm{d}z, \tag{3.6}$$

where $\mathrm{d}t = \frac{\mathrm{d}Q^2}{Q^2}$, α_{abc} is the coupling constant that governs the splitting (either α_S or α_{QED}), $P_{a\to bc}(z)$ is called splitting kernel.

The splitting kernels are derived from the Feynman amplitude in the limit of collinear splitting. They are universal and have the following form:

$$\begin{aligned} P_{q\to qg} &= \frac{4}{3}\frac{1+z^2}{1-z}, \\ P_{g\to gg} &= 3\frac{(1-z(1-z))^2}{z(1-z)}, \\ P_{g\to q\bar{q}} &= \frac{n_f}{2}(z^2 + (1-z)^2). \end{aligned} \tag{3.7}$$

where n_f is the number of quark flavours.

Eq. (3.6) suffers from all the divergences of the original matrix element, because it is just an approximation of the exact matrix element in the collinear emission limit. In particular, the collinear divergence resides in the $1/Q^2$ dependence, and the soft divergence arises from the splitting kernels of Eq. (3.8) for $z = 1$.

The parton shower machinery handles these divergences manually imposing the conservation of total probability. First of all, the branching probability between t and $t + \mathrm{d}t$ is obtained integrating Eq. (3.6) over the z allowed range, $z_{min}(t) < z < z_{max}(t)$.

$$\mathrm{d}\mathcal{P}_a(t) = \Big(\sum_{bc} \int_{z_{min}(t)}^{z_{max}(t)} \frac{\alpha_{abc}}{2\pi} P_{a\to bc}(z)\mathrm{d}z \Big)\mathrm{d}t. \tag{3.8}$$

This branching probability, however, does not conserve the total branching probability. Then, the probability of branching between t and $t + \mathrm{d}t$ is obtained as the probability given by Eq. (3.8) times the probability that no branching occurred between the starting shower scale t_0 and t. Thus the expression of the branching probability that conserves total probability is rather

$$\mathrm{d}\mathcal{P}_a^{FSR}(t) = \mathrm{d}\mathcal{P}_a(t) \cdot \exp\Big(- \sum_{b,c} \int_{t_0}^{t} \mathrm{d}t' \int_{z_{min}(t')}^{z_{max}(t')} \frac{\alpha_{abc}}{2\pi} P_{a\to bc}(z')\mathrm{d}z' \Big). \tag{3.9}$$

The exponential term in Eq. (3.9) is called *Sudakov form factor*; it represents the non-emission probability between scale t_0 and t.

As said, the Sudakov form factor can be regarded as a term that conserves the total branching probability. However the Sudakov has also an interpretation in terms of Feynman diagrams; while the branching probability given in Eq. (3.6) can be regarded as an approximation of the exact matrix element for real emission, the Sudakov factor is an approximation of the complete virtual loop corrections.

With this machinery a consistent cascade of successive branchings can be evolved. This is achieved in the following way; for each branching the scale t of the branching, the type of the branching and the value of z have to be assigned. The scale t is chosen with the probability given by Eq. (3.9); once t is fixed the type of the branching is assigned; if more than one branching is possible, the type is assigned with a probability proportional to the integrated splitting kernels. Finally z is chosen according to the unintegrated splitting function selected at the previous step. The parton cascade is evolved down to a certain virtuality, of the order of 1 GeV^2. After that, non perturbative effects take place and hadronization is applied.

It should be noticed that the parton shower machinery relies on a collinear approximation of the matrix element, thus it should perform well in the description of the evolution of jets, but one cannot expect it to give a precise answer for the description of well separated parton configurations. We will come back on this point again in Sec. 3.4

3.3.2 Initial State showers

The evolution of initial state radiation is more complicated than that of final state radiation. Quarks and gluons are continuously produced and reabsorbed in the colliding hadrons. This means that when the hard scattering occurs the initial state radiation is already there, and the fact that a parton is removed from the initial state due to the hard process prevents recombination processes.

Naively, one might think of simulating the initial state radiation starting from the on shell partons well before the interaction, evolving them to higher and higher scales up to the scale of the hard event and, if the evolution has been such that the hard event can occur, generating the hard event. While this approach is rather realistic, the efficiency can be quite low, especially if the hard event has a tiny phase space available.

A different approach is used in most event generators: the hard event is produced first, then one tries to reconstruct what might have happened before to the initial state partons. This approach is often referred to as *backward evolution*.

Following the same approach outlined for the final state radiation we consider splittings of the type $a \to bc$, we write down the splitting probability and then construct a Sudakov factor to conserve total probability.

The parton density function for a parton of type b evolves according to the DGLAP equation [24]:

$$\frac{\mathrm{d}f_b(x,t)}{\mathrm{d}t} = \sum_{a,c} \int \frac{\mathrm{d}x'}{x'} f_a(x',t) \frac{\alpha_{abc}}{2\pi} P_{a\to bc}\Big(\frac{x}{x'}\Big), \tag{3.10}$$

where $f_{a(b)}(x,t)$ is the parton distribution function for a parton of type a (b) carrying a momentum fraction x of the hadron momentum at scale $t = \ln(Q^2/\Lambda^2)$; $P_{a\to bc}$ is the splitting kernel given by Eq. (3.8).

In the backward evolution, the probability for parton b to "recombine" into parton a, whatever a is allowed to be, going from scale t to $t - \mathrm{d}t$ is [60]:

$$\mathrm{d}\mathcal{P}_b(t) = \frac{\mathrm{d}f_b(x,t)}{f_b(x,t)} = |\mathrm{d}t| \sum_{a,c} \int \frac{\mathrm{d}x'}{x'} \frac{f_a(x',t)}{f_b(x,t)} \frac{\alpha_{abc}}{2\pi} P_{a\to bc}\Big(\frac{x}{x'}\Big). \tag{3.11}$$

Recalling the final state radiation treatment, we compute the non-branching probability between scale t_{max} and $t < t_{max}$:

$$S_b(x,t;t_{max}) = \exp\Big(- \int_t^{t_{max}} \mathrm{d}t' \sum_{a,c} \int \frac{\mathrm{d}x'}{x'} \frac{f_a(x',t')}{f_b(x,t')} \frac{\alpha_{abc}}{2\pi} P_{a\to bc}\Big(\frac{x}{x'}\Big)\Big). \tag{3.12}$$

The probability for parton b to recombine into a during the backward evolution, between t and $t - \mathrm{d}t$ is thus:

$$\begin{aligned} \mathrm{d}\mathcal{P}_b^{ISR}(t) &= -\frac{\mathrm{d}S_b(x,t;t_{max})}{\mathrm{d}t}\mathrm{d}t \\ &= \Big\{ \sum_{a,c} \int \frac{\mathrm{d}x'}{x'} \frac{f_a(x',t)}{f_b(x,t)} \frac{\alpha_{abc}}{2\pi} P_{a\to bc}\Big(\frac{x}{x'}\Big)\Big)\Big\} \cdot S_b(x,t;t_{max}) \cdot \mathrm{d}t. \end{aligned} \tag{3.13}$$

The ISR Sudakov factor $S_b(x, t; t_{max})$ is quite different from the one of the FSR, as it contains the PDFs.

In order to produce an initial state parton cascade three quantities have to be determined for each branching: the scale t, the flavour a, the fraction $z = x/x'$; they are determined with the Monte Carlo method in this way:

1. t value is selected with the probability given by Eq. (3.13);
2. the flavour a is selected according to the ratio of the integrated splitting kernels over z for different allowed choices of a;
3. finally z is chosen with probability given by the unintegrated splitting kernels.

The backward evolution technique leaves some ambiguity about the interpretation of z. When the two colliding partons are on-shell, each carrying a fraction x_i of the hadron momentum, the center of mass energy for the two partons system is

$$\hat{s} = x_1 x_2 s. \tag{3.14}$$

A common approximation is to require that Eq. (3.14) is true for each splitting. This means that for each splitting in the backward evolution $\hat{s}$ takes a factor $1/z$. Besides it also means that we have to fit the splittings of the two partons contributing to the hard interaction into a unique monotonic series. Further details can be found in [61]. This complication is not present in FSR.

3.3.3 Parton Showers and resummation

When calculating a QCD observable perturbatively, the expansion in terms of α_S contains terms with form $\alpha_S^n L^k$ ($k \leq 2n$), where $L = \ln q_{cut}/s$, q_{cut} being the cutoff for resolved emission. When small values of q_{cut} are considered, even if the perturbative regime is still valid (α_S is small), the logarithms that appear in the perturbative expansion may be large and spoil the convergence of the series. In other words, the order n in the perturbative expansion is meaningful only if successive terms in the series are small: this may not be the case if L is large.

In order for the perturbative expansion to recover predictive power, large logarithms need to be treated. The treatment of large logarithms is called "resummation"; resummation is performed organizing the terms in the perturbative series according to the degree of divergence ($\alpha_S^n L^{2n}$ are the leading-log terms, LL; $\alpha_S^n L^{2n-1}$ are the next-to-leading log terms, NLL) and then resumming them to all orders in α_S. Resummed calculation are available for many processes at the NLL accuracy, which means that terms up to NLL in the perturbative expansion are resummed to all orders.

As discussed in [62, 63] the Sudakov factor in Parton Showers effectively reproduces the effect of resummation. Actually the coherent, angular ordered Parton Shower is proven to resum large logarithms up to the NLL accuracy. In `HERWIG` coherence effects are naturally included, while in `PYTHIA` and in the shower module of `SHERPA` the coherence is manually imposed.

3.4 Merging Matrix Elements and Parton Shower

Matrix element and parton shower calculations have different virtues and different applicability limits. We can summarize some of the main facts about the ME calculations as follows:

- as long as tree-level is concerned, these calculations can be performed up to several (order of six) partons in the final state;
- ME are good at describing well separated parton configurations;
- ME calculations are exact to a given order in perturbation theory.

However:

- ME cross sections have divergences in the soft and collinear regions, thus they can not describe the internal structure of a jet;
- since hadrons is what we observe in experiments, fragmentation models need to be applied to the partons. To use bare ME partons would imply the need to tune these models for each center of mass energy; this limits the applicability of bare ME calculations.

On the other hand, Parton Showers:

- are universal; given the basic hard process, the parton shower technique will produce realistic parton configurations;
- are derived in the collinear limit, and handle divergences with the use of Sudakov form factors. This makes them particularly suited to describe the evolution of jets;
- can be used to evolve partons down to a common scale; this removes the need of tuning fragmentation models at different scales.

However, since they are derived in the collinear approximation, they may fail in efficiently filling the phase space for well separated parton configurations.

From the above description it is clear that a combined use of ME and PS would make it possible to take advantages of the qualities of the two approaches in the phase space regions where each performs better.

Several prescriptions exist to perform ME-PS matching avoiding double-counting or holes in the phase space. The aim of a ME-PS matching prescription can be formulated as follows: describe a final state with n well separated partons with the corresponding n-partons tree-level matrix elements, but also including the large logs resummation that characterizes the parton shower. So, matching prescriptions foresee a criterion to fill the phase space by making a combined use of ME and PS; care must be taken to avoid double counting, i.e. to avoid that a configuration with

n partons emerging from the ME is produced also by an $(n-1)$-partons ME plus an additional hard emission coming from the PS.

The aim of a ME-PS merging prescription is described in Fig. 3.3. α_S orders and large logarithm powers for a QCD perturbative calculation are shown on the x and y axis respectively. As said in Sec. 3.3.3, PS calculations are able to re-sum Leading and Next-to-Leading Large Logarithms. This means that parton showers "live" on the diagonal green lines in Fig. 3.3. To produce a four jet event with a matched ME-PS approach means that we want to take into account all the blobs labeled by 4 plus all the green blobs to the right of the four jets vertical line; the latter are produced by the parton shower. The matching prescription has to make sure that the green blobs labeled 4 are taken into account only once.

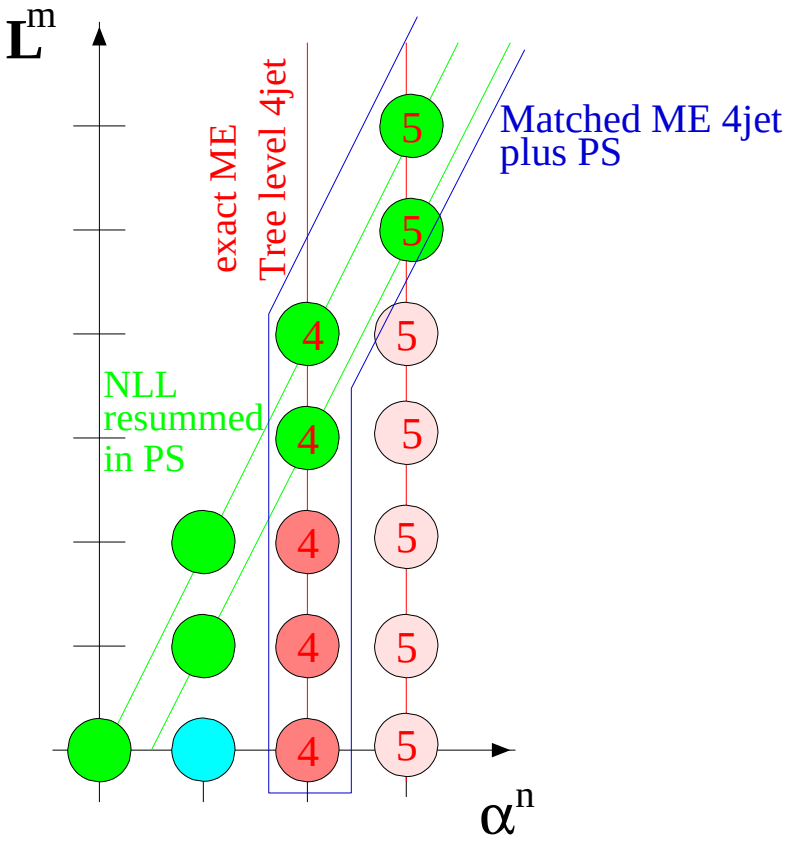


Figure 3.3: A schematic description of α_S orders and large logarithm powers that enter a multi-parton calculation. The terms within the blue contour are those included in a matched ME+PS treatment of a 4 jet event.

In the following we will describe in detail three different matching prescriptions, implemented in the three programs used to produce $Z + jets$ events in this work: `PYTHIA`, `AlpGen` and `SHERPA`. `PYTHIA` implements a corrected PS approach, in which the first emission from the shower is ME-corrected [64–66]; `AlpGen` implements the so called MLM prescription [67], while `SHERPA` uses the CKKW prescription [68,69].

3.4.1 Parton shower reweighting

The basic idea in this approach is to start from the lowest order process and then reweight the parton shower emission as if it were produced by a ME. For simplicity, let's consider for the moment the $e^+e^- \to jets$ process, where only the Final State radiation is present; the lowest order process is $e^+e^- \to q\bar{q}$, with $e^+e^- \to q\bar{q}g$ being the first order real correction. We know that the cross section for the first order correction is

$$\frac{1}{\sigma}\frac{\mathrm{d}^2\sigma}{\mathrm{d}x_1\mathrm{d}x_2} = \frac{2}{3}\frac{\alpha_S}{\pi}\frac{x_1^2+x_2^2}{(1-x_1)(1-x_2)}, \tag{3.15}$$

where $x_i = 2E_i/\sqrt{s}$ are the energy fractions of the quark and the antiquark. In a virtuality ordered PS, like `PYTHIA`, the first emission from the shower is [64]

$$\frac{1}{\sigma}\frac{\mathrm{d}^2\sigma}{\mathrm{d}x_1\mathrm{d}x_2} = \frac{2}{3}\frac{\alpha_S}{\pi}\frac{A_{PS}(x_1,x_2)}{(1-x_1)(1-x_2)} \tag{3.16}$$

with

$$\begin{aligned} A_{PS}(x_1,x_2) &= 1 + \frac{1-x_1}{(1-x_1)+(1-x_2)}\cdot\frac{x_1^2}{(2-x_2)^2} \\ &\quad + \frac{1-x_2}{(1-x_1)+(1-x_2)}\cdot\frac{x_2^2}{(2-x_1)^2}. \end{aligned} \tag{3.17}$$

Eq. (3.15) can be written in the same form as Eq. (3.16) if we substitute $A_{PS}(x_1,x_2)$ with

$$A_{ME}(x_1,x_2) = x_1^2 + x_2^2. \tag{3.18}$$

In the collinear limit, $x_1 \to 1$ or $x_2 \to 1$, $A_{PS}(x_1,x_2) \to A_{ME}(x_1,x_2)$. In the remaining part of the phase space $A_{ME}(x_1,x_2) \leq A_{PS}(x_1,x_2)$, as shown in Fig. 3.4. We can use this feature of the virtuality ordered PS to re-weight the PS emission through a veto algorithm. To do this one just has to produce the PS emission and veto the emission according to A_{ME}/A_{PS}.

In `PYTHIA` the first emission from the shower is corrected with this technique. This approach does not change the cross section, which remains the lowest order one, but it improves the population of the phase space in a way that is corrected to the first order matrix element.

A problem may arise in the above prescription if the first emission is not the hardest, as pointed out in [65]. In virtuality ordered PS the first emission is usually the hardest, so the problem is not serious, but in angular ordered showers, like `HERWIG`, it is likely to have several soft, large angle gluons before the first hard emission, that is the one that deserves the correction, so a modified treatment is needed, as described in [65].

The technique described above can be used also in hadron collisions, where ISR is present, as described in [66] for the case of inclusive W production. ME corrections for the Initial State Radiation will be further discussed in Sec. 4.1.

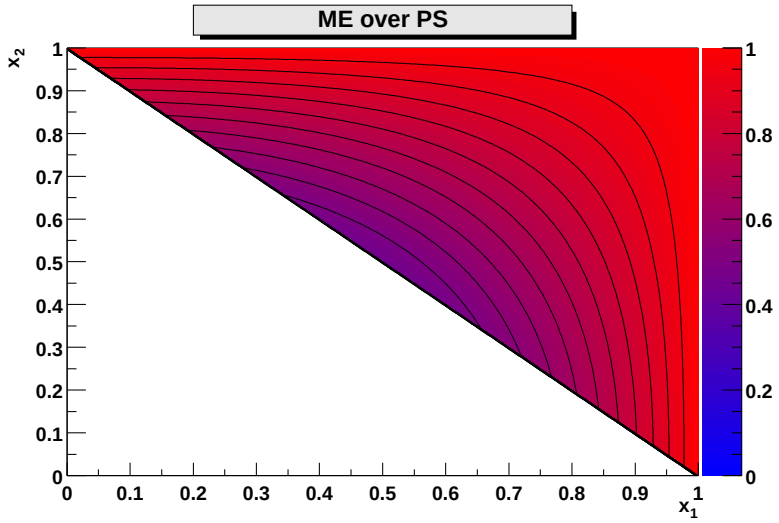


Figure 3.4: The ratio A_{ME}/A_{PS} as a function of x_1 and x_2 [69].

3.4.2 The CKKW matching prescription

The CKKW (Catani-Krauss-Kuhn-Webber) matching prescription is based on the separation of the phase space in two regions, through a $k_\perp$ measure cutoff y_{cut}: a "jet production" region, which is filled by a weighted ME and a "jet evolution" region, which is filled by the PS.

The algorithm was originally proposed for e^+e^- collisions in [68], then an extension to hadron collisions has been proposed in [69] that takes ISR into account. In the original implementation, the $k_\perp$ distance between two partons was defined according to the Durham $k_\perp$ algorithm as

$$y_{ij} = \frac{2\min\{E_i^2, E_j^2\}(1-\cos\theta_{ij})}{s}, \tag{3.19}$$

where $E_{i(j)}$ is the energy of parton i (j) and θ_{ij} is the angular separation between the two. A $k_\perp$ resolution variable y_{cut} is defined as:

$$y_{cut} = q_{cut}^2/s\ . \tag{3.20}$$

The method is based on the calculation of jet rates. Let's consider again the process $e^+e^- \to jets$ for simplicity. The lowest order contribution to this process is $e^+e^- \to q\bar{q}$. Recalling that the Sudakov form factor $S_i(q_{cut}, \sqrt{s})$ (i=q, $\bar{q}$) can be interpreted as a non-emission probability when going from scale $\sqrt{s}$ to scale q_{cut}, the two-jet rate $\mathcal{R}_2(q_{cut}, \sqrt{s})$, i.e. the probability of having two jets resolvable at scale q_{cut}, is nothing but the product of the probability that nor the quark neither the antiquark emitted resolvable radiation between $\sqrt{s}$ and q_{cut}:

$$\mathcal{R}_2(q_{cut}, \sqrt{s}) = \left[S_q(q_{cut}, \sqrt{s})\right]^2. \tag{3.21}$$

Besides, considering that in general, for a parton of type i the probability of evolving from scale $\sqrt{s}$ to scale $t > q_{cut}$ without branching is

$$S_i(t, \sqrt{s}) = \frac{S_i(q_{cut}, \sqrt{s})}{S_i(q_{cut}, t)}, \tag{3.22}$$

it can be shown that the three-jet rate, that corresponds to the emission of an additional gluon resolvable at scale q_{cut}, can be written as

$$\mathcal{R}_3(q_{cut}, \sqrt{s}) = 2\big[S_q(q_{cut}, \sqrt{s})\big]^2 \int_{q_{cut}}^{\sqrt{s}} \mathrm{d}t \Gamma_{q\to qg}(t; \sqrt{s}) S_g(q_{cut}, t), \tag{3.23}$$

where $\Gamma_{q\to qg}(t; \sqrt{s})$ is the integral over the allowed z range for the splitting functions $P_{q\to qg}$ given in Eq. (3.8). The factor 2 in Eq. (3.23) comes from the fact that the additional gluon can be emitted either from the q or the $\bar{q}$. Rates for every jet multiplicity can be written using this same technique.

The CKKW prescription aims at giving a more refined description of jet rates with respect to that of parton shower. The basic idea is to replace the $\Gamma_{q\to qg}(t; \sqrt{s})$ functions that appears in the rate expressions with the corresponding ME. Let's consider again the three-jet rate. Let's suppose we produced a momentum configuration according to the three-parton matrix element, with a y_{cut} cutoff; this means that the parton momenta produced are such that 3 partons are resolved at scale q_{cut}. The scale t at which the gluon emerged from either the q or $\bar{q}$ lines can be calculated as $\min(y_{qg}, y_{\bar{q}g}) = t^2/s$. Then, according to equation Eq. (3.23), the weight to be applied to this ME parton configuration is

$$F_{q\bar{q}g} = \big[S_q(q_{cut}, \sqrt{s})\big]^2 S_g(q_{cut}, t). \tag{3.24}$$

This weight, applied on an event by event basis, allows to reproduce the three jet rate of Eq. (3.23), with the exact matrix element in place of the collinear approximation splitting function $\Gamma_{q\to qg}(t; \sqrt{s})$.

This can be easily extended to any parton multiplicity and it improves the description of configurations with n partons resolved at scale y_{cut} making use of the n-parton ME.

So far we took care of the phase space region above y_{cut}. The description of the phase space below the cut is done with plain parton shower. The shower is however vetoed so that any hard emission above y_{cut} is forbidden. It was shown in [68] that the combined ME weighting plus vetoed PS cancels the dependency on the parameter y_{cut} up to the NLL accuracy.

The CKKW approach has been extended also to hadron-hadron collision and the cancellation of the y_{cut} dependency is reasonable, even if no proof exists so far that demonstrates at which order the cancellation is achieved. The main difference with respect to the e^+e^- case is that the $k_\perp$ measure is replaced with its longitudinally invariant version

$$y_{ij} = \frac{\min(p_{T,i}^2, p_{T,j}^2)(\Delta y^2 + \Delta\phi^2)/D^2}{\hat{s}}, \tag{3.25}$$

where Δy^2 is the rapidity distance and $\Delta\phi^2$ is the distance along the azimuthal angle; D is a parameter usually taken to be 1; $\hat{s}$ is the energy in the center of mass of the colliding partons. The whole CKKW prescription can be summarized in the following steps:

1. ME cross sections $\sigma_{n,i}$ are calculated for each parton multiplicity n and for each different combination i of partons that contributes to multiplicity n. A cutoff y_{cut} on the separation of partons is applied to avoid divergences. A fixed ${\alpha_S}^{ME}$ is used.

2. One among configurations n, i is selected with probability $P_{n,i} = \sigma_{n,i} / \sum_{m,j} \sigma_{m,j}$.

3. Parton momenta are generated according to the corresponding matrix element squared.

4. The scales at which the splittings happened are reconstructed: this is achieved through a $k_\perp$ clustering of the partons emerging from the ME. The clustering is stopped when the core $2 \to 2$ process is found. This leads to a series of $n-2$ clusterings with associated values of the $k_\perp$ distance $y_2 ... y_n$. Once the values $y_i = t_i^2/s$ are known we can finally calculate the ME event weight, that comes in two factors:

   - an α_S correction: for each clustering i an α_S correction $\alpha_S(t_i)/\alpha_S^{ME}$ is applied;
   - a Sudakov form factor correction is applied, like the one exemplified in Eq. (3.24) for the 3 jet case.

5. Events are accepted or rejected according to their weight.

6. The accepted events are showered with a veto on the emission above y_{cut}.

The matrix element calculation is implemented in `SHERPA` by the module called `AMEGIC++` [70], which uses the helicity amplitudes method. Parton showering is done by the `APACIC++` [71] shower module, that implements a virtuality ordered PS with coherence effects explicitly included (similar to the `PYTHIA` PS).

3.4.3 The MLM matching prescription

The MLM (M. Mangano) prescription implemented in `AlpGen` does not require a dedicated PS program, and can be used both with `PYTHIA` and `HERWIG` showers.

The MLM prescription is similar to the CKKW up to the α_S correction. As in CKKW, ME cross sections are calculated up to the maximum parton multiplicity that the user wants in the final state; a minimum p_T cut for final state partons is used to cutoff ME divergences and a fixed α_S is used. As in CKKW a "PS history" is reconstructed and a splitting sequence is identified, with corresponding scales; an

α_S correction is applied as in the CKKW. From this point the two prescriptions become different. As said, in the MLM approach a conventional PS program is used (PYTHIA or HERWIG) and ME partons are showered without any constraint. The parton collection that results from the PS step is clustered using a jet algorithm (a cone in the AlpGen implementation, but also other options have been investigated, e.g. a $k_\perp$ algorithm is used in MADGRAPH/MADEVENT); the resulting jets are matched in angle to the ME partons and only those events in which all the jets match to the ME partons without any extra unmatched jets are retained (for the maximum ME parton multiplicity additional jets, softer than the matched ones, are allowed).

This procedure tries to reproduce in one go the effect that in the CKKW is achieved in two steps: the Sudakov re-weighting and the vetoed shower. Indeed the rejection of events with additional jets should, at the same time, reject ME configuration in a similar way as the Sudakov weight does and prevent additional emission from the shower, thus reproducing the effect of the CKKW PS veto.

The MLM prescription is really convenient because it does not require modifications in the PS program, but it just requires a veto routine to kill events not fulfilling the matching criteria.

While the CKKW prescription contains one parameter (the q_{cut} of the internal $k_\perp$ clustering algorithm), in the MLM the user has to choose different parameters. The cone algorithm used for the matching has three parameters, namely the minimum jet p_T the cone radius R, and the jet maximum pseudorapidity η. The minimum p_T used in the cone clustering ($p_{T\,min}^{jet}$) is not the same as the minimum p_T used in the ME step to cutoff divergences ($p_{T\,min}^{ME}$): usually it is recommended to have $p_{T\,min}^{jet} > p_{T\,min}^{ME}$; this is needed because events that are below the cut at the ME level could fall above after the PS. For this reason a process dependent tuning for $p_{T\,min}^{jet}$ with respect to $p_{T\,min}^{ME}$ is needed; for the $Z/\gamma^* + jets$ production AlpGen authors recommend to choose the jet finder minimum p_T to be 5 GeV higher than the ME minimum p_T.

3.5 Hadronization

After the Parton Shower step of the event generation we are left with a set of partons with virtualities of the order of the cutoff scale at which the shower was stopped. Hadronization is the step in which partons are turned into hadrons. The process is non-perturbative and at present is described by models. Currently, two models are the most effective at describing data: the Lund string model, implemented in PYTHIA, and the cluster model implemented in HERWIG. We shall now briefly describe the main ideas behind the two approaches.

3.5.1 Lund string model

Quark and antiquarks produced in the shower move apart from each other transferring part of their energy to the colour field that connects them. As they move apart the color field lines tighten and acquire a string shaped configuration; the energy stored per unit length in the colour field tends to be uniform, as shown in Fig. 3.5. When enough energy is stored in the string it can break up into a quark

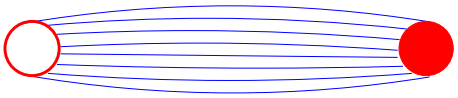


Figure 3.5: Schematic representation of the color field as the string forms.

antiquark pair. With this simple mechanism the formation of mesons is described. The flavour of the $q\bar{q}$ pair that results from the string break up is assigned with probabilities tuned to data. The formation of baryons is more complex and it requires considering a three quark final state in which two of the quarks are close and form one of the two end points of the string.

3.5.2 Cluster model

The cluster model is based on the observation that colour connected partons have a very rapidly falling invariant mass spectrum. This effect is often referred to as "pre-confinement" and it suggests a hadronization model in which color singlets, called "clusters", are formed out of the partons emerging from the Parton Shower. Clusters then decay to lighter resonances; in the decay the spin information is assumed to be washed out; heavier hadrons and baryons are automatically suppressed kinematically. This hadronization model has fewer parameters with respect to the Lund model.

3.6 Multiple parton interactions

In each hadron hadron collision several (∼4-5 at LHC) parton scatterings happen simultaneously. The cross section for parton-parton scatterings is dominated by t-channel $2 \rightarrow 2$ low p_T gluon exchange. When simulating realistic events, the generation of the hard, large momentum transfer, process has to be accompanied by additional soft parton-parton interactions that build up the underlying structure of the event ("underlying event"). The number of additional interactions is chosen assuming a Poissonian distribution, with mean number of events tuned to data.

The underlying event physics is the least understood and its simulation is heavily relying on models, so extrapolation to different energies (Tevatron to LHC for example) is difficult and uncertain.

Generators used in this work describe the underlying event in different ways. While PYTHIA does not apply the PS to the partons emerging from underlying event simulation SHERPA does. This leads to a very different amount of radiation produced by the two generators for the same number of multiple parton interactions, with SHERPA being the hardest. For this reason SHERPA needs a fewer mean number of multiple interactions than PYTHIA to reproduce presently available data.

Chapter 4

First order corrections to Z/γ^* inclusive production

In this chapter we compare different matching algorithms on the ground of first order real corrections to the inclusive Z/γ^* production.

We consider parton level observables, which means that we stop the simulation of the event before the hadronization step; in this way we intend to avoid the smearing effects due to the hadronization and thus we can look in detail at the perturbative phase of event generation.

We consider the first order corrections to the inclusive Z/γ^* production. The lowest order graph for this process is $q\bar{q} \to Z/\gamma^*$, with the first order correction given by the graphs $q\bar{q} \to Z/\gamma^* g$, and $qg \to Z/\gamma^* q$: in both cases the additional radiation comes from the initial state. As pointed in Sec. 3.4.1, `PYTHIA` is able to fully take into account matrix element corrections for the first emission.

We investigate the effect of deactivating ME-corrections in `PYTHIA` for a number of observables. We also compare distributions obtained with `PYTHIA` with those obtained with `HERWIG`, as the latter generator implements ME corrections in a way that is similar to `PYTHIA`.

We use ME-corrected `PYTHIA` production as a reference for a consistency check of the two other matching prescriptions used in this work. Both CKKW and MLM claim that they can describe multiple parton emission corrected for the corresponding multi-parton matrix element. If just one additional parton emission from the matrix element is permitted, those prescriptions should give results compatible with `PYTHIA`. This test is particularly sensitive to possible dependences of CKKW and MLM on the resolution parameter used to separate ME and PS regions. The ME correction prescription implemented in `PYTHIA` does not depend on any parameter and thus it provides us with the "correct" reference to spot dependences on the ME-PS separation parameter used in the two other prescriptions.

Results presented in most of this and the next chapters were obtained running an analysis coded with the Rivet [72] Monte Carlo validation software. Rivet is a tool that provides a general framework for event generators validation and tuning.

More information on Rivet will be provided in Sec. 5.4.

4.1 Matrix element corrections in PYTHIA

Matrix element corrections for inclusive Z/γ^* production in PYTHIA modifies the first emission from the PS so that the effect of the first order correction is reproduced. The lowest order graph contributing to Z/γ^* production is shown in Fig. 4.1. The processes that contribute to the first order correction are $q\bar{q} \to Z/\gamma^* g$, represented by the graphs in Fig. 4.2, and $qg \to Z/\gamma^* q$, represented by the graphs in Fig. 4.3.

As already mentioned in Sec. 3.4.1 the correction comes as a re-weighting factor of the parton shower. The initial state parton shower is weighted with two factors, one to reproduce the matrix elements of Fig. 4.2, $W_{q\bar{q}\to Z/\gamma^* g}$, and one to reproduce the matrix element of Fig. 4.3, $W_{qg\to Z/\gamma^* q}$ [66]. The PS emission closest to the hard

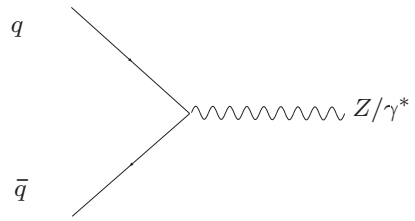


Figure 4.1: Lowest order contribution to the Z/γ^* production.

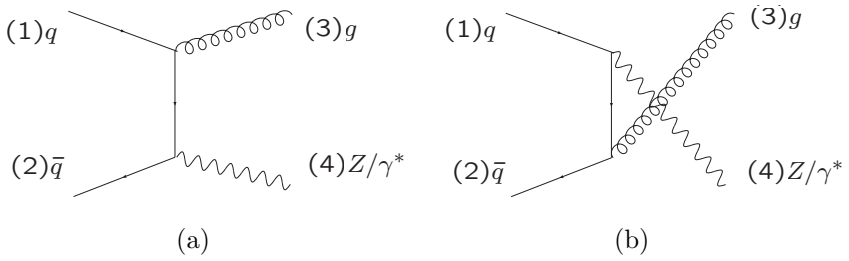
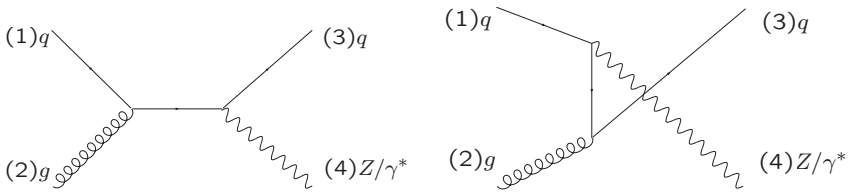


Figure 4.2: Graphs contributing to the $q\bar{q} \to Z/\gamma^* g$ process.

$q\bar{q}$ process is the one that gets the correction. In order to correct the PS we need to classify this emission as either $(q\bar{q} \to Z/\gamma^* g)$-like or $(qg \to Z/\gamma^* q)$-like. A PS branching like the one depicted in Fig. 4.4 is considered $(q\bar{q} \to Z/\gamma^* g)$-like, while a branching like the one in Fig. 4.5 is considered $(qg \to Z/\gamma^* q)$-like.

The ratio of the ME and PS differential cross sections, classified as described



(a) (b)

Figure 4.3: Graphs contributing to the $qg \to Z/\gamma^* q$.

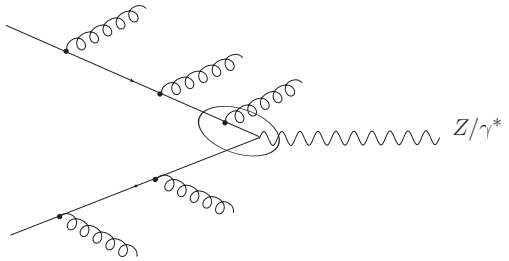


Figure 4.4: The circled PS branching is considered of type $q\bar{q} \to Z/\gamma^* g$.

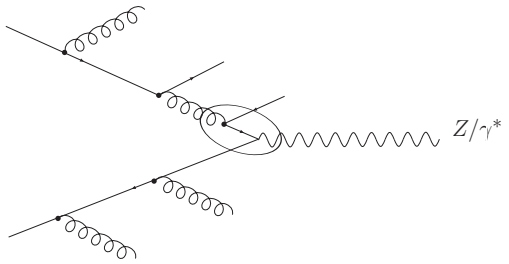


Figure 4.5: The circled PS branching is considered of type $qg \to Z/\gamma^* q$.

above, leads to the following expressions for $W_{q\bar{q}\to Z/\gamma^* g}$ and $W_{qg\to Z/\gamma^* q}$ [66]:

$$W_{q\bar{q}\to Z/\gamma^* g} = \frac{\hat{t}^2 + \hat{u}^2 + 2m_Z^2\hat{s}}{\hat{s}^2 + m_Z^4}, \tag{4.1}$$

$$W_{qg\to Z/\gamma^* q} = \frac{\hat{s}^2 + \hat{u}^2 + 2m_Z^2\hat{t}}{(\hat{s} - m_Z^2)^2 + m_Z^4}, \tag{4.2}$$

where $\hat{s}$, $\hat{t}$ and $\hat{u}$ are the Mandelstam variables. It can be shown that

$$\frac{1}{2} < W_{q\bar{q}\to Z/\gamma^* g} < 1, \tag{4.3}$$

and

$$1 < W_{qg\to Z/\gamma^* q} < 3. \tag{4.4}$$

This means that the Parton Shower alone would overestimate the $q\bar{q} \to Z/\gamma^* g$ contribution and underestimate the $qg \to Z/\gamma^* q$ contribution. One explanation for the latter is that the s-channel graph in Fig. 4.3 (a) is not present in the PS.

We will now see in detail the effect of ME-corrections on various observables. We should stress that the ME corrected distributions are the best description that PYTHIA can give of existing data. Uncorrected result does not describe existing data equally well, so only the ME corrected result should be trusted when making comparisons to data. The reason why we are interested in evaluating how distributions change switching off the ME corrections is only to characterize this effect. Another reason for looking at the uncorrected result is that, as we shall point in Sec. 4.4, when PYTHIA is used to shower events produced in AlpGen, ME corrections are switched off because AlpGen itself is going to introduce its own corrections.

4.1.1 ME correction effect on lepton observables

The p_T distribution for the lepton pair for inclusive Z/γ^* production in PYTHIA is shown in Fig. 4.6. The lepton pair invariant mass has been generated around the Z resonance, between 66 and 116 GeV. Only the electron decay channel has been selected.

The three curves correspond to three different configurations: one is with ME corrections activated, the other two are obtained without matrix element corrections, but with different starting scales for the shower: the total hadronic center of mass energy and the invariant mass of the lepton pair respectively.

As already mentioned in Secs. 3.3.1 and 3.3.2, PYTHIA implements a virtuality ordered parton shower. The starting scale of the shower marks the maximum allowed virtuality in the shower evolution. If the starting scale is set to M_Z the hardest parton transverse momentum cannot exceed M_Z, and thus also the Z p_T cannot exceed 90 GeV approximately, as shown in Fig. 4.6. When the starting scale is raised, the spectrum gets harder. When ME corrections are activated the spectrum gets even harder. The reason why the ME corrected spectrum is harder than the uncorrected one can be explained considering the relative amount of the two corrections at the LHC. The graph $gq \to Z/\gamma^* q$ contributes more than $q\bar{q} \to Z/\gamma^* g$, because $\bar{q}$ is not a valence quark at LHC. We recall that Eq. (4.4) states that $W_{gq \to Z/\gamma^* q} > 1$, meaning that the first emission from the ME-corrected shower is done with a splitting probability higher than that of the uncorrected shower. Since the PS emission is always ordered, a higher splitting probability means that the probability for emitting a harder parton is higher than that of uncorrected shower. This explains why the corrected spectrum is harder than the uncorrected one. We also notice in passing that at the Tevatron $p\bar{p}$ collider [73–75] exactly the opposite holds: in that case $q\bar{q} \to Z/\gamma^* g$ dominates because $\bar{q}$ is a valence quark; since the correction is < 1 for this graph (Eq. (4.3)) the ME corrected result is softer than the corrected one, as shown in [66].

One might naively expect that ME corrections should change the shape of spec-

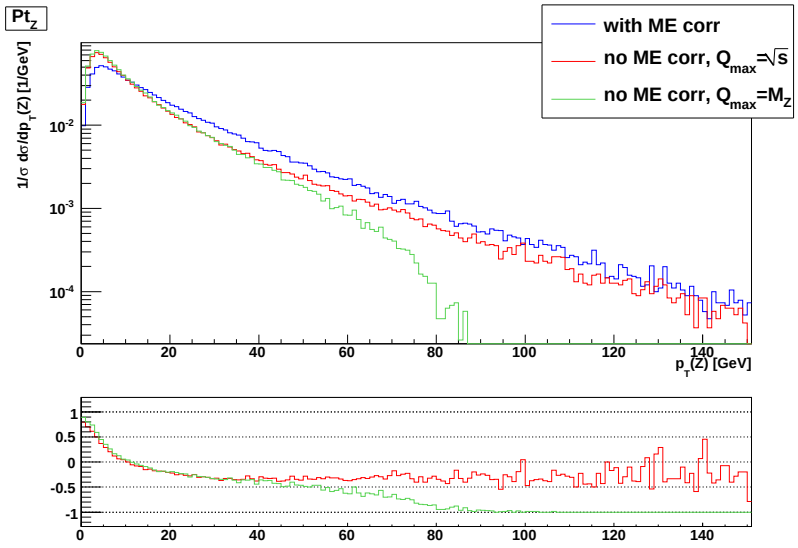


Figure 4.6: p_T spectrum for the lepton pair in PYTHIA for three different settings of the ISR: with ME corrections, without ME corrections and with the starting scale of the shower set to $\sqrt{s}$, without ME corrections and with the starting scale of the shower set to M_Z. The relative difference with respect to the curve with ME corrections is shown in the lower plot.

tra only at high p_T; low p_T region should be well described by the parton shower alone. Actually the Z p_T spectrum is altered by ME corrections all over the p_T range, as shown in Fig. 4.6. The three distributions are normalized, but the difference at low p_T is not only due to normalization; a change in shape is also present; such a change is testified by the relative difference plot, that does not flatten as p_T approaches zero. The reason for the change in shape at low p_T is that ME corrections change the Sudakov form factor used in the shower [76].

Fig. 4.7 shows the pseudorapidity distribution of the Z boson for the three configurations. The configuration with ME-corrections shows the most central Z, while the configuration with starting shower scale set to M_Z and without ME corrections shows the less central distribution; this is consistent with what was observed on the p_T spectrum.

The p_T distribution for the positive lepton is shown in Fig. 4.8. This distribution confirms that the ME corrected spectrum is the hardest while the uncorrected one with starting scale set to M_Z is the softest.

4.1.2 ME correction effect on jet observables

We will now see directly how the ME corrections change the jet observables. First of all we describe how jets are reconstructed throughout this work. Jets are reconstructed with the longitudinally invariant $k_\perp$ algorithm, as described in [77] and in Sec. 1.4.4. We used the inclusive variant of the algorithm, which means that

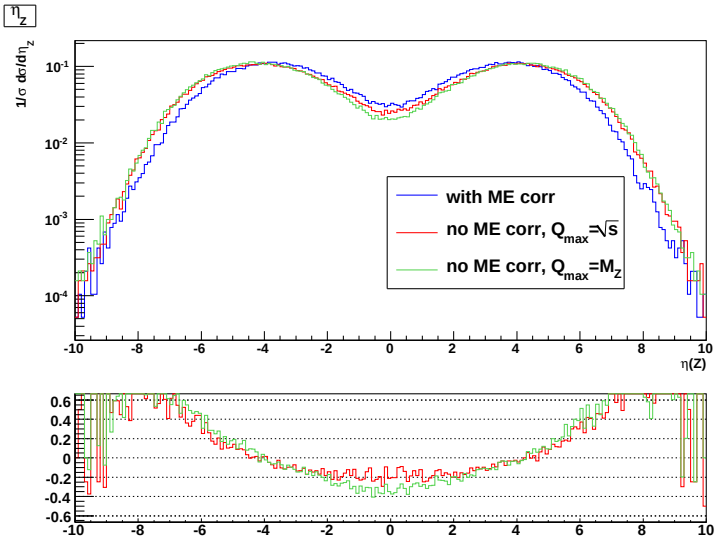


Figure 4.7: Pseudorapidity (η) distribution for the lepton pair in PYTHIA with three different ISR settings: with ME corrections, without ME corrections with starting scale set to $\sqrt{s}$, without ME corrections with starting scale set to M_Z. The relative difference with respect to the ME corrected curve is also shown.

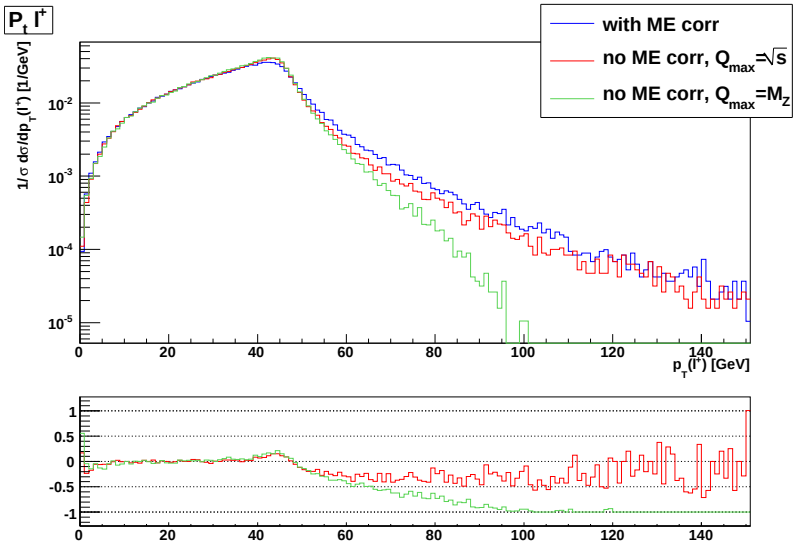


Figure 4.8: p_T distributions for the positively charged lepton.

clustering is performed until all the particles have been assigned to a jet and then a minimum p_T cut for jets is applied. Here we recall the $k_\perp$ distance definitions; a beam distance d_{iB} is defined for object i (i can be either a particle or a pseudo-jet that has been created during the clustering procedure summing the four-momenta of several particles):

$$d_{iB} = p_T^2; \tag{4.5}$$

a distance between to particles or pseudo-jets i and j, d_{ij}, is defined as:

$$d_{ij} = \min({p_T}_i^2, {p_T}_j^2)\frac{\Delta y^2 + \Delta\phi^2}{D^2}, \tag{4.6}$$

where Δy and $\Delta\phi$ are the rapidity and azimuthal differences between the two objects respectively. The D constant marks the minimum distance between two jets and has been set to 0.4 for this study; the minimum p_T for jets has been set to 30 GeV.

Fig. 4.9 shows the jet multiplicity for the three configurations described above. The fact that the number of jets is minimum for the sample without ME corrections

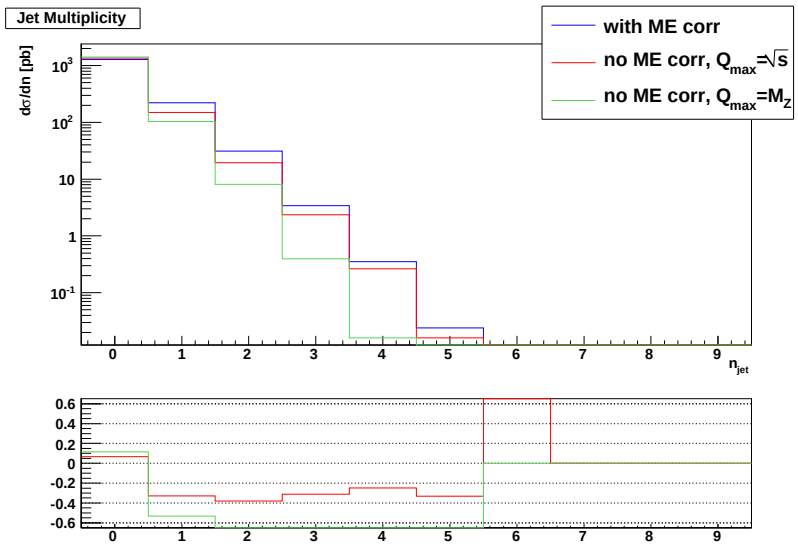


Figure 4.9: Jet multiplicity for the three different settings of the shower and ME corrections in PYTHIA.

and low starting shower scale, and maximum for the sample with ME corrections is consistent with what has been observed on the Z p_T spectrum; the largest the radiation the highest the number of jets; thus also the Z p_T spectrum is the hardest since the Z recoils against the QCD radiation.

Fig. 4.10 shows the p_T spectrum of the leading jet; in this case we plot the differential cross section distribution, instead of the normalized distribution, as we are interested in estimating the total amount of radiation. The behavior is consistent with the expectations; the ME corrected sample shows the highest amount of radiation, the uncorrected sample with scale set to M_Z shows the lowest.

To test how the phase space available for QCD radiation is filled we looked at differential jet rates. The differential jet rate $n \to n-1$ is the distribution of the $n \to n-1$ transition value, $Q_{n\to n-1}$, which is the value of the resolution parameter d_{cut} (in an exclusive $k_\perp$ algorithm) for which an n jet event turns into an $n-1$ jet event. Naively the $n \to n-1$ transition value could be found running the $k_\perp$ algorithm in exclusive mode (Sec. 1.4.4) with increasing d_{cut} and finding the value for which the requested transition happens. Actually it is enough to run the

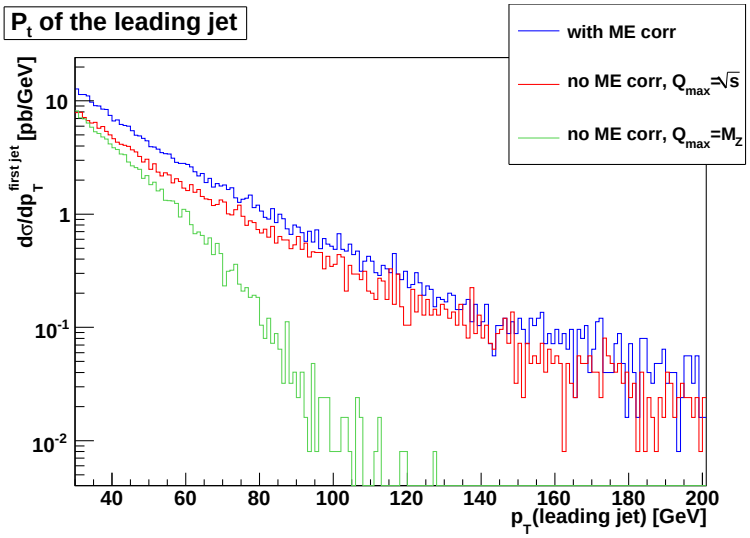


Figure 4.10: Leading jet p_T spectrum for three different settings of the shower and of the ME corrections in PYTHIA.

algorithm in inclusive mode and look at the clustering sequence and find the step for which two pseudo-jets get recombined into one or one pseudo-jet is closest to the beam line than to any other jet, and thus would be recombined with the beam-jet in an exclusive approach.

To understand what a differential jet rate is let's consider a simple example with three particles in the final state. Let the particles be 1, 2, 3. Suppose that the sequence of ordered distances looks like this:

$$d_{1,B} < d_{2,B} < d_{1,2} < d_{3,B} < d_{1,3} < d_{2,3}. \tag{4.7}$$

In an exclusive calculation the first jet that would be recombined with the beam line would be 1, thus the $3 \to 2$ transition value is $Q_{3\to 2} = d_{1,B}$. Then particle 2 is the next one to be recombined with the beam, thus $Q_{2\to 1} = d_{2,B}$. At this point, if the d_{cut} is raised to be at least $d_{2,B}$ only $d_{3,B}$ survives in Eq. (4.7), thus $Q_{1\to 0} = d_{3,B}$.

Let's now consider another example sequence:

$$d_{1,B} < d_{1,3} < d_{1,2} < d_{2,3} < d_{2,B} < d_{3,B}. \tag{4.8}$$

As before, $Q_{3\to 2} = d_{1,B}$. If we raise d_{cut} to at least $d_{1,B}$ particle 1 gets clustered with the beam and the new sequence looks like

$$d_{2,3} < d_{2,B} < d_{3,B}, \tag{4.9}$$

so particles 2 and 3 are going to be clustered in the next step. Thus if d_{cut} is set to be at least equal to $d_{2,3}$ particles 2 and 3 are clustered, thus passing from 2 to 1 jet. Thus $Q_{2\to 1} = d_{2,3}$ After they are clustered the sequence will have one only element, namely $d_{(2,3),B}$; thus $Q_{1\to 0} = d_{(2,3),B}$.

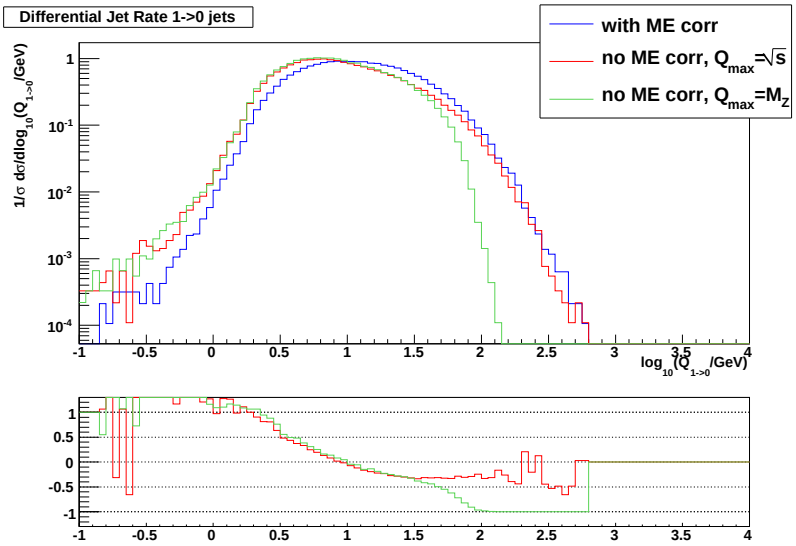


Figure 4.11: Distribution for the $1 \to 0$ differential jet rate.

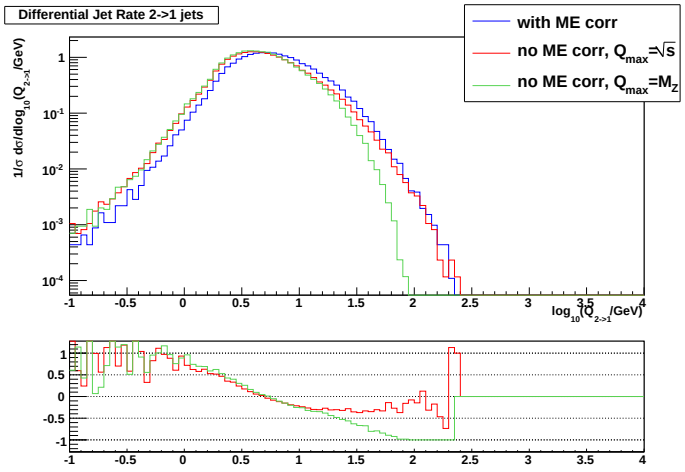
Let's now see how the differential jet rates look for the three setting of ME corrections. The $1 \to 0$ differential jet rate is shown in Fig. 4.11. The main differences between the three settings are in the region for high values of $Q_{1\to 0}$. We see that the sample without ME corrections and with low starting shower scale is the one that dies first. This means that it is unable to fill the phase space for hard parton emission, which is responsible for filling the rightmost part of the plot. We see that the plot ends at about 2.1, which corresponds to a $Q_{1\to 0}$ $\sim$130 GeV, which is where also the leading jet p_T distribution in Fig. 4.10 ends. For what concerns the other two settings we observe that they are similar, but while the sample without ME corrections tends to fill the region below 1, the one with ME corrections fills the region above 1 more, thus allowing for more radiation to be emitted.

Fig. 4.12 show the differential jet rates for the transition $2 \to 1$ and $3 \to 2$; In this case the differences among the three settings are a bit tamed due to the fact that additional emission other than the first one is anyway uncorrected for ME effects.

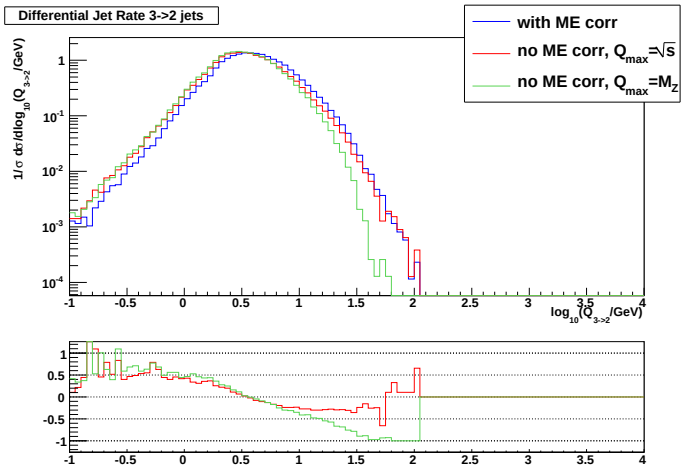
4.2 Matrix element corrections in HERWIG

HERWIG implements ME corrections in a way that is similar to PYTHIA. There are two main differences between HERWIG and PYTHIA that are relevant for ME corrections:

- in HERWIG the PS is angular ordered, which means that the early emissions are soft large angle gluons; thus, the emission that deserves the corrections is not the first like in PYTHIA, as pointed in [65];



(a)



(b)

Figure 4.12: Differential jet rates for the transitions $2 \to 1$ (a) and $3 \to 2$ (b). The effect of different ME correction settings is shown.

- in HERWIG the PS cannot fill the phase space for values above the hard scale of the process; in this respect it is similar to PYTHIA with starting scale set to M_Z.

For these reasons a two-step approach is used in HERWIG to implement ME corrections [78]. In the phase space region covered by the PS, corrections are applied as in PYTHIA, the only difference being that not the first emission, but rather the *hardest emission so far* during the shower evolution gets the correction [65]. These corrections are referred to as "soft ME corrections". In the "dead zones", that are left completely uncovered by the PS, the exact ME for one additional parton is used, with subsequent PS. These corrections are referred to as "hard ME corrections".

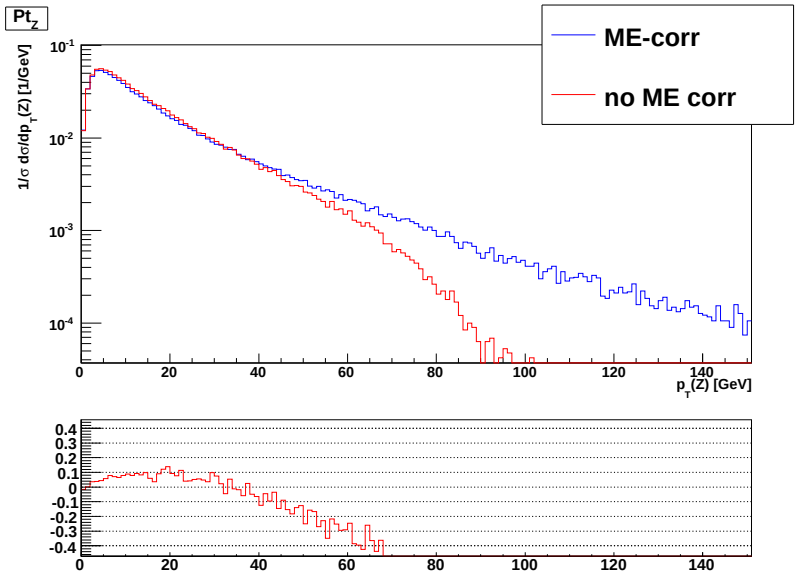


Figure 4.13: ME corrections effect in HERWIG in the Z p_T distribution. The effect at low p_T is small, while in PYTHIA the shape was different at low p_T as well.

The effect of deactivating ME corrections in HERWIG is shown in Fig. 4.13 for the Z p_T. While ME corrections in PYTHIA change the whole shape of the distribution, also at low p_T, the effect of ME corrections at low p_T in HERWIG is small, the change in shape is small.

A comparison between ME corrected Z p_T distribution in PYTHIA and HERWIG is shown in Fig. 4.14. The agreement is very good all over the p_T spectrum.

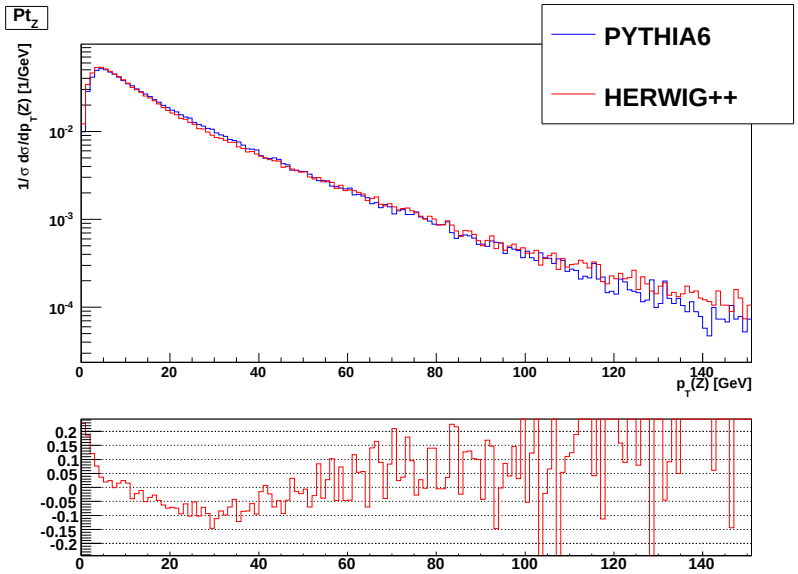


Figure 4.14: ME corrected Z p_T distribution in PYTHIA and HERWIG.

4.3 A comparison between SHERPA and PYTHIA

Since PYTHIA is able to correct the first emission for ME effects it represents a good test-bed for procedures that claim to be able to correct also for further emission; if only the matrix element for one additional parton is allowed to enter the calculation the result should be close to PYTHIA. Besides, while both the CKKW and MLM prescriptions depend on a separation parameter between the region filled by the ME and the one filled by the PS, the ME correction algorithm in PYTHIA is parameter free, and thus it represents a good reference to spot dependencies on the separation parameter in the two other prescriptions. A dependency on the separation parameter in CKKW and MLM is unavoidable, but it should not be heavy; we will evaluate to what extent the dependency on the parameter is cancelled. We recall here that for e^+e^- collisions the parameter dependency in CKKW cancels to the NLL accuracy according to [79]. For hadron collision no theoretical proof exists so far that establishes the limit to which the parameter dependency is cancelled either in CKKW or in MLM. So it's worth checking directly if such cancellation occurs.

4.3.1 Total cross section

In PYTHIA the total cross section for inclusive Z/γ^* is calculated at Leading Order accuracy, i.e. it is calculated from the process $q\bar{q} \to e^+e^-$; the subsequent PS, either corrected or uncorrected, happens with probability 1, so it cannot modify the cross section. In SHERPA the cross section comes as a sum over the different selected final state parton multiplicities.

$$\sigma_{\texttt{SHERPA}} = \sum_{i=0}^{N} \sigma_i < w_i > \tag{4.10}$$

where σ_i is the cross section for i additional partons in the final state and $< w_i >$ is the average Sudakov weight for that configuration.

Table 4.1 shows the cross section values as obtained in PYTHIA and in SHERPA for four different values of the resolution cut q_{cut} that steers the separation between the ME and the PS regions.

The difference in the total cross section with respect to PYTHIA is up to about 10%, for the sample with the lowest value of q_{cut}. Cross sections for both SHERPA and PYTHIA are formally LO; some differences are due to the q_{cut} dependency mainly. If one takes a very high value for q_{cut}, this makes SHERPA get closer and closer to PYTHIA. In fact, as q_{cut} is increased the contribution to the total cross section from the configuration with one additional parton vanishes, thus leaving the leading order contribution alone, that is the only one considered in PYTHIA for the cross section calculation. In the sample with q_{cut}=500 GeV the difference with respect to PYTHIA goes down to 5%, and this residual difference is most probably due to different scale choices.

SHERPA		σ_i [pb]	$< w_i >$	Total σ [pb]
q_{cut}=10GeV	0 jet	838.9	0.7489	1383
	1 jet	998.7	0.7559	
q_{cut}=20GeV	0 jet	1059.5	0.9301	1405
	1 jet	484.6	0.8657	
q_{cut}=40GeV	0 jet	1271.2	0.9992	1434
	1 jet	177.2	0.9267	
q_{cut}=500GeV	0 jet	1926.6	0.7540	1453
	1 jet	0.038	0.9802	
PYTHIA				Total σ [pb]
inclusive				1528

Table 4.1: Cross sections for SHERPA and PYTHIA.

4.3.2 Lepton observables

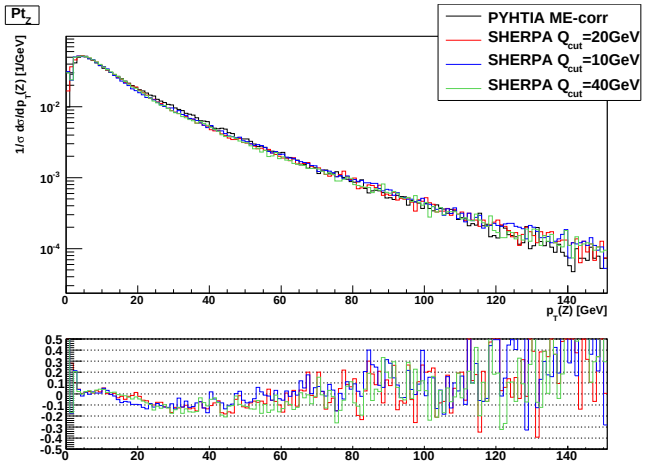
The p_T spectrum for the lepton pair in SHERPA and PYTHIA is shown in Fig. 4.15 (a). As mentioned above, SHERPA has been run such that only one additional parton can be emitted from the matrix element.

SHERPA has been run with three different values of the parameter q_{cut} that governs the separation between the phase space region filled by the ME and the region filled by the PS. The values used were q_{cut}= 10, 20, 40 GeV. As described in Sec. 3.4.2 the region above q_{cut} is filled by a modified ME, in which Sudakov form factors are attached to the ME, while the region below the cut is filled by a vetoed PS. SHERPA appears to closely follow the PYTHIA spectrum. Some discrepancies are observed in the first few bins. These are most probably due to a different tuning for the primordial p_T[1] distribution of partons in the protons. We notice in particular that the high p_T tail, that is sensitive to additional hard emission, appears to be correctly reproduced. We also observe a small dependence on the value of q_{cut}: the three curves for SHERPA agree within few percent.

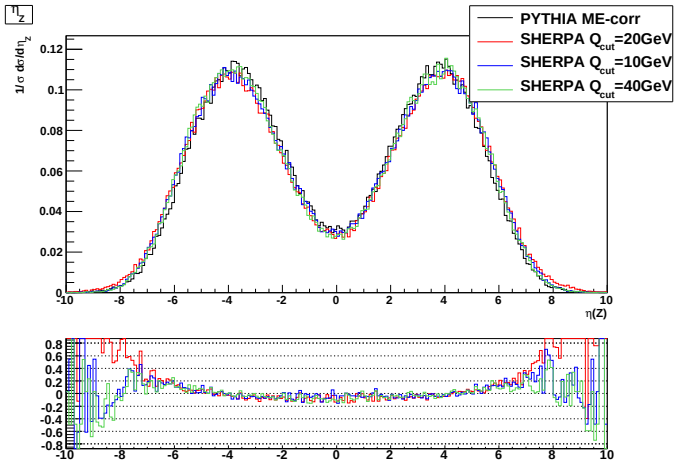
The pseudorapidity η distribution for the lepton pair is shown in Fig. 4.15 (b). The distributions from SHERPA well agree with PYTHIA in the central region, where the difference is within 10-15%; in the tails of the distribution difference is more evident.

The p_T distribution for the positive lepton is shown in Fig. 4.16. In this case too SHERPA agrees with PYTHIA, without strong dependency on the resolution cut q_{cut}.

[1]The primordial p_T distribution of partons in protons, often referred to as primordial $k_\perp$ is the transverse momentum distribution of partons in the hadrons entering the collision.



(a)



(b)

Figure 4.15: p_T spectrum (a) and η distribution (b) for the lepton pair in PYTHIA and SHERPA. The latter has been run with at most one additional parton from the ME; three different values for the separation cut between ME and PS regions have been used: q_{cut}= 10, 20, 40 GeV.

4.3.3 Jet observables

For what concerns the jet observables a powerful test of the matching algorithm is the differential jet rate for the transition $1 \to 0$. Fig. 4.17 shows the distribution for the rate $1 \to 0$ in PYTHIA and in SHERPA. For SHERPA we tried three different values for the matching parameter q_{cut}: 10, 20, 40 GeV. The vertical lines shown in the plots indicate the position of the resolution parameter q_{cut}. In SHERPA, the phase space above q_{cut} is filled by the Matrix Element, while the region below q_{cut} is filled by the vetoed shower.

SHERPA agrees well with PYTHIA. The transition between the ME- and PS-

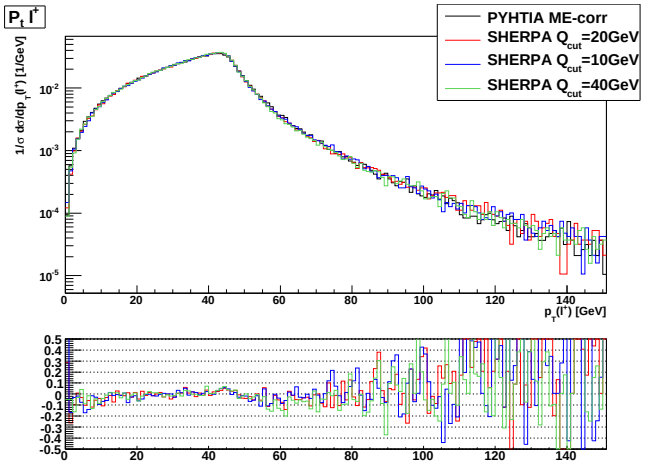


Figure 4.16: p_T distributions for the positive lepton.

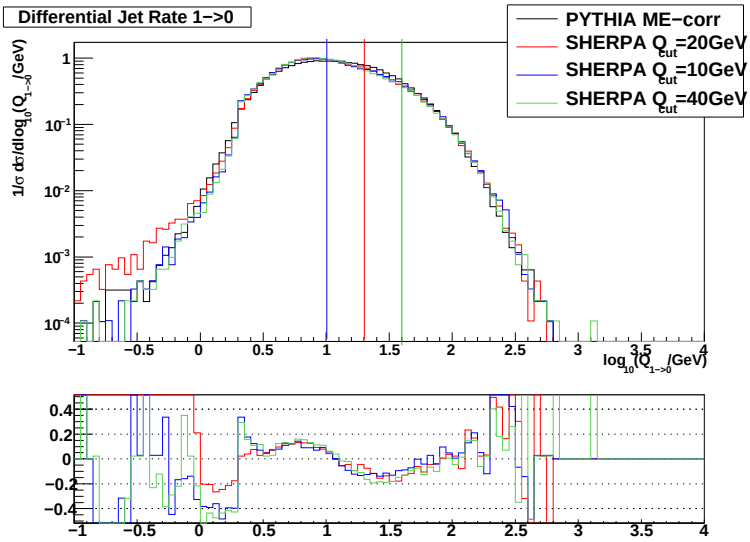
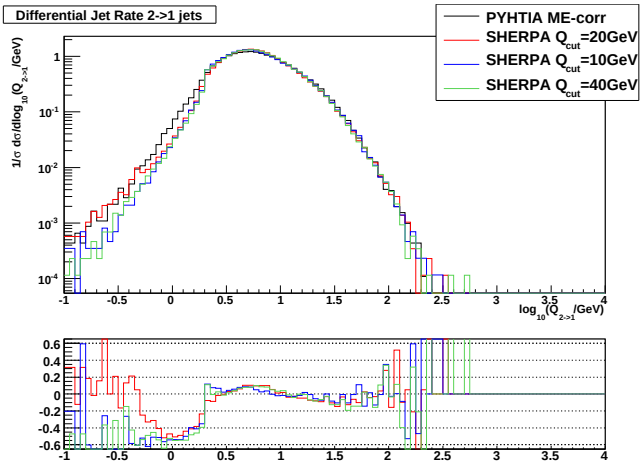


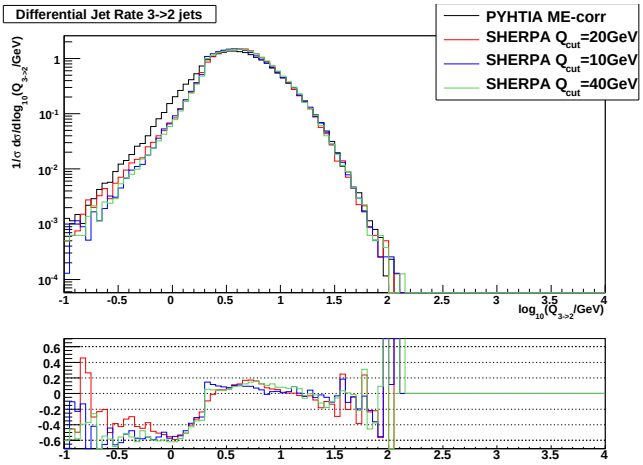
Figure 4.17: Differential jet rate $1 \to 0$ in PYTHIA and SHERPA. For SHERPA we used three values for the matching parameter q_{cut}: 10, 20, 40 GeV. Relative differences are calculated with respect to PYTHIA.

populated regions is quite smooth for all the three values used for q_{cut}. The relative difference with respect to PYTHIA is at most 20%, in the vicinities of the q_{cut} region. The tail for very low values of $Q_{1\to 0}$ shows some differences, that can be due to slightly different settings for the primordial $k_\perp$.

For what concerns the rates $2 \to 1$ and $3 \to 2$, those are shown in Fig. 4.18. The difference with respect to PYTHIA are of the same order or less than in the case of the $1 \to 0$ transition; apart from a discrepancy in the low Q region, that is motivated by the same effect as in the rate $1 \to 0$, the three curves for SHERPA agree very well, and the relative difference with respect to PYTHIA is within 15%.



(a)

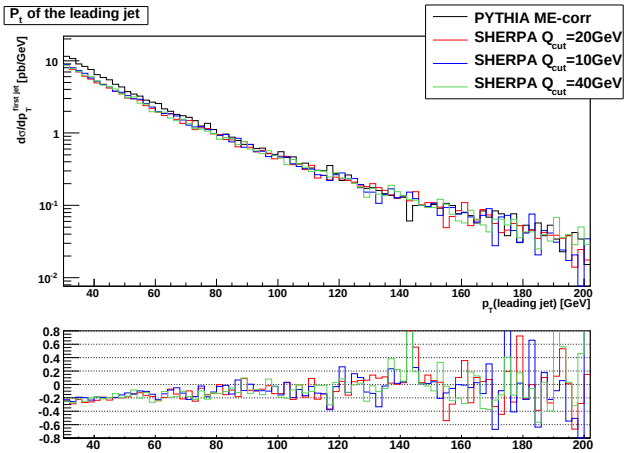


(b)

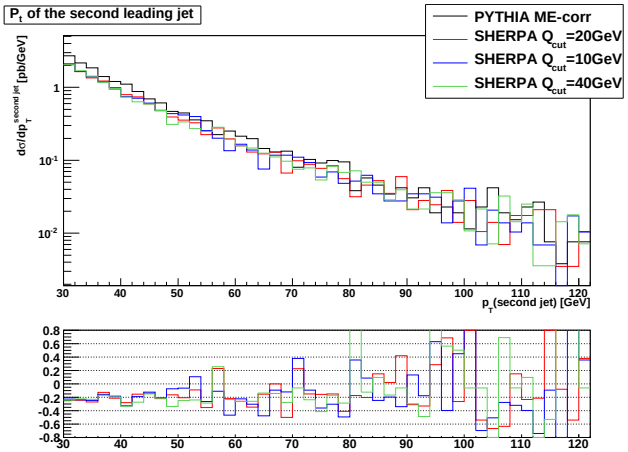
Figure 4.18: Differential jet rate for the transition $2 \to 1$ (a) and $3 \to 2$. Relative differences are calculated with respect to PYTHIA.

This is consistent with what is expected, since the second and third emission are governed by the PS alone both in PYTHIA and SHERPA.

Fig. 4.19 shows the p_T spectra for the hardest and the second hardest jet in PYTHIA and in SHERPA for the three different values used for q_{cut}. In this case, since we are interested in the amount of radiation we normalized plots to their respective cross section. The shape of the distributions is similar in all cases and also the total cross section for the emission of at least one jet is similar.



(a)



(b)

Figure 4.19: p_T spectra for the hardest (a) and second hardest (b) jet in PYTHIA and in SHERPA. The latter generator has been run with three different values of q_{cut}. Relative differences are calculated with respect to PYTHIA.

4.4 A comparison between AlpGen and PYTHIA

We made the same test with AlpGen, looking for differences with respect to ME corrected PYTHIA. We considered the contribution from up to one additional parton from the matrix element, and we used PYTHIA to shower the ME events generated by AlpGen.

4.4.1 Total cross section

Table 4.2 summarizes the total cross section for AlpGen with up to one additional parton from the ME. Results for three values of the ME cutoff for the generation of the additional jet are shown. The difference with respect to PYTHIA is 5% at most.

AlpGen		σ_i [pb]	Total σ [pb]
ME cutoff=10GeV	0 jet	1092	1609
	1 jet	517	
ME cutoff=20GeV	0 jet	1303	1594
	1 jet	291	
ME cutoff=40GeV	0 jet	1452	1558
	1 jet	106	
PYTHIA			Total σ [pb]
inclusive			1528

Table 4.2: Cross sections for AlpGen and PYTHIA.

4.4.2 Lepton observables

The p_T spectrum and the η distribution for PYTHIA (with and without ME corrections) and AlpGen are shown in Fig. 4.20. Concerning PYTHIA, the sample without ME corrections has the shower starting scale set to $\sqrt{s}$. For AlpGen we used a minimum p_T for the additional ME generated parton of 20 GeV, and the minimum p_T for the cone algorithm that steers the matching was set to 25 GeV, as recommended by the authors.

AlpGen lepton pair p_T spectrum appears to be softer than ME-corrected PYTHIA. This also translates into a broader η distribution for AlpGen. If we compare AlpGen to both ME corrected and to uncorrected PYTHIA, it appears that AlpGen follows uncorrected PYTHIA for low p_T values and then agrees with corrected PYTHIA in the high p_T tail.

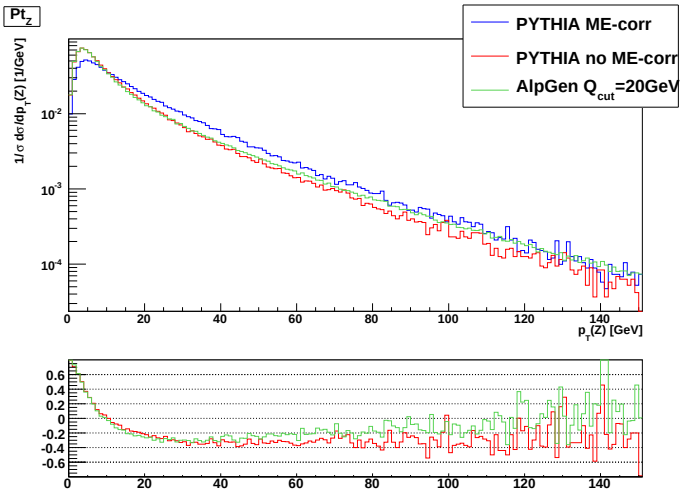
Also the p_T spectrum for the leptons from the Z boson decay (Fig. 4.21) shows the same behavior.

The dependency on the resolution cut that separates the ME from the PS region is very limited. The lepton pair p_T and η spectra and the p_T spectrum for the electron from Z are shown in Fig. 4.22 for three different choices of the Matrix Element cutoff in AlpGen (and correspondingly of the minimum p_T of the internal cone algorithm): q_{cut}=10, 20, 40 GeV.

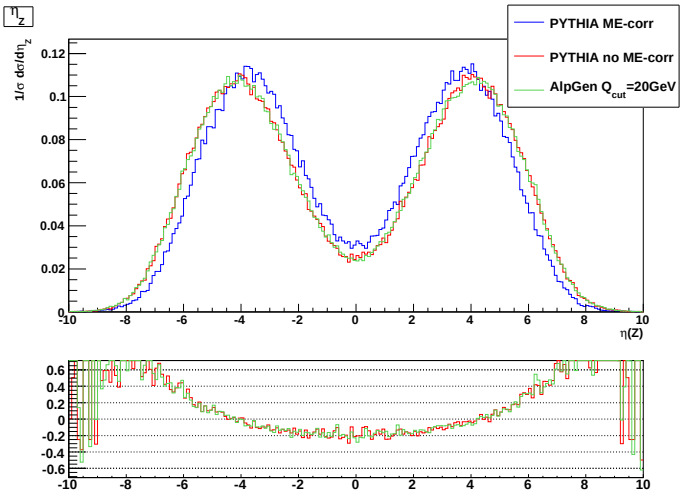
4.4.3 Jet observables

Differential jet rate plots appear to confirm the trend observed for letponic observables. Fig. 4.23 shows the $1 \to 0$ differential jet rate. Both ME corrected and uncorrected PYTHIA are shown as a reference. We see that AlpGen closely follows uncorrected PYTHIA in the low $Q_{1\to 0}$ region, then it starts to agree with corrected PYTHIA for high values of $Q_{1\to 0}$.

Fig. 4.24 shows the differential jet rates $1 \to 0$, $2 \to 1$ and $3 \to 2$ in AlpGen for three different values of the ME cutoff in AlpGen. Also ME-corrected PYTHIA is shown as a reference. The dependency on the cut is very limited, as already



(a)



(b)

Figure 4.20: p_T spectrum (a) and η distribution (b) for the lepton pair. `AlpGen` is compared with `PYTHIA`, with and without ME corrections. Relative difference with respect to ME corrected `PYTHIA` is shown for each plot.

observed for the leptonic observables.

4.5 Summary

We studied the effect of ME corrections in `PYTHIA` and `HERWIG`. Both programs can fully take into account ME corrections for one additional parton emission. The implementation is slightly different.

- `PYTHIA` modifies the shower in two steps: first the starting scale is raised so that any hard emission from the shower is kinematically possible; then the

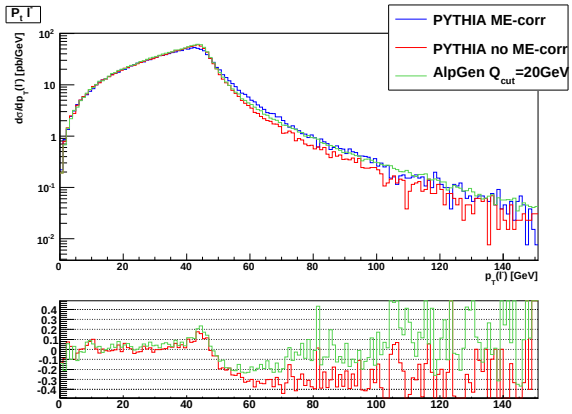
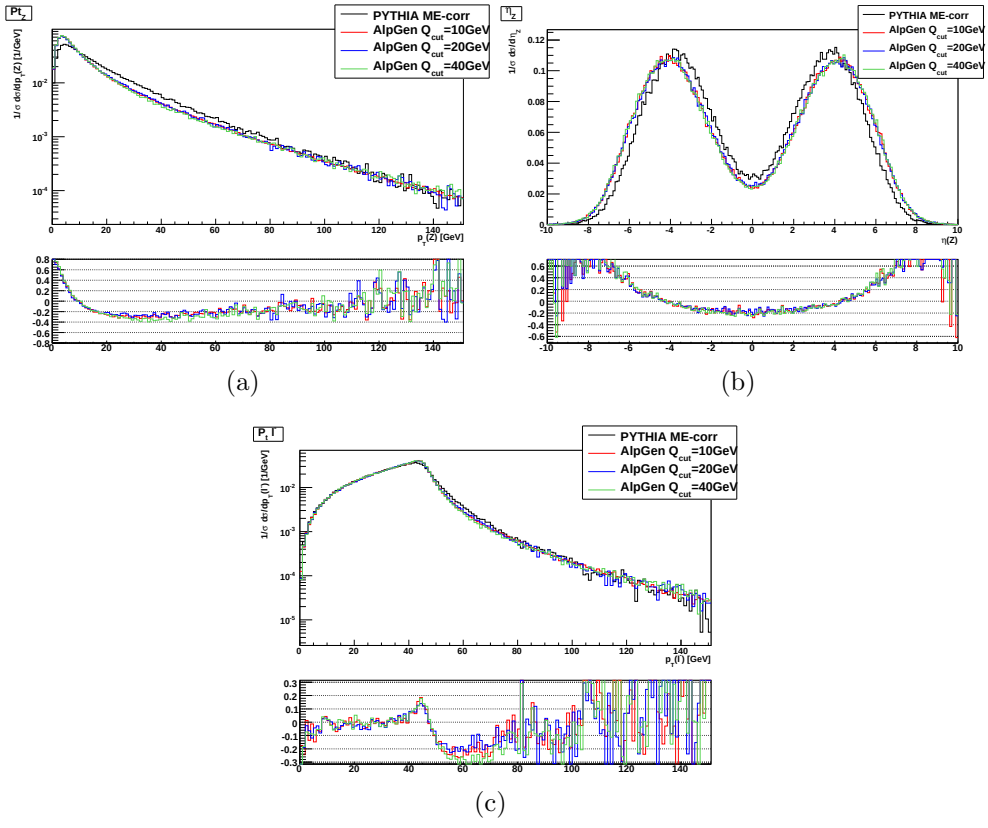


Figure 4.21: p_T spectrum of the electron from Z decay. `AlpGen` is compared with `PYTHIA`, with and without ME corrections. Relative differences are with respect to ME-corrected `PYTHIA`.



(a) (b)

(c)

Figure 4.22: (a) p_T spectrum and (b) η distribution for the lepton pair. (c) p_T spectrum for the electron. `AlpGen` has been run with three different values of the Matrix Element cutoff. Also ME-corrected `PYTHIA` is shown as a reference. Relative differences are calculated with respect to ME-corrected `PYTHIA`.

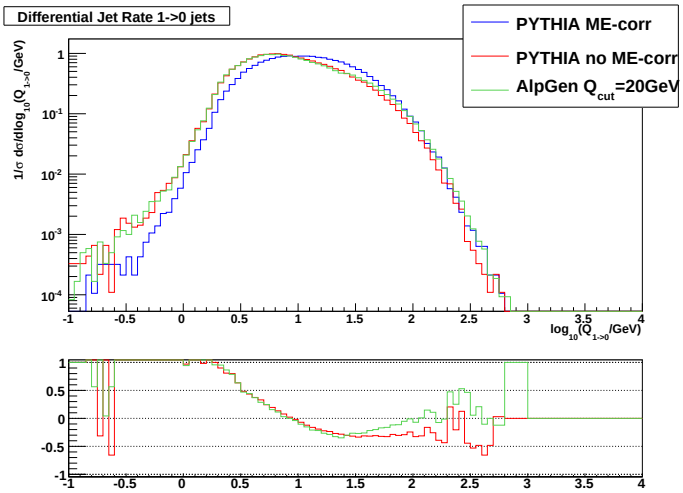


Figure 4.23: Differential jet rate for the $1 \to 0$ transition. `AlpGen` has been run with a ME cutoff of 20 GeV. Both ME corrected and uncorrected (starting shower scale set to $\sqrt{s}$) `PYTHIA` are shown as a reference. Relative differences are with respect to ME corrected `PYTHIA`.

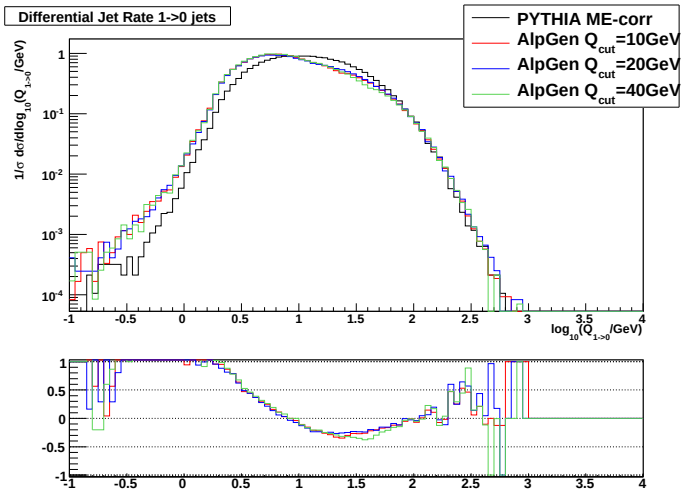
emission probability for the first emission is modified to include ME correction effect.

- In `HERWIG` the shower leaves the phase space for hard emission completely uncovered. For this reason the correction is performed in two steps: in the region already filled by the shower the same approach as `PYTHIA` is used. In the remaining part of the phase space the ME for one additional parton emission is directly used.

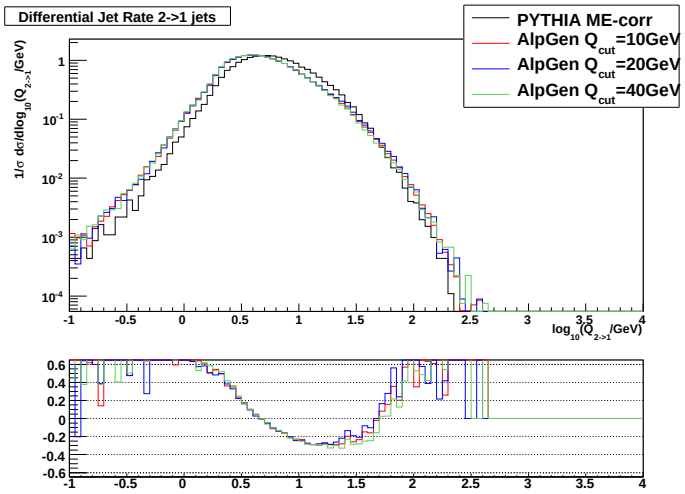
Both approaches give similar results. When ME corrections are switched off the Z p_T spectrum in `PYTHIA` changes also at low p_T, while the low p_T shape in `HERWIG` remains unchanged.

We used ME corrected `PYTHIA` as a test bed for `SHERPA` and `AlpGen`. When those programs are allowed to emit at most one parton from the ME calculation they should give results similar to `PYTHIA`. Actually `SHERPA` follows `PYTHIA` quite well, both on lepton and jet observables. On the other hand `AlpGen` appears to follow uncorrected `PYTHIA` at low p_T, for example it shows softer Z p_T spectrum with respect to ME corrected `PYTHIA`. This can be traced down to the fact that when `PYTHIA` is used to shower events produced by `AlpGen` ME corrections are switched off. This is done because `AlpGen` is going to introduce its own corrections. The side effect of this is that the low p_T shape of the Z p_T spectrum changes, and `AlpGen` cannot do anything in that region, which is entirely determined by the PS alone. `AlpGen` can only modify the high p_T tail of the distribution.

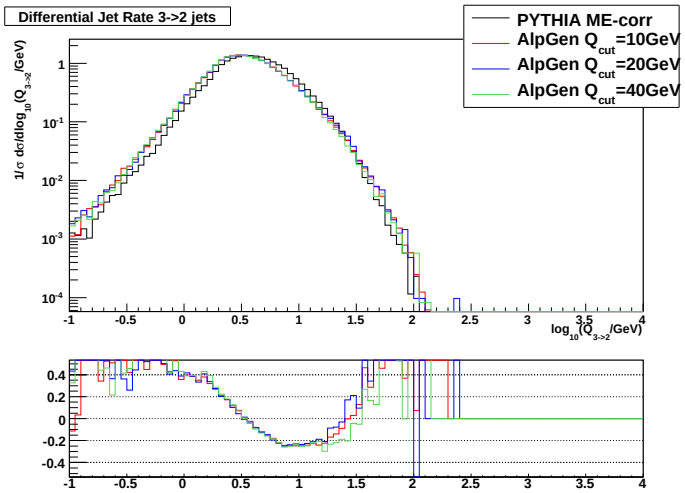
The dependency on the cut used to separate the ME and PS regions is limited both in `AlpGen` and in `SHERPA`.



(a)



(b)



(c)

Figure 4.24: Differential jet rates for the transitions $1 \to 0$ (a), $2 \to 1$ (b) and $3 \to 2$ (c). `AlpGen` has been run with three different values of the Matrix Element cutoff. Also ME-corrected `PYTHIA` is shown as a reference. Relative differences are with respect to ME corrected `PYTHIA`.

Chapter 5

AlpGen and SHERPA in $Z/\gamma^* + jets$

In this chapter we explore the features of AlpGen and SHERPA in events in which multiple emission from the ME is considered. A consistent matching of ME and PS in those events is a much more complicated task than the one described in the previous chapter.

We shall describe the behavior of the two generators when run with up to three additional partons emerging from the ME. Both generators were run with CTEQ6L [80] parton density functions. Results presented in this chapter are obtained running the full event generation chain, including the hadronization step, but still without the simulation of the Multiple Parton Interactions. We check at generator level how event observables for leptons and jets change as the matching parameters or the scale definition are changed. Then, we compare the two generators running them with the default parameter choice.

Finally we see how the two generators reproduce real data published in three recent papers from the Tevatron D0 collaboration.

5.1 Characterization of SHERPA

Results presented in this section, and in most of this chapter, were obtained with an analysis coded within the Rivet [72] Monte Carlo validation framework. The analysis has been run on the final state particles with $|\eta| < 5$. Jets were reconstructed with the longitudinally invariant $k_\perp$ algorithm as implemented in the FastJet [81] package, that is interfaced to Rivet. The D constant used in the $k_\perp$ algorithm was chosen to be 0.4. The minimum p_T for jets was set to 30 GeV. We required the presence of a lepton pair with invariant mass within 66 and 116 GeV.

In Fig. 5.1 we show the p_T (a) and η (b) distributions of the lepton pair in SHERPA, as obtained when running SHERPA with up to three partons emerging from the ME and with a ME-PS separation cut q_{cut}=20 GeV. The inclusive contribution is in black, while the contribution from the different jet multiplicities is shown in color. The high p_T tail of the distributions is built up with the contribution from

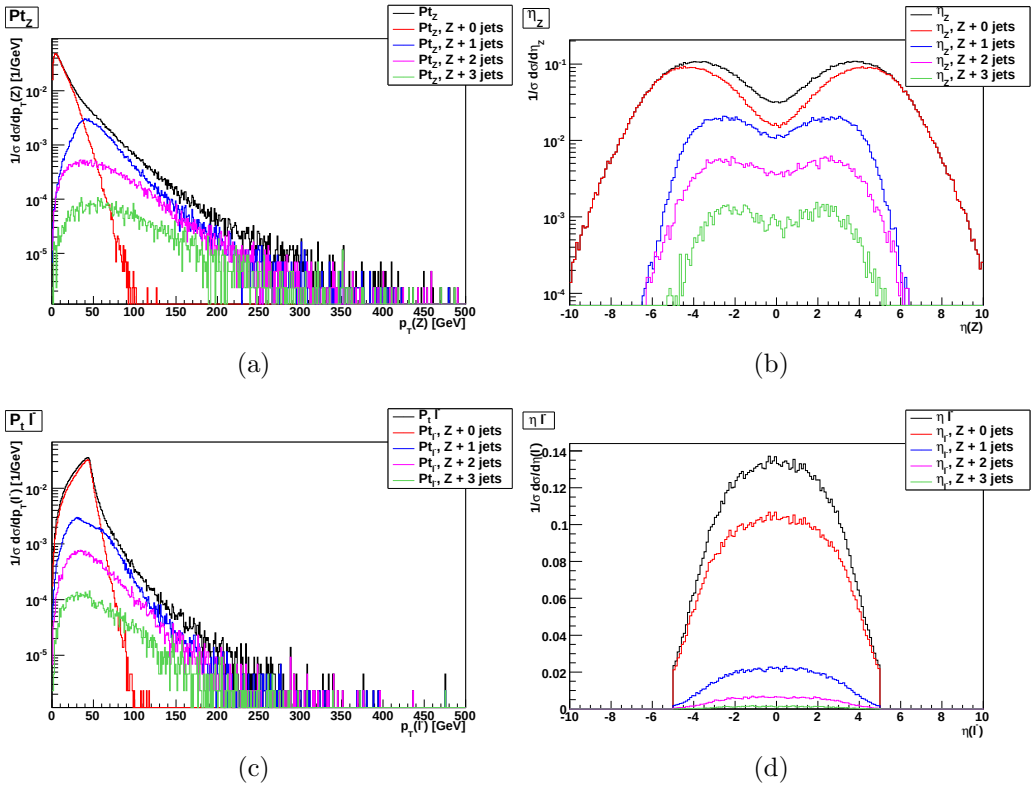


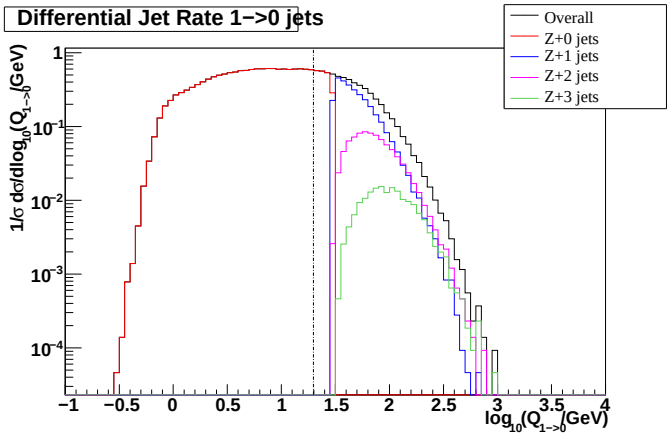
Figure 5.1: p_T (a) and η (b) distribution for the lepton pair, and p_T (c) and η (d) distributions for the negative lepton. Events were generated with SHERPA. The inclusive contribution is in black, while the contribution from the different jet multiplicities is in color.

multi-jet emission. This translates into the η distribution, where it is shown that the lepton pair is more central in case of multiple jet emission.

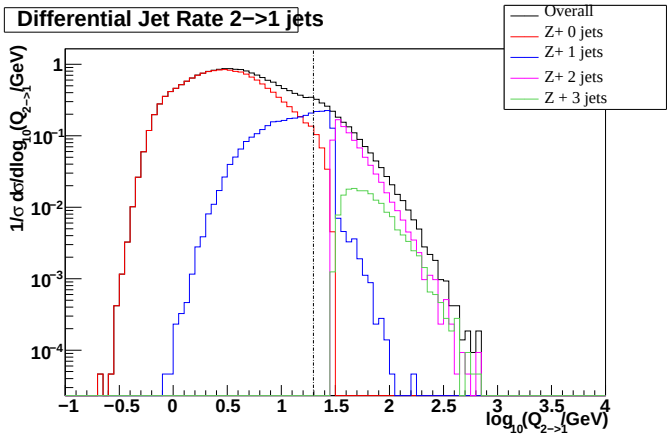
The p_T and η distributions for the negatively charged lepton are shown in Fig. 5.1. Also in this case we observe that there is contribution to the high p_T tail from multiple jet emission.

Differential jet rates are shown in Fig. 5.2 for the transitions $1 \to 0$ (a), $2 \to 1$ (b), $3 \to 2$ (c). In each plot a vertical line signals the position of q_{cut}. The contribution from different jet multiplicities is put into evidence in color. As already noticed in the previous chapter differential jet rates provide us with a very detailed description of how the phase space is filled. The distribution for the differential jet rate $n \to n-1$ is filled by the n parton ME in the region above q_{cut}, and by the PS plus the ME up to $n-1$ partons in the region below q_{cut}. This makes the differential jet rate a very informative diagnostic tool for this matching prescription.

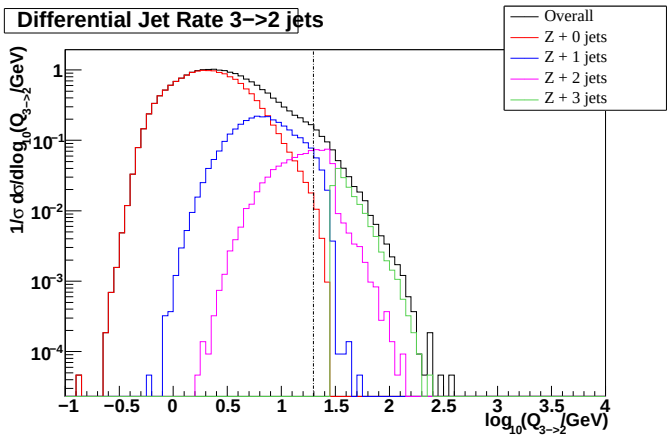
We notice that while the $1 \to 0$ distribution shows a very smooth transition around q_{cut}, little bumps are observed in this region in the $2 \to 1$ and $3 \to 2$ plots. This is expected to some extent, because the matching prescription is not exact.



(a)



(b)



(c)

Figure 5.2: Differential jet rates in SHERPA. (a) $1 \to 0$, (b) $2 \to 1$, (c) $3 \to 2$. The vertical dashed line signals the position of q_{cut}.

This effect is probably due to configuration mismatch that can occur in the vicinities of the cut: parton configurations that are below the cut as they emerge from the ME can migrate above due to the way the PS alters the ME kinematics. The effect is anyway moderate.

Jet multiplicity is shown in Fig. 5.3. The leading (a) and second (b) jet p_T distributions are shown in Fig. 5.4. Also in this case we see that the tail of the p_T distributions is filled by the multiple jet emission.

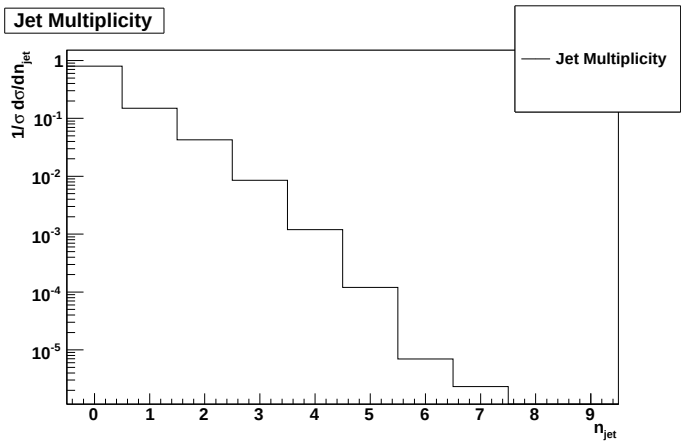
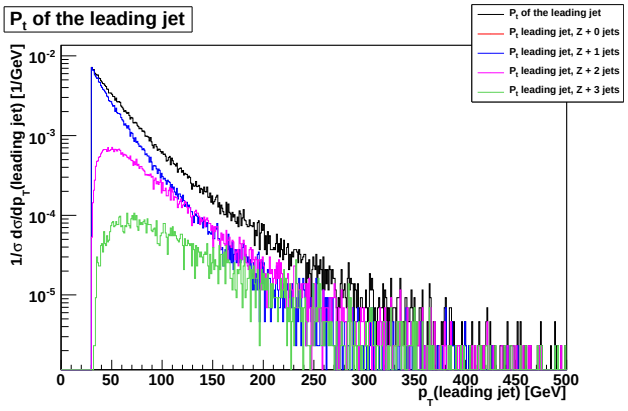


Figure 5.3: Jet multiplicity in SHERPA.

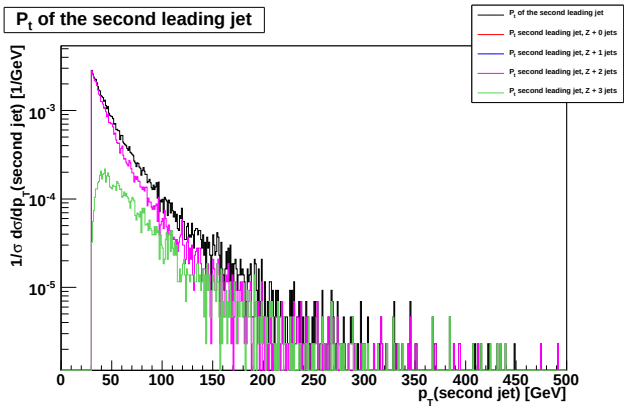
In order to quantify the effect of the inclusion of multiple parton emission from the ME we made some comparisons with ME corrected PYTHIA. The p_T distribution for the lepton pair (a) and for the negative lepton (b) in PYTHIA and SHERPA are shown in Fig. 5.5. In both cases the distribution for SHERPA is harder at high p_T. We also notice that the low p_T shape in the Z transverse momentum distribution is unaffected, as expected.

The origin of these harder spectra can be identified looking at the differential jet rates in Fig. 5.6. SHERPA fills the region for high Q significantly more than PYTHIA. Consequently, the rate of events with at least one jet is higher in SHERPA, and it shows a harder spectrum, as shown in Fig. 5.7.

Another interesting feature is the distribution of the difference in the azimuthal angle between the leading and the second jet, shown in Fig. 5.8. SHERPA shows a more evident tendency in producing back to back jets. A possible explanation is that the 2-jet ME is present in SHERPA but not in PYTHIA.

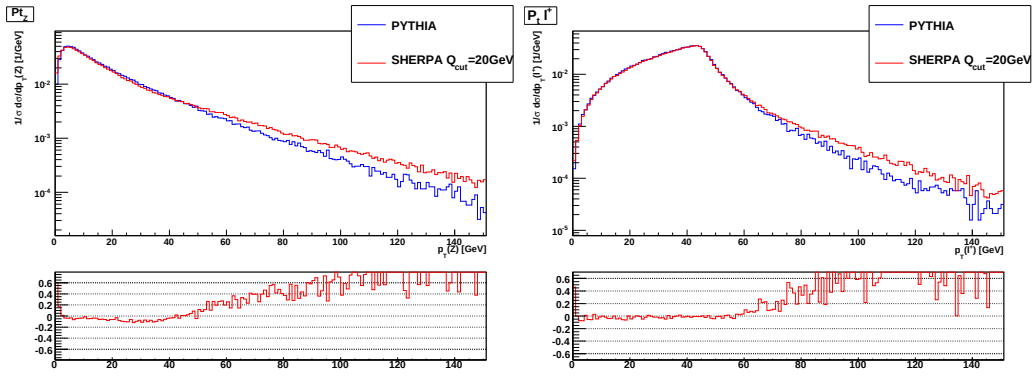


(a)



(b)

Figure 5.4: (a) Leading jet p_T, (b) second jet p_T in SHERPA.



(a) (b)

Figure 5.5: p_T distribution for the lepton pair (a) and the lepton (b) in SHERPA and PYTHIA.

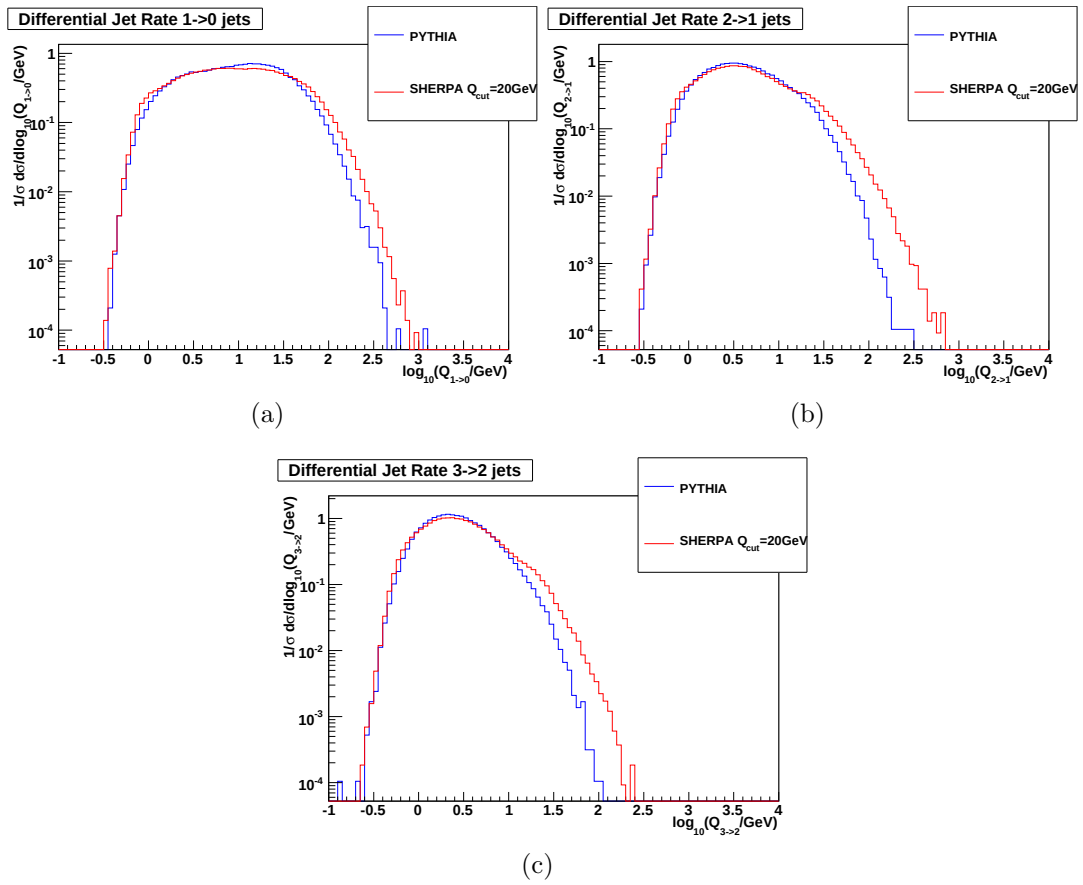


Figure 5.6: Differential jet rates in SHERPA and PYTHIA.

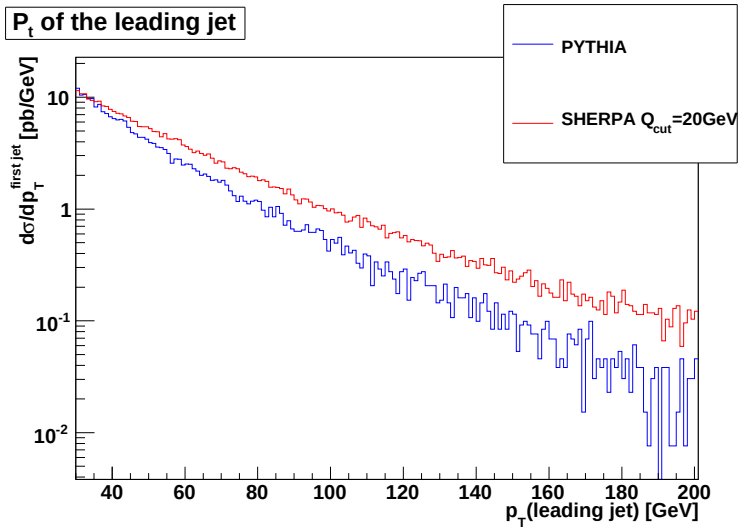


Figure 5.7: p_T spectrum for the leading jet in SHERPA and PYTHIA.

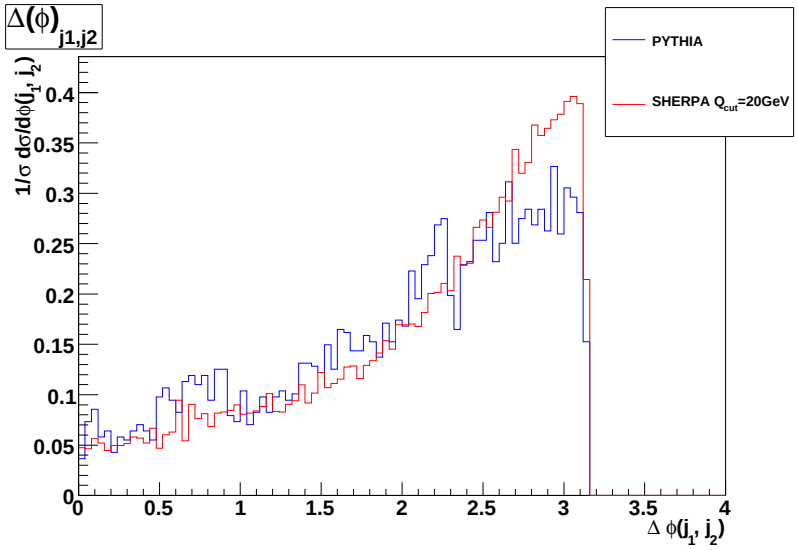


Figure 5.8: $\Delta\phi$ distribution between the two leading jets.

5.1.1 Changing the matching parameter in SHERPA

As already noticed in the previous chapter it's very important to evaluate the dependency of observables on the matching parameter q_{cut}. This parameter is unphysical, so we shall check that there is little dependency of event observables on it. We tried three different values of q_{cut}: 20 GeV, 30 GeV, 50 GeV. In the following we shall see how lepton and jet observables change as q_{cut} changes.

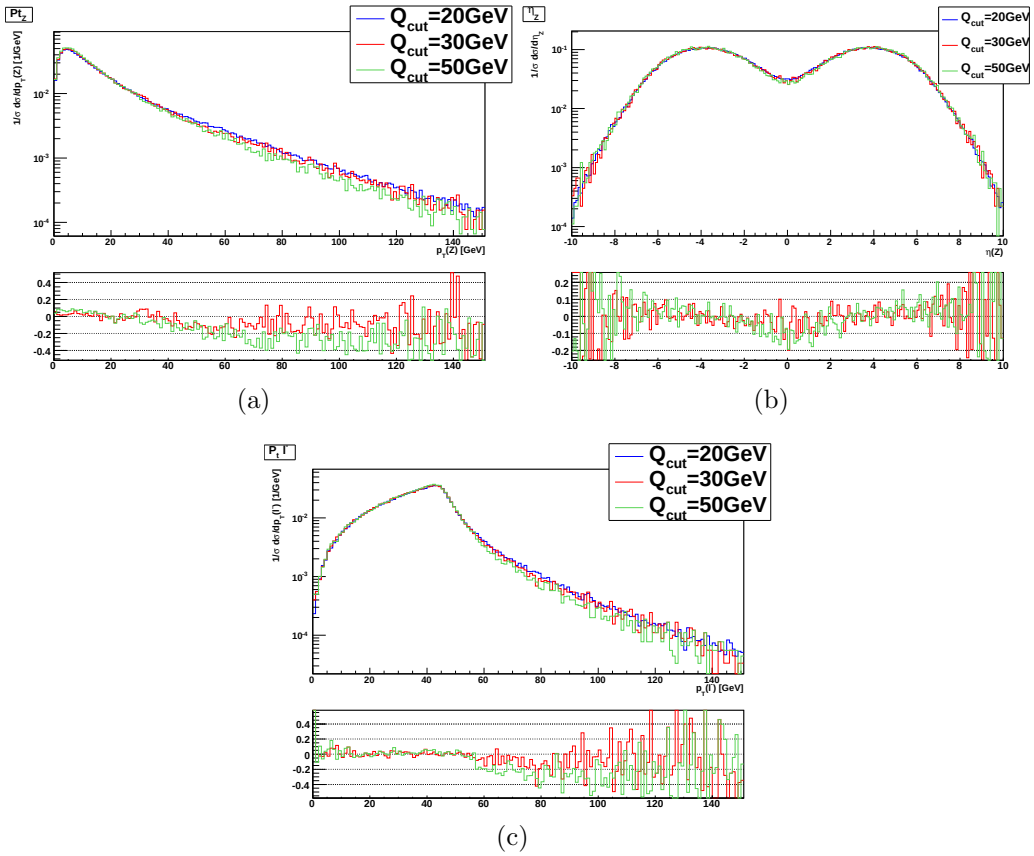
Table 5.1 summarizes the total cross section values for the three above mentioned choices of q_{cut}. The difference between the maximum and minimum value is about 15%. Fig. 5.9 shows the p_T and η distribution of the lepton pair. The effect on the shape of these distribution is moderate. We notice however that the spectra tend to be softer as q_{cut} is increased. This is due to the fact that raising q_{cut} means to reduce the phase space available for the ME. Since the ME is responsible for the hardest parton kinematics, softer spectra appear.

Differential jet rates in SHERPA are shown in Fig. 5.10 for three different choices of q_{cut}. The position of the cut is signaled with a vertical line. Relative differences with respect to the q_{cut}=20 GeV curve are shown. The relative difference is up to 40% in the region around the cut.

The jet multiplicity (a) and the leading jet p_T spectrum (b) are shown in

SHERPA	Total σ [pb]
q_{cut}=20GeV	1594
q_{cut}=30GeV	1411
q_{cut}=50GeV	1400

Table 5.1: Cross sections for SHERPA.



(a) (b)

(c)

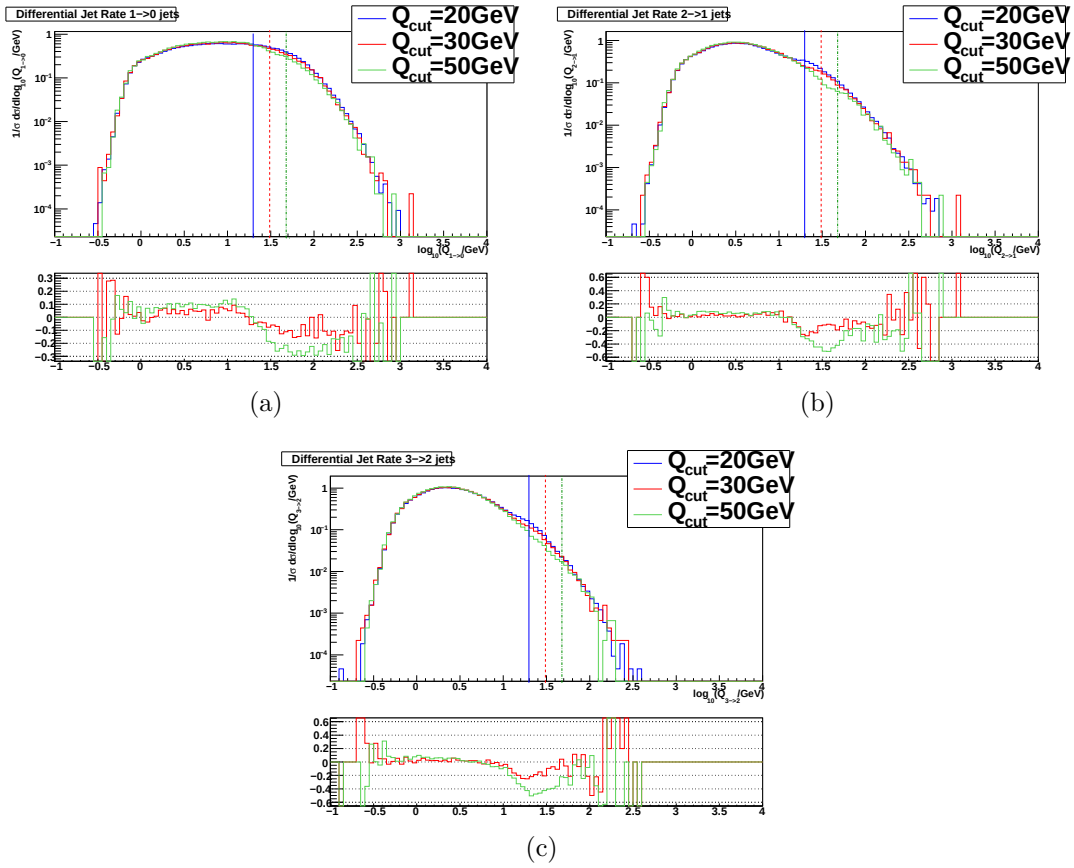
Figure 5.9: (a) p_T and (b) η spectrum for the lepton pair and (c) p_T distribution for the electron with three different choices of q_{cut}. Relative differences with respect to the q_{cut}=20 GeV curve are shown.

Fig. 5.11. The mean jet multiplicity is highest for the q_{cut}=20 GeV sample and lowest for the q_{cut}=50 GeV sample. This in in agreement with what we already observed on the lepton spectra: softer lepton spectra are accompanied by lower mean jet multiplicity. The shape of the p_T spectrum is almost unchanged.

5.1.2 Changing the scales in SHERPA

Another source of systematic uncertainties is the scale definition. In CKKW the scale at which the strong coupling constant is evaluated is $q_i = y_i \cdot \hat{s}$, where y_i identifies a particular clustering step i in the $k_\perp$ clustering procedure that is used in the CKKW matching prescription and $\hat{s}$ is the scale of the $2 \to 2$ event, as described in Sec. 3.4.2.

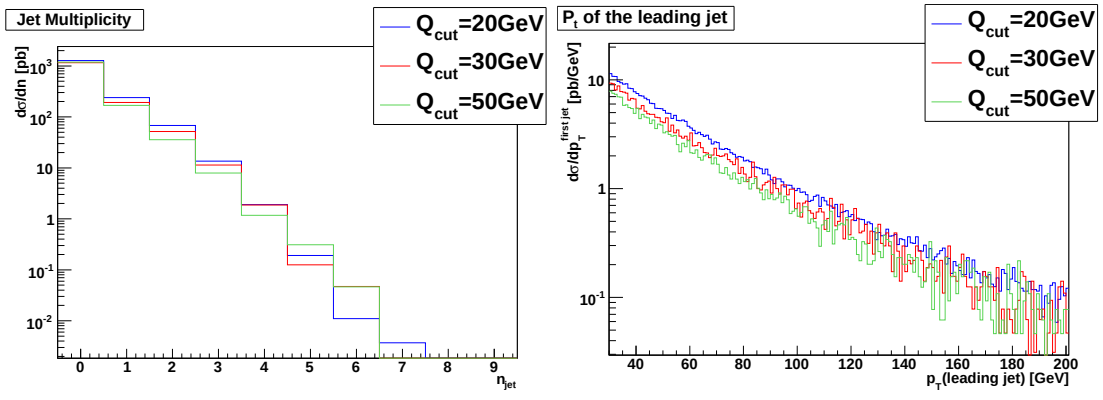
Since the scale is an unphysical parameter the dependency on the scale for the reconstructed quantities should be limited. In SHERPA it is possible to change the scale at which the coupling constant is evaluated by a factor. To evaluate to what



(a) (b)

(c)

Figure 5.10: Differential jet rates for different choices of q_{cut} in SHERPA. (a) $1 \to 0$, (b) $2 \to 1$, (c) $3 \to 2$. Relative differences with respect to the q_{cut}=20 GeV curve are shown.



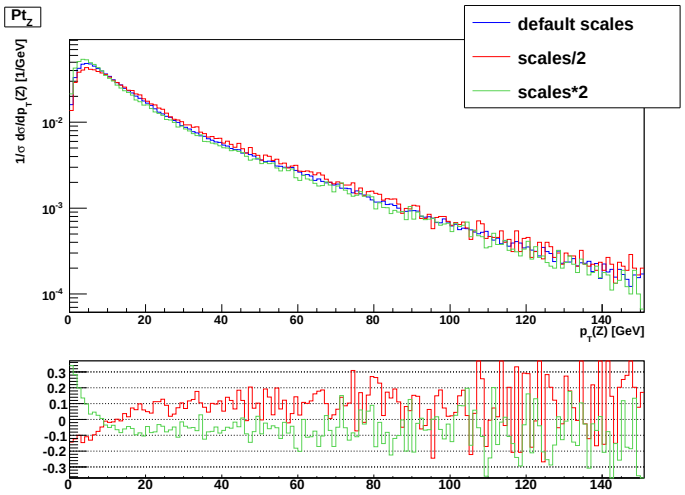
(a) (b)

Figure 5.11: (a) Jet multiplicity and (b) leading jet p_T spectrum in SHERPA for three different choices of q_{cut}.

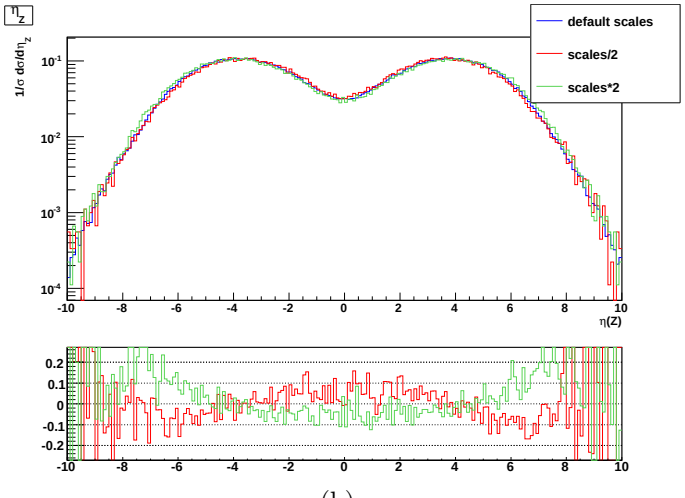
SHERPA	Total σ [pb]
default scales	1594
scales $\div$ 2	1292
scales $\times$ 2	1646

Table 5.2: Cross sections for SHERPA with different scale choices.

extent observables depend on the scale choice we generated samples in which the scale at which the coupling constant is evaluated is multiplied by two or divided by two. The choice of a factor of two is rather arbitrary, but it is quite common when evaluating uncertainties due to the scale definition.



(a)



(b)

Figure 5.12: p_T (a) and η distribution for the lepton pair in SHERPA for three different scale choices. Relative differences with respect to the default scales are shown.

The total cross section for the three the different scale definitions is summarized in Table 5.2. The difference between the two extremes is nearly 30%.

The scale dependency effect in the p_T and η distribution for the lepton pair is shown in Fig. 5.12. The lepton pair p_T distribution shows that most of the difference between the three samples is concentrated in the first bins, where the 0 jet contribution dominates. This is in agreement with the prediction from [82], where it is shown that the inclusive LO cross section decreases with scales, while the LO cross section for the emission of at least one jet is much less sensitive to scale variations. Thus, since the biggest contribution to the inclusive cross section comes from the 0 jet contribution the effect is mostly visible in the low p_T region in the lepton pair p_T spectrum.

The effect translates into the η distribution, where the sample with scales divided by two shows the most central lepton pair.

5.2 Characterization of AlpGen

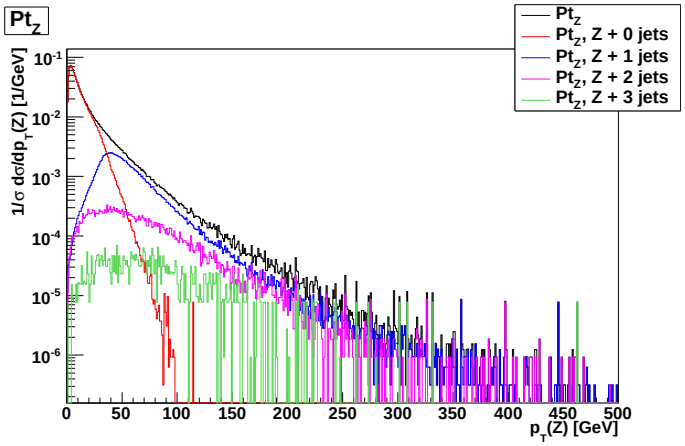
In this section we describe events produced with AlpGen. We produced up to three partons from the ME, with a minimum p_T cut of 20 GeV. Events were showered with PYTHIA, using the MLM matching prescription. The minimum E_t for the cone algorithm that is used in AlpGen to steer the matching was set to 25 GeV, as recommended by the AlpGen authors for this kind of events. Exactly the same analysis that was run on the SHERPA sample was run on AlpGen to produce plots shown in this section.

Distributions for the p_T and η of the lepton pair in events produced with AlpGen are shown in Fig. 5.13. Also in this case we observe how the multi-jet emission sums up to produce the high p_T tail.

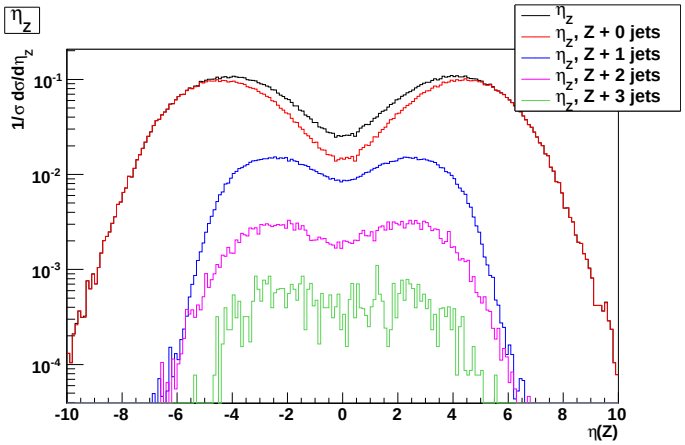
Differential jet rates in AlpGen are shown in Fig. 5.14. For AlpGen the measure used to separate the ME and the PS filled regions of the phase space is not $k_\perp$, hence the interpretation of the differential jet rates is more complex. In Fig. 5.14 the overall contribution is in black, while in color the contribution from different jet multiplicities is shown. Even if we cannot identify a boundary between the ME and the PS filled regions, the behavior of the curve around the minimum p_T cut used in the AlpGen cone algorithm (signaled with a vertical line in the plots) is expected to be the region where possible matching effects show up. For all the three plots the transition is rather smooth in this region.

5.2.1 Changing the matching parameter in AlpGen

As for SHERPA we modified the parameters of the matching to see how the total cross section and the shape of the distributions change. The total cross section in AlpGen for two different values of the p_T cut of the cone algorithm used for the MLM matching is summarized in Table 5.3. Practically no difference is observed.



(a)



(b)

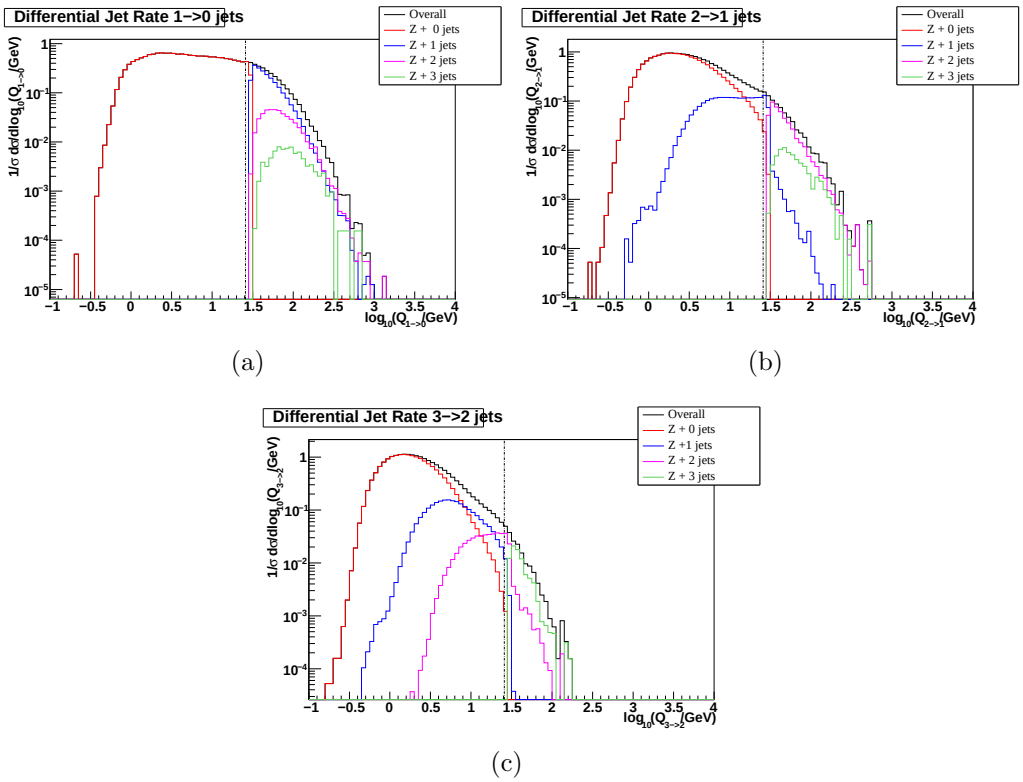
Figure 5.13: (a) p_T distribution and (b) η distribution for the lepton pair in AlpGen.

AlpGen	Total σ [pb]
p_T^{min}=25GeV	1534
p_T^{min}=40GeV	1516

Table 5.3: Cross sections for AlpGen with different choices for the p_T^{min} in the internal cone algorithm.

The lepton observables distributions for the two choices of p_T^{min} are shown in Fig. 5.15. The effect is very moderate in all the distributions.

For what concerns jet observables, differential jet rates are shown in Fig. 5.16 for the two choices of p_T^{min}. The relative difference plot shows the difference with respect to the p_T^{min}=25 GeV sample. The shape of these distribution is mildly affected in the region around the p_T^{min} cutoff.



(a) (b)

(c)

Figure 5.14: Differential jet rates in AlpGen.

5.2.2 Changing the scales in AlpGen

As for SHERPA we tried to change the scale where the strong coupling is evaluated. We used a factor two and one half as in SHERPA. The effect on the total cross section predicted by AlpGen is summarized in Table 5.4. The effect is of the same order as in SHERPA.

The effect of a change in the scale choice on the p_T and η spectrum for the lepton pair is shown in Fig. 5.17 . In this case the low p_T shape is unaffected, while there are not negligible differences in the high p_T tail.

AlpGen	Total σ [pb]
default scales	1534
scales $\div$ 2	1449
scales $\times$ 2	1672

Table 5.4: Cross sections for AlpGen with different scale choices.

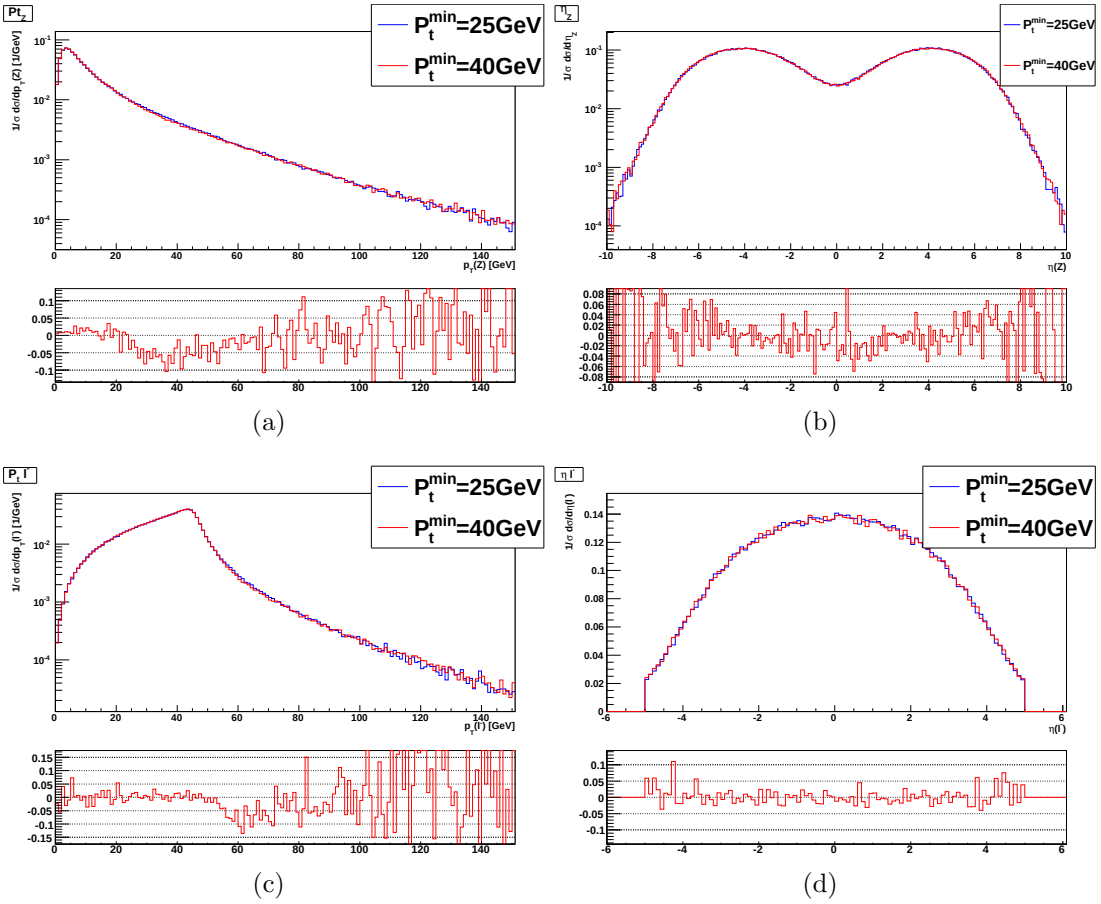


Figure 5.15: (a) p_T distribution and (b) η distribution of the lepton pair; (c) p_T and (d) η distribution of the negatively charged lepton in AlpGen. Plots are shown for two different values of the p_T^{min} used in the MLM cone algorithm. The relative difference is calculated with respect to the sample with p_T^{min}=25 GeV.

5.3 Comparison between AlpGen and SHERPA

We compared AlpGen and SHERPA; both generators were run with the default parameter choice (default scale choice and default matching parameters, i.e. q_{cut}=20 GeV in SHERPA and minimum p_T for the cone algorithm used in AlpGen set to 25 GeV); the AlpGen sample has been showered both with PYTHIA and with HERWIG.

The transverse momentum distribution and the η distribution for the lepton pair are shown in Fig. 5.18. The shape of the distribution for the transverse momentum shows not negligible differences; clearly SHERPA has the hardest spectrum, while AlpGen+PYTHIA shows the softest. We also notice that the low p_T shape of AlpGen+HERWIG is similar to SHERPA (the relative difference between AlpGen+HERWIG and SHERPA is flat as $p_T \to 0$), while the shape for AlpGen+PYTHIA is different all over the p_T spectrum. We believe that the different low p_T behavior

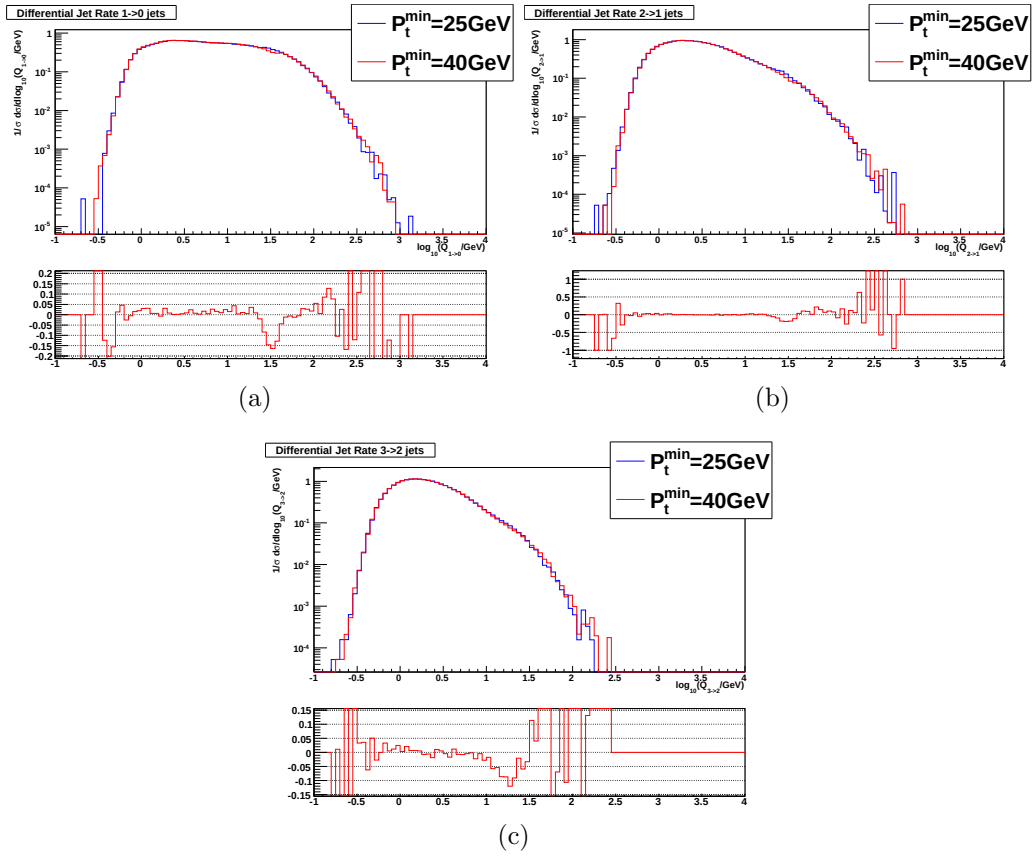


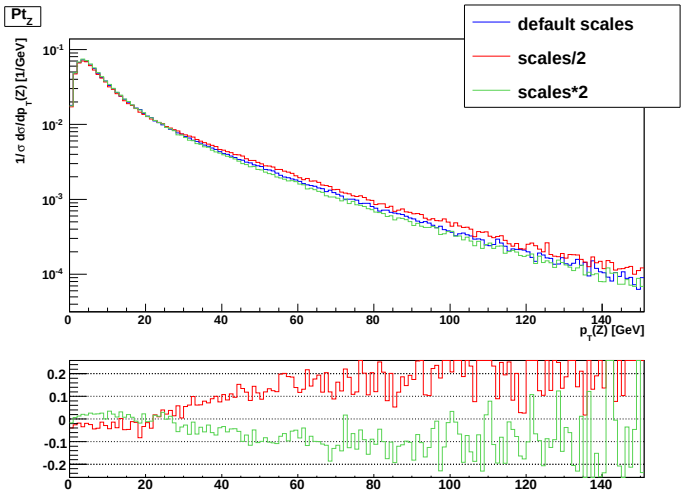
Figure 5.16: Differential jet rates in AlpGen for two choices of p_T^{min}: 25 GeV and 40 GeV; (a) $1 \to 0$, (b) $2 \to 1$, (c) $3 \to 2$. The relative difference plot shows difference with respect to the p_T^{min}=25 GeV sample.

in AlpGen+HERWIG and AlpGen+PYTHIA is connected to the different effect that the deactivation of ME corrections has on HERWIG and PYTHIA, as already discussed in Secs. 4.1.1 and 4.2. Differences are also present in the η distribution, where SHERPA shows the most central Z.

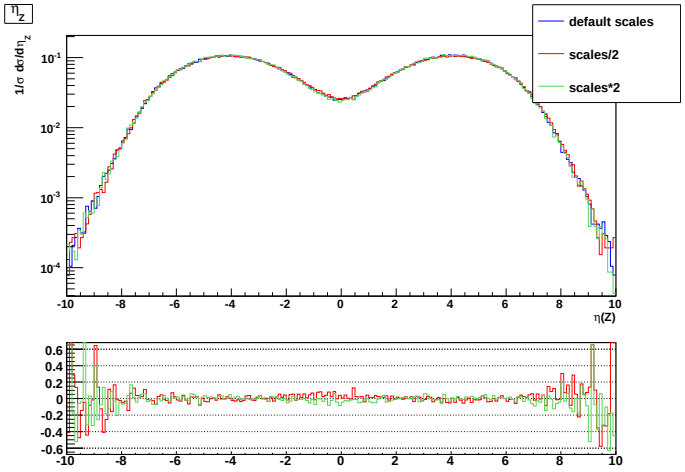
Both plots are consistent with a picture in which SHERPA is responsible for the emission of more radiation than AlpGen. In fact, the more QCD radiation is emitted the highest the Z boson transverse momentum is, because the Z recoils against the QCD radiation.

The same picture is confirmed also from the spectra of the individual leptons: the p_T spectrum for the lepton from Z is shown in Fig. 5.19 (a); also in this case the spectrum for SHERPA is the hardest. The η spectrum for the lepton from Z is shown Fig. 5.19 (b); in this case no big differences are observed.

Let's now have a look at the jet observables. The jet multiplicity for the three generators is shown in Fig. 5.20. SHERPA has the highest mean jet multiplicity, while AlpGen+PYTHIA the lowest. Nevertheless AlpGen samples show a somewhat longer



(a)



(b)

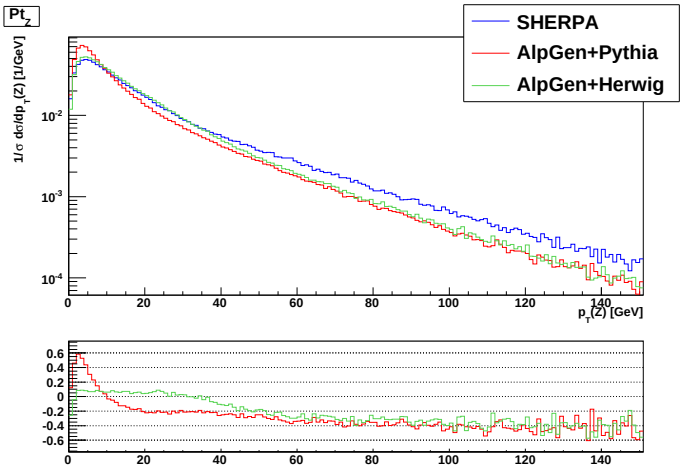
Figure 5.17: p_T (a) and η (b) distribution for the lepton pair with two different scale choices.

tail for high multiplicity.

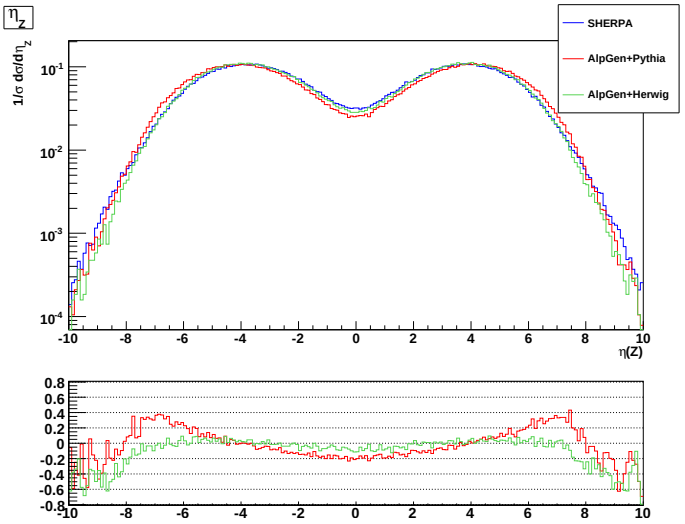
Fig. 5.21 shows the p_T spectrum for the first (a) and second (b) jet. Both plots are normalised to the total cross section for at least one (a) or two (b) jets. SHERPA shows a higher cross section in both plots. For what concerns the leading jet SHERPA also shows a slightly harder spectrum.

Fig. 5.22 shows angular correlations between the two leading jets: (a) is the $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$, (b) is $\Delta\phi$ and (c) is $\Delta\eta$. These observables do not show big differences.

This comparison shows that when running AlpGen and SHERPA with the default parameter choice some not negligible differences arise. While the total cross section is rather similar, it looks like SHERPA has more hard radiation than AlpGen. This is



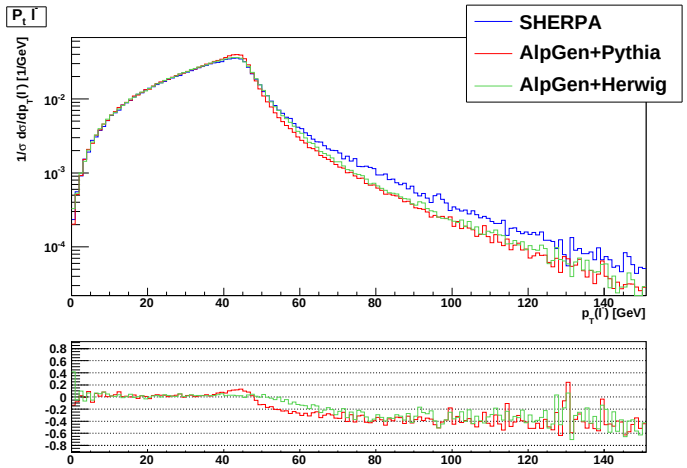
(a)



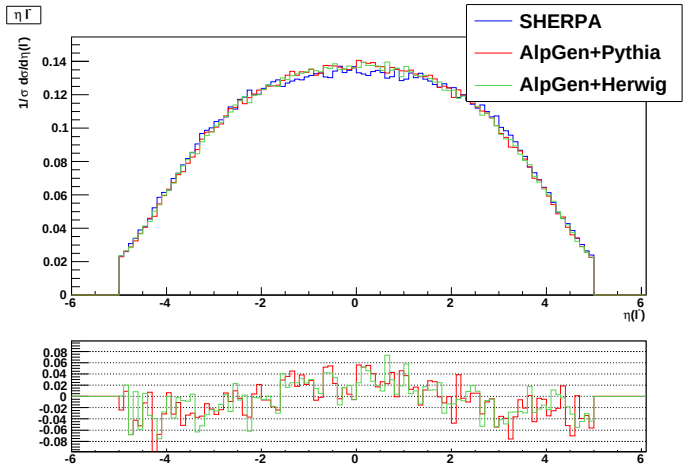
(b)

Figure 5.18: (a) p_T distribution for the lepton pair; (b) η distribution for the lepton pair. Events are produced with SHERPA, AlpGen+PYTHIA and AlpGen+HERWIG. The relative difference is with respect to SHERPA.

confirmed by the mean jet multiplicity and by the jet p_T spectra, and has an effect on the Z and lepton spectra.



(a)



(b)

Figure 5.19: (a) p_T distribution for the lepton; (b) η distribution for the lepton. Events are produced with SHERPA, AlpGen+PYTHIAand AlpGen+HERWIG. The relative difference is with respect to SHERPA.

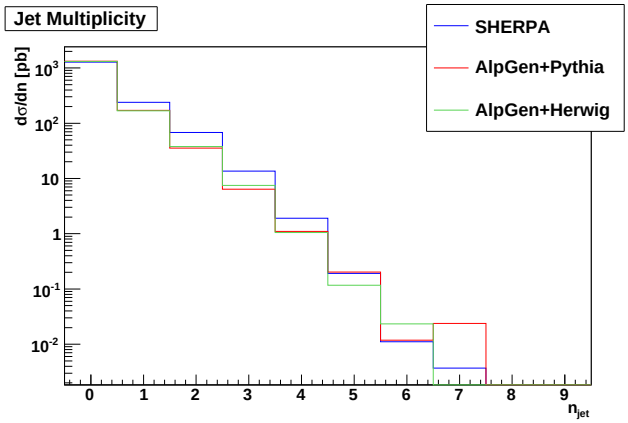
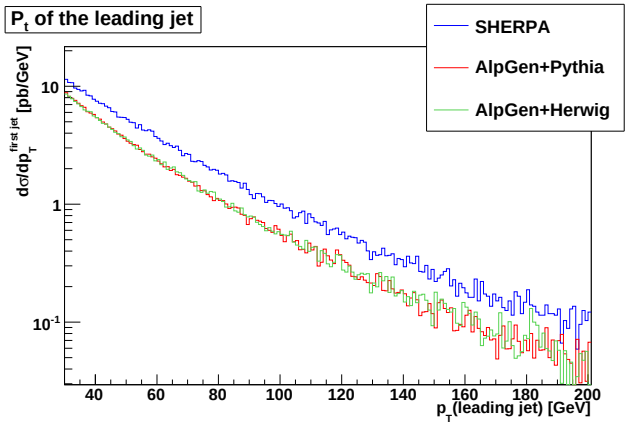
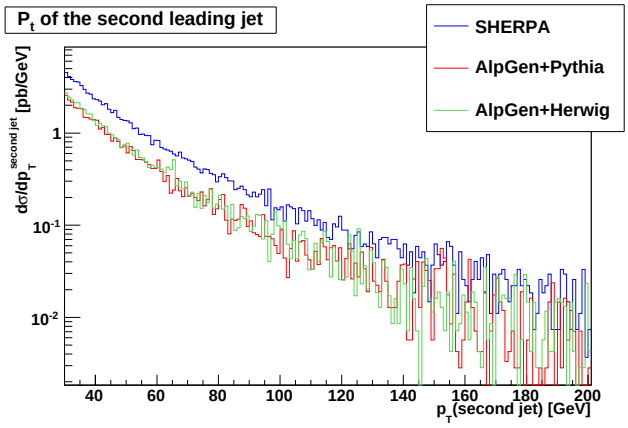


Figure 5.20: Jet multiplicity in SHERPA, AlpGen+PYTHIA, AlpGen+HERWIG.

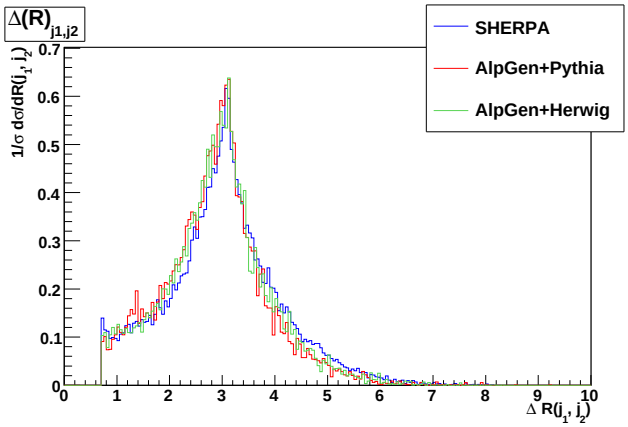


(a)

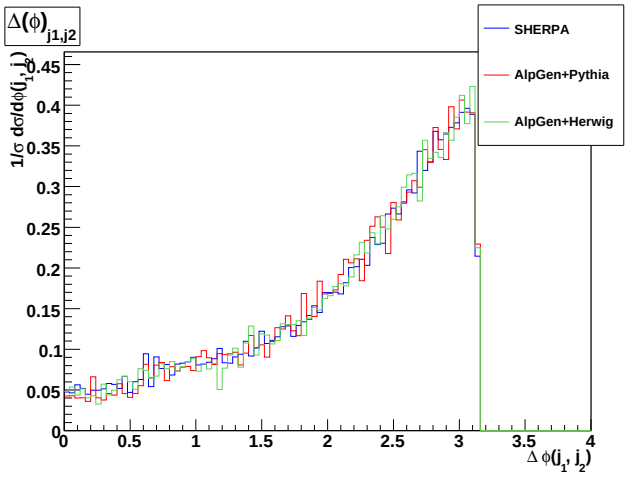


(b)

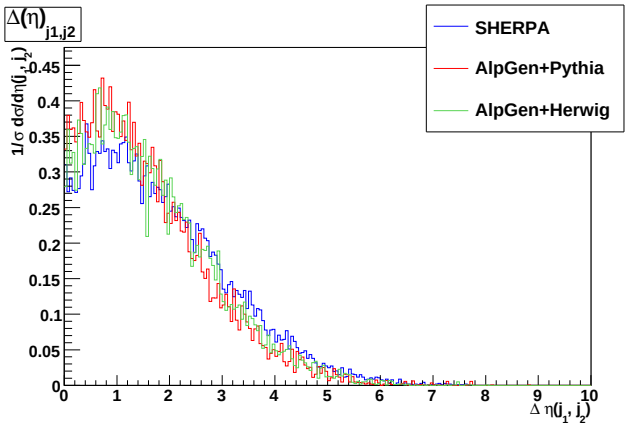
Figure 5.21: (a) Leading jet and (b) second jet p_T distributions in SHERPA, AlpGen+PYTHIA and AlpGen+HERWIG.



(a)



(b)



(c)

Figure 5.22: Angular correlations between the two leading jets: (a) $\Delta R{=}\sqrt{\Delta\phi^2 + \Delta\eta^2}$, (b) $\Delta\phi$, (c) $\Delta\eta$.

5.4 Comparison to Tevatron data

In this section we compare SHERPA, AlpGen and PYTHIA with data collected at the Tevatron $p\bar{p}$ collider [73]. We used the Rivet analysis software for this comparison. While Rivet can be used as an analysis framework for general studies on Monte Carlo event generators, its main feature is that it comes as a very straightforward tool to make comparisons to data.

Rivet implements a variety of algorithms that are used in experimental analyses; for example it implements all the jet finding algorithms that are used in Tevatron analyses. If one wants to see how an event generator reproduces results published in an experimental analysis, he just needs to implement in Rivet a clone of the experimental analysis, so that it can be run at generator level. Experimental data need to be corrected for detector effects. Experimental data the user wants to make comparisons to are stored in a data base called HEP-DATA [83]; for every published analysis it is possible to request the inclusion of results and plots in the HEP-DATA database. Once plots are in this database it is possible to export them in a low weight histogramming framework called AIDA [84]; Rivet output is AIDA histograms as well. Rivet implements a histogramming feature, called "autobooking" that ensures that the histograms created have the same binning as the original plots from the published paper.

5.4.1 Measurement of Z boson p_T measurement at Tevatron Run I by D0 collaboration

We made comparisons with data published by the Tevatron D0 collaboration in [85], regarding the measurement of the differential production cross section of Z boson as a function of Z's transverse momentum, at $\sqrt{s}$=1.8 TeV. Data were collected between 1994 and 1996, with an integrated luminosity of about 110 pb^{-1}.

We produced simulated samples with PYTHIA, SHERPA and AlpGen (using PYTHIA for Parton Shower and hadronization). The comparison among all the simulates sampled and D0 data is shown in Fig. 5.23. The total cross section for all simulated samples has been normalized to the measured cross section, so the comparison concerns the shape of the distribution and not the overall normalization.

We notice that PYTHIA underestimates the high p_T tail of the distribution. This is due to the fact that only the first emission in PYTHIA is corrected for ME effect, while the other emissions come from pure parton shower. Since the high p_T tail of the distribution is dominated by the emission of at least one additional parton from the ME, the fact that PYTHIA sits below the data is reasonably explained.

On the other hand, SHERPA shows a harder spectrum in the high p_T region, and seems to follow experimental data quite well.

Also AlpGen seems to do a better job than PYTHIA, even if it appears to underestimate the last bin of the distribution.

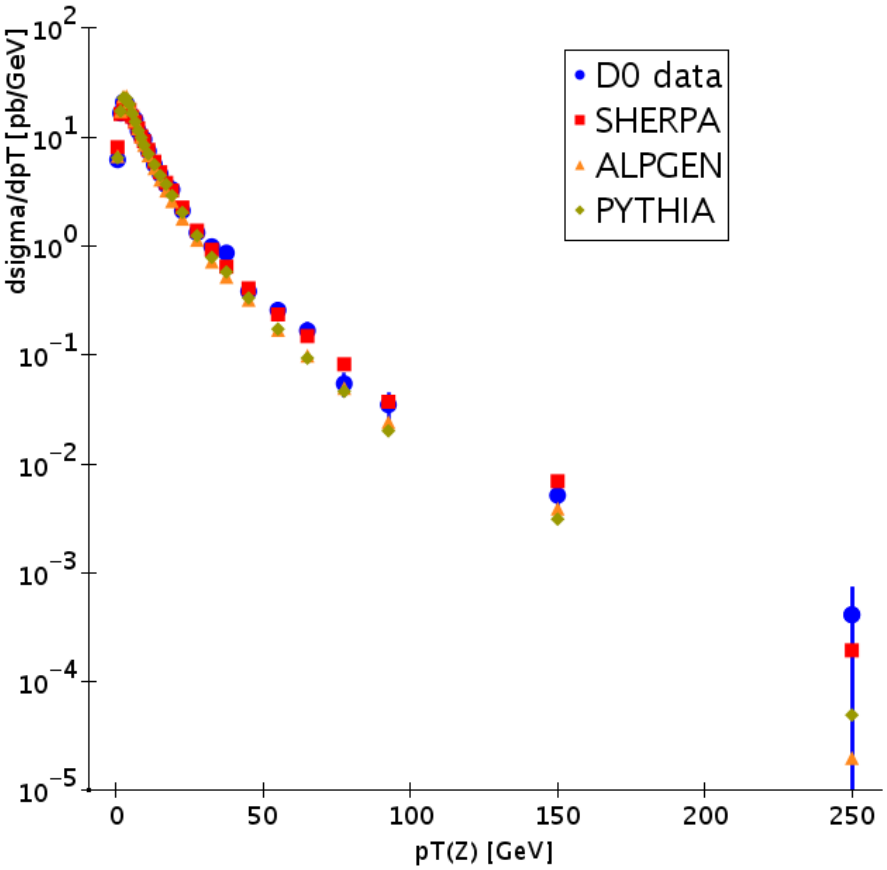


Figure 5.23: p_T distribution for the Z boson at Tevatron ($\sqrt{s}$=1.8 TeV). Results obtained from different MC generators are shown. The total cross section for the MC generators is normalized to the measured cross section.

5.4.2 Measurement of $\sigma(Z/\gamma^* + \geq njet)/\sigma(Z/\gamma^* + X)$ at Tevatron Run II by the D0 collaboration

We made a comparison with a recent measurement of the D0 collaboration on the ratio between the production cross section for $Z/\gamma^* + \geq n\ jet$ and the inclusive Z/γ^* production cross section. Data collected with the D0 detector correspond to an integrated luminosity of 400 pb^{-1}, and were collected with $\sqrt{s}$=1.96 TeV; these data were published in [86].

Fig. 5.24 shows a comparison between D0 data and the predictions of PYTHIA and SHERPA. Results from both generators were normalized to the ratio $\sigma(Z/\gamma^* + \geq 1\ jet)/\sigma(Z/\gamma^* + X)$ as obtained from data. This is the same choice that was made also in the original paper. Jets were reconstructed with D0 Run II Infrared Legacy cone algorithm [21], that is implemented in Rivet.

We notice that PYTHIA underestimates the ratio, already for two jets. SHERPA

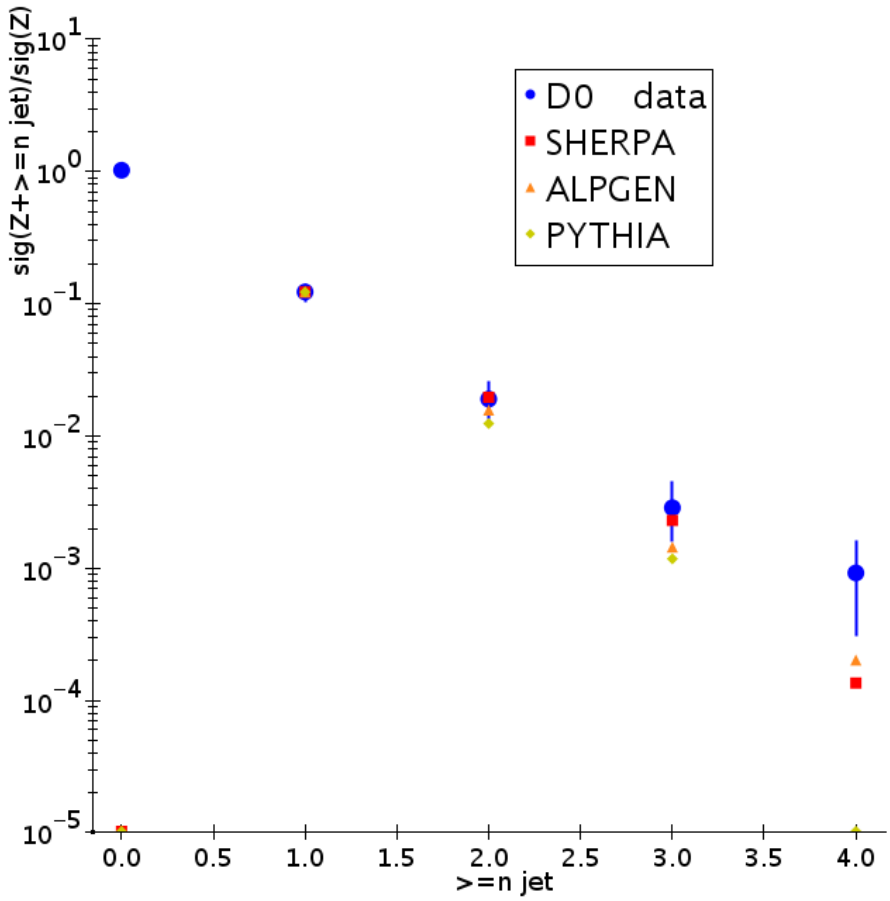


Figure 5.24: Measurement of the $\sigma(Z/\gamma^* + \geq njet)/\sigma(Z/\gamma^* + X)$ ratio by the D0 collaboration. Comparison to `PYTHIA`, `AlpGen` and `SHERPA` are shown. The three generators were normalized to the ratio for $\sigma(Z/\gamma^* + \geq 1\, jet)$, as done in the original paper.

on the other hand correctly reproduces the ratio up to the third jet. The ratio for $\geq$ 4 jets is underestimated in `SHERPA`. This is most probably due to the fact that in our simulation we allowed for up to three jets emerging from the ME. `AlpGen` (with `PYTHIA` for PS and hadronization) is better than `PYTHIA` but worse than `SHERPA` for the second and third jet; it is better than `SHERPA` for the fourth jet.

5.4.3 Measurement of $Z/\gamma^* + jet + X$ at Tevatron Run II by D0 collaboration

In [87] the D0 collaboration measured the Z boson and leading jet p_T and rapidity distributions from Run II data with an integrated luminosity of about 1 fb^{-1} in events with at least one jet.

Jets were reconstructed with D0 Infrared Legacy Cone [21] with a radius of 0.5,

a minimum p_T of 20 GeV and rapidity $|y| < 2.8$. Leptons were required to have an invariant mass between 65 GeV and 115 GeV, and rapidity $|y| < 1.7$.

Distributions obtained from the three generators were normalized to the measured cross section. The differential cross section as a function of Z p_T and Z rapidity y is represented in Fig. 5.25. Data points and predictions from SHERPA, AlpGen+PYTHIA and PYTHIA are shown in each plot. These comparisons are consistent with comparisons shown in the original paper. AlpGen is the best at describing the Z p_T spectrum. SHERPA seems to be too hard, while PYTHIA is too soft. PYTHIA is expected to be softer than the two other generators because it lacks higher order real emission corrections. The three generators describe y distribution reasonably.

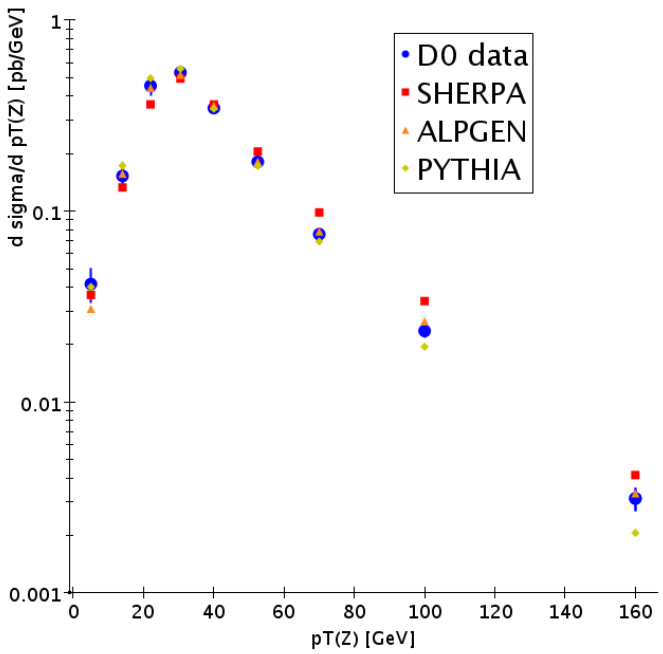
The differential cross section for the leading jet p_T and y distribution are shown in Fig. 5.26. The p_T distribution is still best reproduced in AlpGen+PYTHIA, while SHERPA appears to be too hard, and PYTHIA too soft. Concerning the y distribution it seems that SHERPA is doing the best job, while both AlpGen+PYTHIA and PYTHIA predict slightly more central distributions.

5.5 Summary

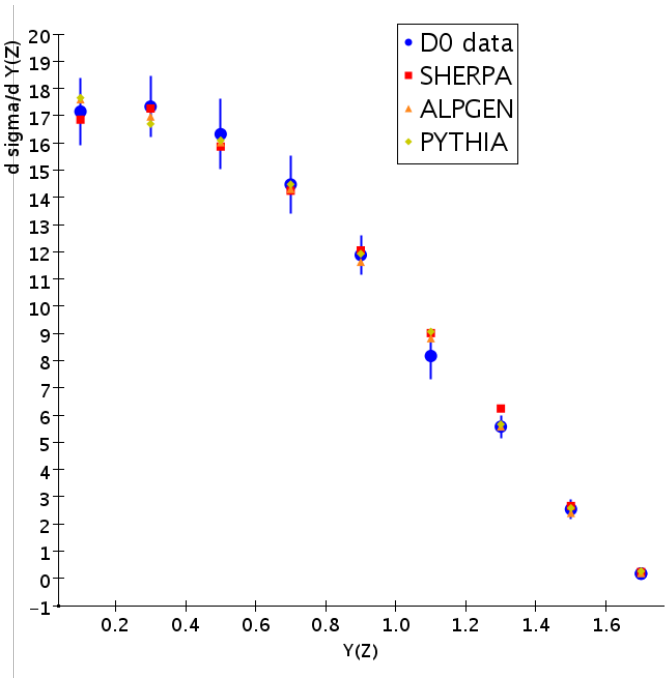
We studied the dependence of AlpGen and SHERPA on the parameters that are used to steer the ME-PS matching and on the scales definition. The change of the matching parameter in SHERPA has a not negligible effect on the total cross section (15% between the two extremes considered). On the other hand the effect on the AlpGen total cross section is almost negligible. Concerning the shape of the distributions, as q_{cut} in increased in SHERPA, p_T spectra tend to be slightly softer; this has been ascribed to the reduced phase space for ME emission. The change in shape in AlpGen is almost negligible.

The change of the scales definition has an effect of about 30% in the total cross section in SHERPA and about 15% in AlpGen. Concerning the change in shape, we observed that the p_T spectrum in SHERPA changes mainly at low p_T, while in AlpGen the low p_T shape is unchanged, and the effect is more evident at high p_T.

Finally we compared AlpGen, SHERPA and PYTHIA with data collected by the D0 collaboration at Tevatron $p\bar{p}$ collider. Concerning the Z p_T measurements, it looks like SHERPA is the best at describing the Z p_T spectrum in Run I data for the inclusive Z production. On the other hand, in newer measurements for $Z+ \geq 1\ jets$, it looks like AlpGen is the best at describing the Z p_T spectrum, while SHERPA is too hard. The same is observed in these newer data for the leading jet p_T. So, concerning the Z and leading jet p_T spectra, probably we still lack the definitive answer whether AlpGen or SHERPA is the best at describing data. We also made comparisons with a paper in which the ratio $\sigma(Z/\gamma^* + \geq njet)/\sigma(Z/\gamma^* + X)$ is measured. In this case it appears that SHERPA is reproducing data better than AlpGen.

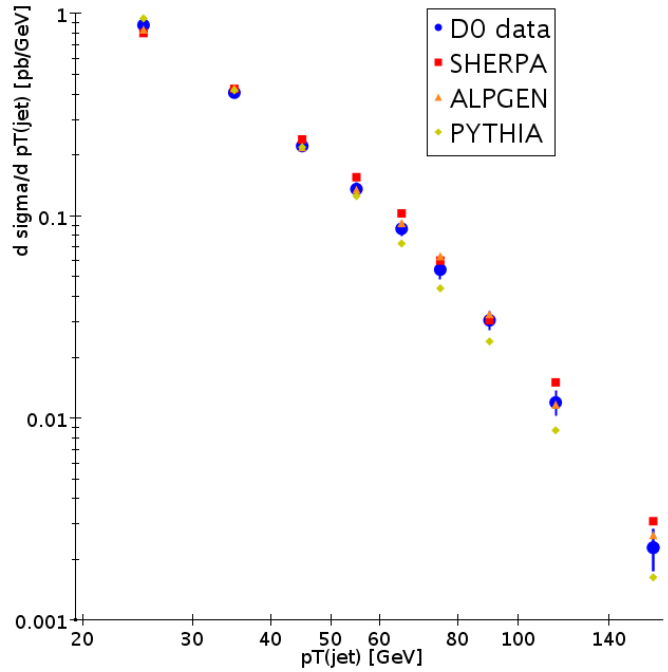


(a)

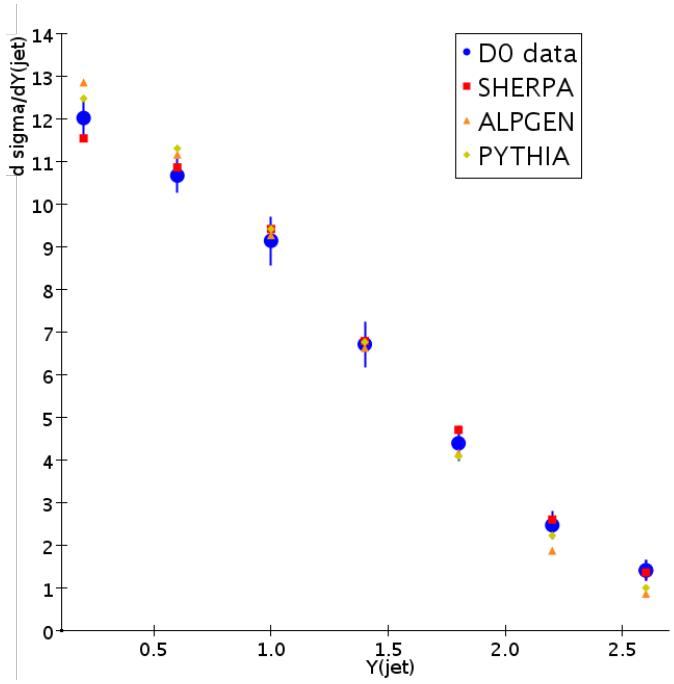


(b)

Figure 5.25: (a) p_T and (b) rapidity differential cross section for the Z boson. Total cross section is normalized to the measured cross section.



(a)



(b)

Figure 5.26: (a) p_T and (b) rapidity differential cross section for the leading jet. Total cross section is normalized to the measured cross section.

Chapter 6

Fast Simulation and Reconstruction of $Z/\gamma^* + jets$ in CMS

In this chapter we use the Fast Simulation of the CMS detector to see how events generated with `AlpGen` and `SHERPA` are reconstructed in our detector; in particular we want to see how the differences spotted at generator level are smeared when looking at the reconstructed quantities.

We shall make comparisons between the quantities reconstructed with CMS reconstruction software and the corresponding generated quantities obtained with a Rivet analysis on the generated sample.

We will finally compare the reconstructed quantities on the samples produced with `AlpGen` and `SHERPA`.

6.1 Event generation with `AlpGen` and `SHERPA`

`AlpGen` and `SHERPA` samples were produced with as similar configuration as possible. Both samples were produced requiring an invariant mass for the lepton pair higher than 50 GeV; no cuts were applied on the lepton or jet η. For both samples the electron and muon Z/γ^* decay channels were considered.

We produced events with up to three partons from the ME.

For both `AlpGen` and `SHERPA` we used the CTEQ6L1 [80] pdf set.

In `AlpGen` we used `PYTHIA` for the showering and hadronization. `PYTHIA` was setup using the so called "D6T Tune" [88] for the underlying event (UE).

For `SHERPA` we used a dedicated UE tune, based on CTEQ6L1 (the default tune is based on CTEQ6L).

For both generators we produced events for an integrated luminosity of approximately $100\,\mathrm{pb}^{-1}$. Pile-up effects, due to additional proton-proton interactions during each bunch crossing were not taken into account, because $100\,\mathrm{pb}^{-1}$ integrated

luminosity is expected to be collected in low luminosity conditions.

CMS reconstruction software (CMSSW) has interface libraries for both `AlpGen` and `SHERPA`. The production of events with `AlpGen` in CMSSW proceeds as follows: first we need to produce ME parton level events with standalone `AlpGen` and then run `PYTHIA` for PS and hadronization through the CMSSW `AlpGen` interface. For `SHERPA` we need to run the preliminary step for the calculation of cross section using standalone `SHERPA`; then, the generation of events happens from within CMSSW, using the `SHERPA` interface.

6.2 Simulation of detector effects and reconstruction

The simulation of detector effects has been done with the Fast Simulation of the CMS detector.

As introduced in Sec. 2.4.2, the Fast Simulation of detector effects skips the time consuming simulation of the interaction of particles with the detecting material (GEANT4 [49] step) and the digitization step. Simulated hits in the tracking detectors and shower evolution in the calorimetry are obtained using a parametrized approximation that aims at reproducing the full GEANT4 result.

Concerning the simulation of the hits in the tracker detector, a simplified geometry is used, which is made of cylinders and disks that are assigned a depth in radiation and interaction lengths that approximates the full geometry result. Propagation of particles between layers is performed analytically; when the particle traverses a layer, the effect of the interaction with the material is calculated according to parametrized functions. The hit position and error are then assigned with a Gaussian smearing with respect to the analytically calculated impact point. Track reconstruction starts with the emulation of the seeding efficiency. The pattern recognition step is skipped, the hit-to-track associations being taken directly from the Monte Carlo information; the track fit is performed using the standard algorithms.

Calorimeter energy deposits are obtained in two steps. First the shower is simulated as if the calorimeter were a homogeneous medium; then detector effects, such as crystal granularity, inefficiency, magnetic field influence, are simulated. Energy deposits are then turned into reconstructed signals simulating noise and zero-suppression[1].

Muon tracks in the muon detector are simulated using a parametrization of resolution and efficiencies, but without the simulation of hits in the muon chambers. The matching with tracker tracks is done using standard algorithms.

To pass from tracks and energy deposits in the calorimeters to physics objects, like muons, electrons or jets, the standard CMS algorithms are used. The output

[1]The zero-suppression is an online algorithm implemented in the calorimetry readout boards to avoid recording the signal of channels with output below a threshold level.

of the reconstruction algorithm is then processed through the so called Physics Analysis Toolkit (PAT) [89]. The PAT step produces high level physics objects like photons, muons, electrons, taus, jets. Those objects embed all the information that was produced in the reconstruction step, plus the matching with the Monte Carlo truth. The PAT is a very recent development of CMS reconstruction and analysis software and it will represent a standard tool for physics analyses.

6.3 Comparison between generated and reconstructed quantities

In this section we will study several observables, making comparisons between what is obtained at generator level and what is reconstructed. To do this, we setup an analysis in CMSSW that runs on the reconstructed quantities, and a Rivet analysis that runs on the same events at generator level.

Both analyses select events in which an e^+e^- or $\mu^+\mu^-$ pair is present with an invariant mass between 50 GeV and 150 GeV. Both leptons are also required to:

- have a p_T greater than 20 GeV;
- have $\eta < 2.4$, which means that they have to be identified by the tracking detectors (silicon tracker for the electrons, silicon tracker plus muon chambers for the muons);
- be isolated. Muons are isolated if the sum of the p_T of tracks reconstructed in the region between a cone with radius $\Delta R = 0.01$ and a cone with radius 0.3 around the muon is less than 3 GeV. Electrons have to fulfill the same requirement as muons for the tracker isolation and in addition their deposits have to be isolated in the calorimeter; this means that the calorimeter deposit between a veto cone with radius 0.1 and a cone with radius 0.5 around the electron has to be less than 5 GeV.

6.3.1 Lepton observables

Fig. 6.1 shows the p_T distribution for the muon pair as obtained at generator level and on the reconstructed quantities for the sample produced with `SHERPA`. The shape of the distribution is very well reconstructed; the overall efficiency is about 95%. Both muons are required to be "global muons", that is they have to be made out of a tracker track and a muon chamber track.

In Fig. 6.2 the η distribution for the muon pair is shown. The agreement in shape is quite good.

In Fig. 6.3 the p_T (a) and η distributions (b) for the positively charged lepton are represented. The shape of the p_T distribution is very well reconstructed, also at high p_T. This is possible thanks to the excellent performances of our tracking system.

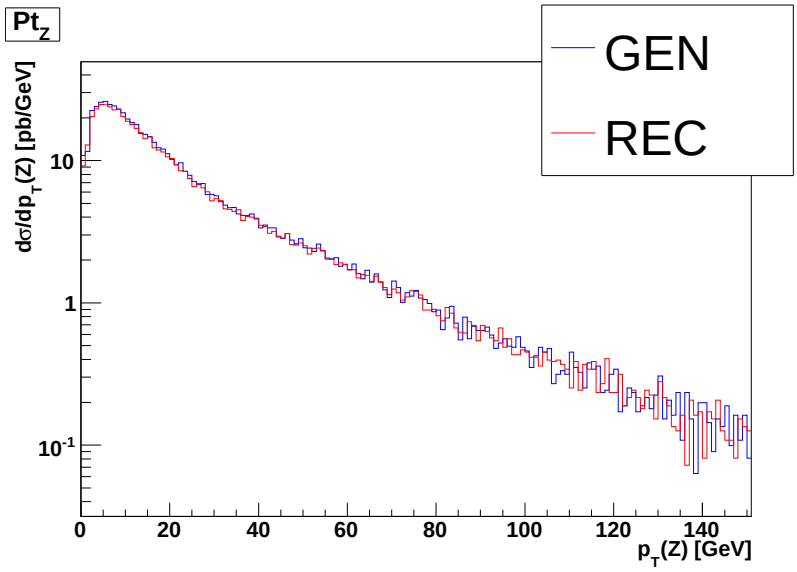


Figure 6.1: p_T distribution for the lepton pair in SHERPA. Generator level result is compared with what is obtained from the reconstructed quantities.

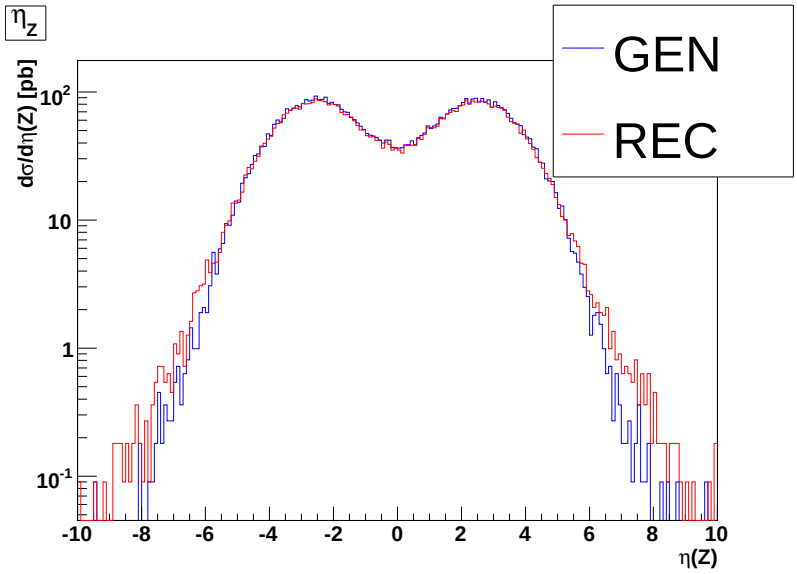
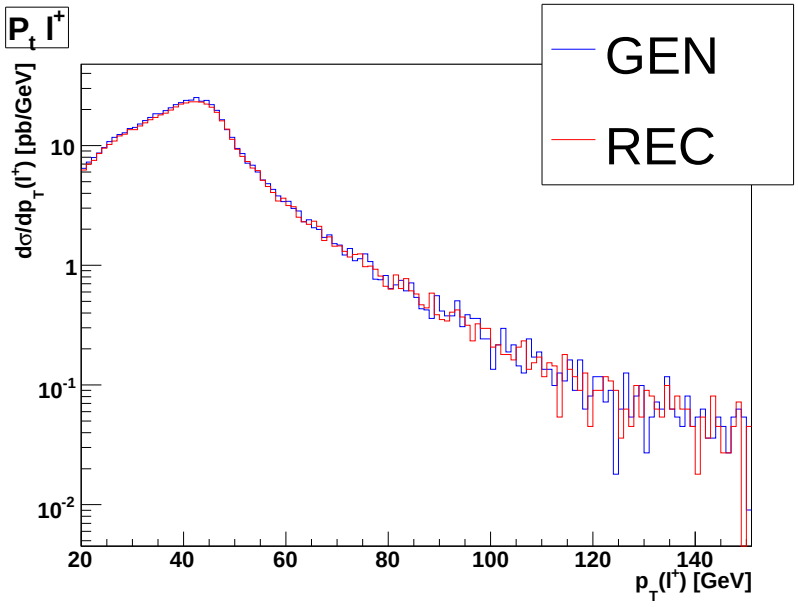


Figure 6.2: η distribution for the muon pair in SHERPA. Generator level result is compared with what is obtained from the reconstructed quantities.

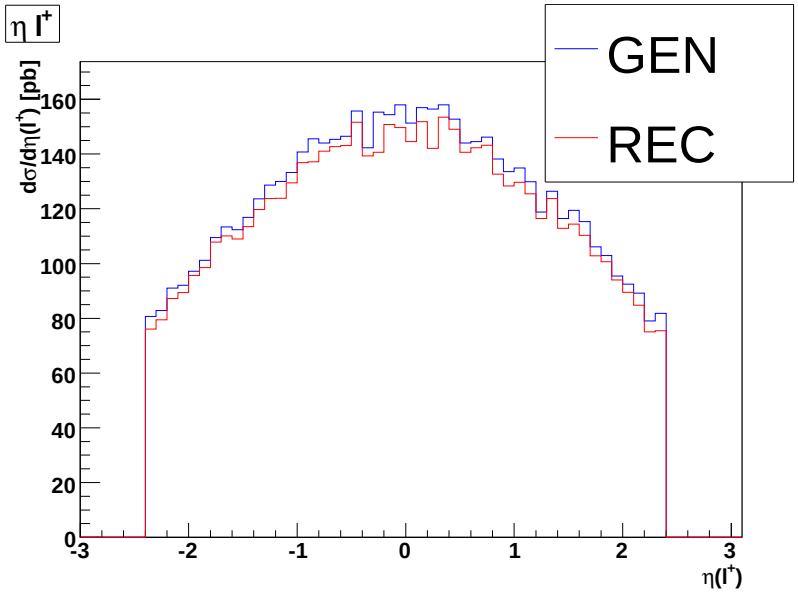
6.3.2 Jet observables

In Fig. 6.4 the jet multiplicity as obtained on the SHERPA sample for the $\mu^+\mu^-$ decay channel is shown. Generator level distribution is compared to the reconstructed one; the agreement is satisfactory.

The p_T distribution for the first and the second jet in the SHERPA sample for the



(a)



(b)

Figure 6.3: (a) p_T and (b) η for the positively charged muon.

muon decay channel is shown in Fig. 6.5. The shape of the distribution as obtained on the reconstructed sample follows the generator level shape fairly well. These results are obtained with the application of energy corrections to the reconstructed jets. These corrections aim at providing the correct energy scale for reconstructed jets. Energy corrections are needed because the calorimeter response is a not trivial function of the jet p_T and η. The main sources of bias in the jet reconstruction are:

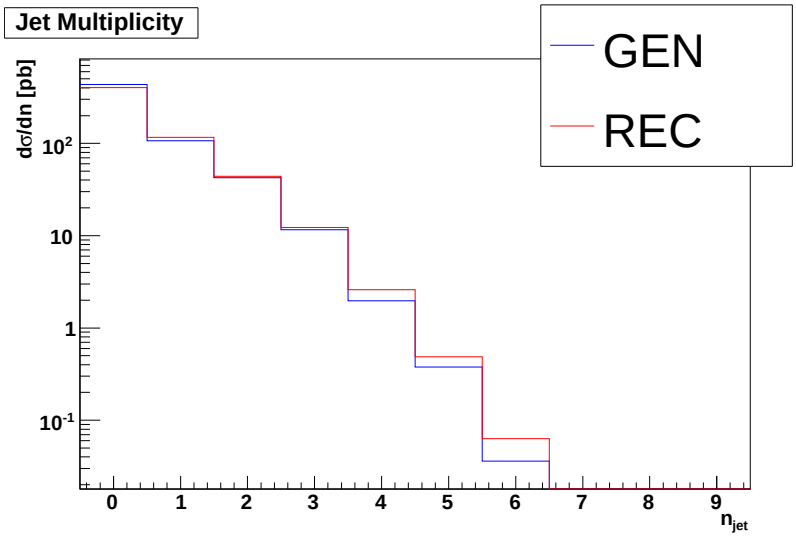
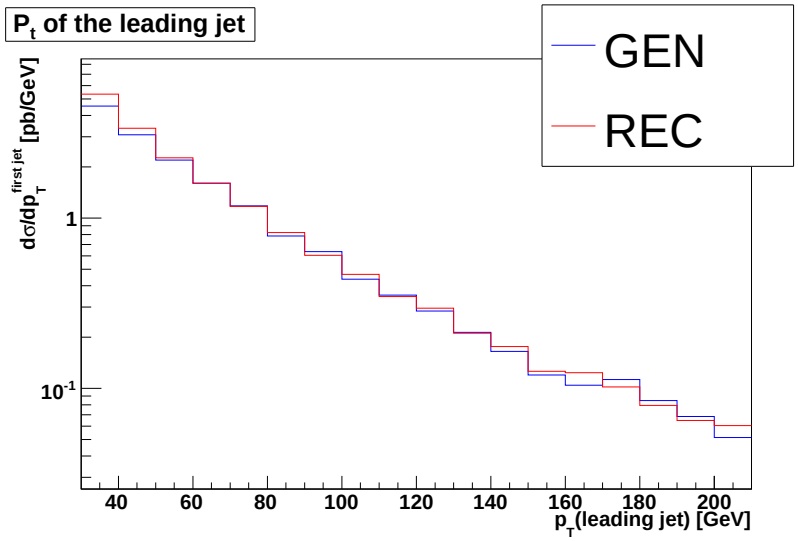


Figure 6.4: Jet multiplicity as reconstructed on the generated sample and on the reconstructed quantities fro the SHERPA sample.

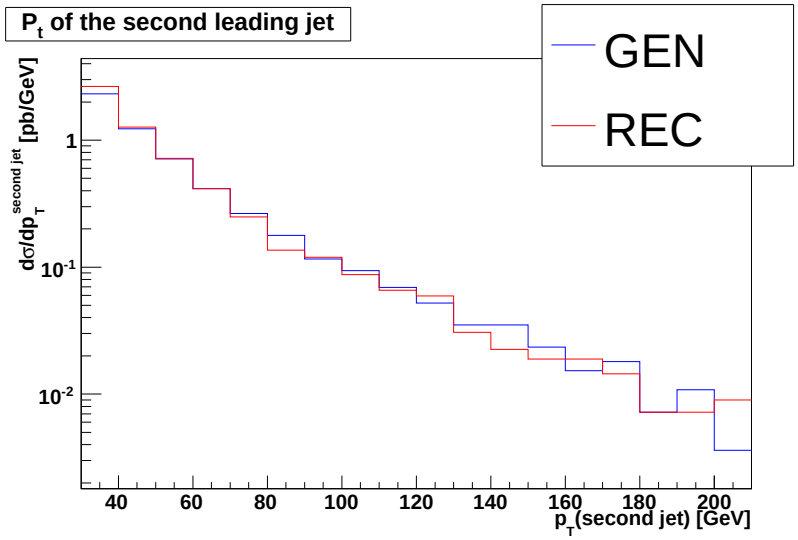
- The jet algorithm, which can be such that only part of the energy is actually included in the reconstructed jets.
- The magnetic field, which can deflect low p_T particles such that they are not included in the jet they belong to.
- The non-flat calorimeter response as a function of p_T, that is due to the fact that the hadron calorimeter is non-compensating.
- The calorimeters granularity.
- The interaction of particles with the material traversed before reaching the calorimeters.

CMS has implemented two kind of jet corrections [90]. The first method is often referred to as Monte Carlo driven method: the basic idea is to compare jets reconstructed on a full simulation sample and the corresponding generator level jets. The correction is given by the ratio between the reconstructed ${p_T}^{\mathrm{REC}}$ and the generator level jet ${p_T}^{\mathrm{GEN}}$; the correction is mapped as a function of p_T and η. These Monte Carlo driven corrections were used in this work.

The second method is referred to as data driven factorized correction. This method is designed to be independent from the Monte Carlo and is used to derive corrections directly from data. The method is called "factorized" because the correction happens in two steps: first the response is corrected to be flat in η (relative correction), then a correction is applied to make the response flat in p_T (absolute correction). The relative η correction is obtained using di-jet events in which one



(a)



(b)

Figure 6.5: p_T distribution for the first (a) and the second (b) jet as obtained at generator level and on the reconstructed quantities in the SHERPA sample.

jet is in a central η region ($|\eta| < 1.3$). The absolute correction is then applied using p_T-balance techniques in $(\gamma + 1\ jet)$ or $(Z + 1\ jet)$ events.

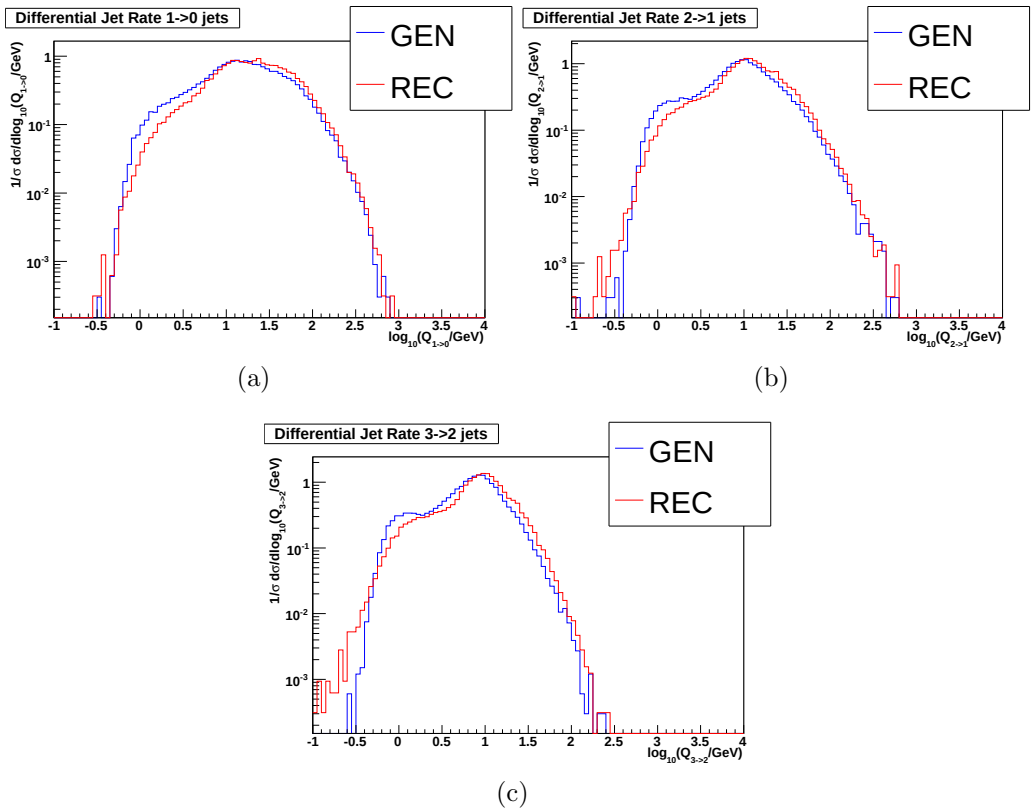
Since energy corrections need to be applied to the reconstructed jets, this makes the task of calculating differential jet rates on reconstructed quantities particularly difficult. We developed a technique to compute differential jet rates using calorimeter towers. On Monte Carlo, as described in Sec. 4.1.2, the calculation is done running the $k_\perp$ clustering algorithm in inclusive mode and then looking for the relevant recombination throughout the clustering sequence. On reconstructed

quantities, the fact that the calorimeter response is not flat in p_T and η alters the clustering sequence with respect to what happens on the MC truth.

The bias in the jet energy scale due to detector effects alters the $k_\perp$ distances d_{ij} (between pseudo-jets) and d_{iB} (between a pseudo-jet and the beam line), thus the clustering sequence is altered. In case the ordering of the clusterings is not heavily altered by detector effects we can try to reconstruct differential jet rates out of calorimeter deposits. To do so, we need a two step procedure; for the calculation of the $n \to n-1$ differential jet rate we set-up the following recipe:

1. run the $k_\perp$ clustering in "exclusive" mode and require exactly n jets in the event. In doing so, the d_{cut} that steers the "exclusive" $k_\perp$ is automatically adjusted so that exactly n jets are built.

2. apply jet corrections to the n reconstructed jets, calculate the distances d_{ij} and d_{iB} for the corrected jets and find the smallest among d_{iB} and d_{ij}: that is the value for the differential jet rate $n \to n-1$.



(a) (b)

(c)

Figure 6.6: Differential jet rates in SHERPA at generator level and reconstructed: (a) $1 \to 0$, (a) $2 \to 1$, (c) $3 \to 2$

Differential jet rates are shown in Fig. 6.6 for the transitions $1 \to 0$ (a), $2 \to 1$ (b), $3 \to 2$ (c). Both the generator level and the reconstructed differential jet

rates are shown in each plot. The shape of the distribution is reasonably well reconstructed, even if there seem to be a shift of the reconstructed distribution to the right of the generator level one. This is likely to be an effect of an energy over-correction.

6.4 Comparison of reconstructed quantities for AlpGen and SHERPA

In this section we make comparisons between AlpGen and SHERPA on the reconstructed quantities, to see to what extent the differences that were spotted in Sec. 5.3 are washed out by detector effects.

6.4.1 Leptonic observables

In Fig. 6.7 the reconstructed Z boson transverse momentum in the muon decay channel is shown. The distributions show the differences that were already observed

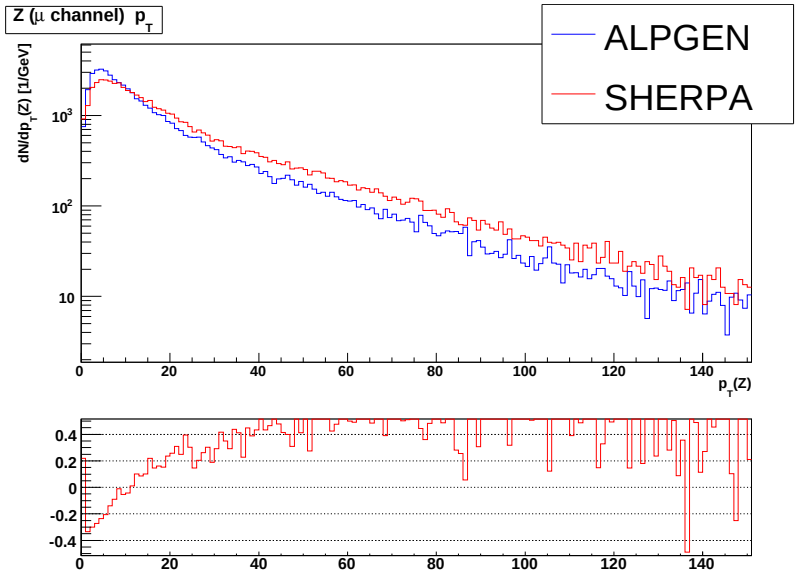


Figure 6.7: Z boson transverse momentum in the muon decay channel as reconstructed in AlpGen and SHERPA samples. The overall normalization is the number of events expected for $100\,\mathrm{pb}^{-1}$ by the two generators respectively.

at generator level, with SHERPA showing a significantly harder spectrum. Also the η spectrum represented in Fig. 6.8 confirms what was observed at generator level, with SHERPA showing a more central spectrum than AlpGen.

The muon momentum and η distribution as reconstructed on the samples produced with AlpGen and SHERPA are shown in Fig. 6.9. These distributions show the

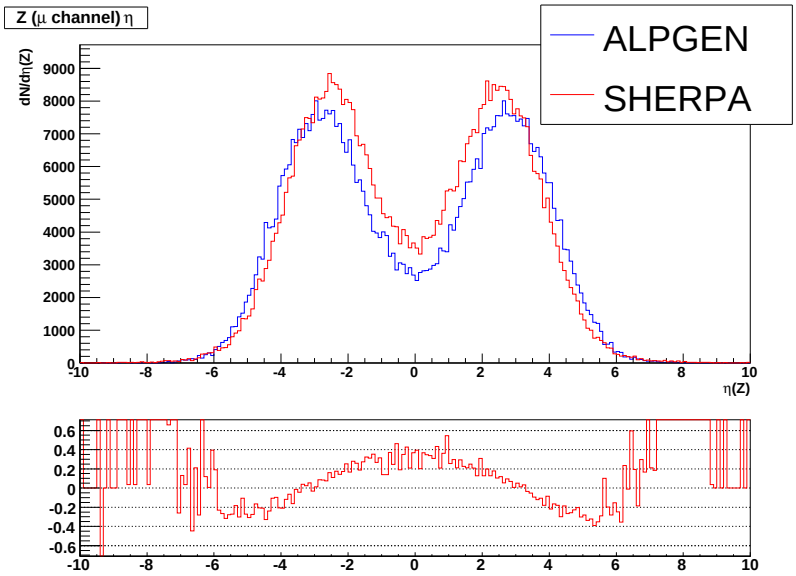


Figure 6.8: Z boson η distribution in the muon decay channel as reconstructed in `AlpGen` and `SHERPA` samples. The overall normalization is the number of events expected for $100\,\mathrm{pb}^{-1}$ by the two generators respectively.

expected behavior too, with `SHERPA` showing a harder spectrum than `AlpGen`. The η distribution is quite similar as already observed at generator level.

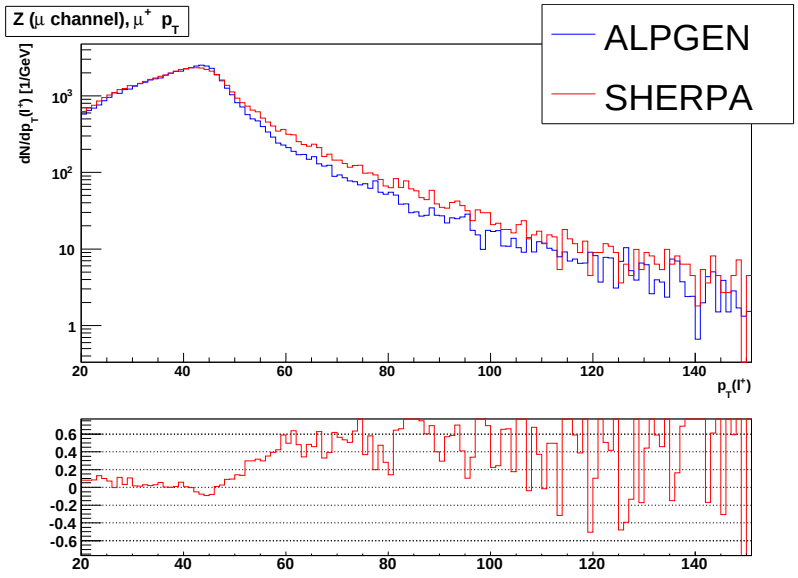
p_T and η distributions for the lepton observables in the e^+e^- decay channel are shown in Fig. 6.10. As expected the same behavior observed in the muon decay channel is observed also in the electron channel.

6.4.2 Jet observables

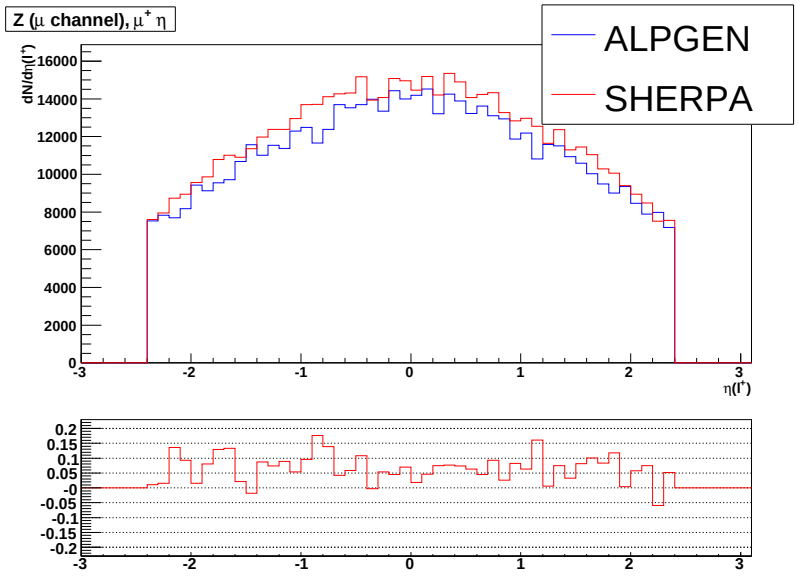
We now turn our attention to the jet observables. The jet multiplicity as reconstructed in the `AlpGen` and `SHERPA` samples for the electron and muon Z decay channels is represented in Fig. 6.11. As expected the mean jet multiplicity is higher in `SHERPA` even if `AlpGen` shows a somehow longer tail.

The leading jet p_T distribution is shown in Fig. 6.12. Even if the total cross section for emission of 1 jet is higher in `SHERPA`, distributions are similar in both the muon (a) and electron (b) decay channels.

The same holds true for the p_T of the second jet as depicted in Fig. 6.13.



(a)



(b)

Figure 6.9: (a) p_T distribution and (b) η distribution for the positively charged muon as reconstructed on the AlpGen and SHERPA samples. The overall normalization is the number of events expected for 100 pb^{-1} by the two generators respectively.

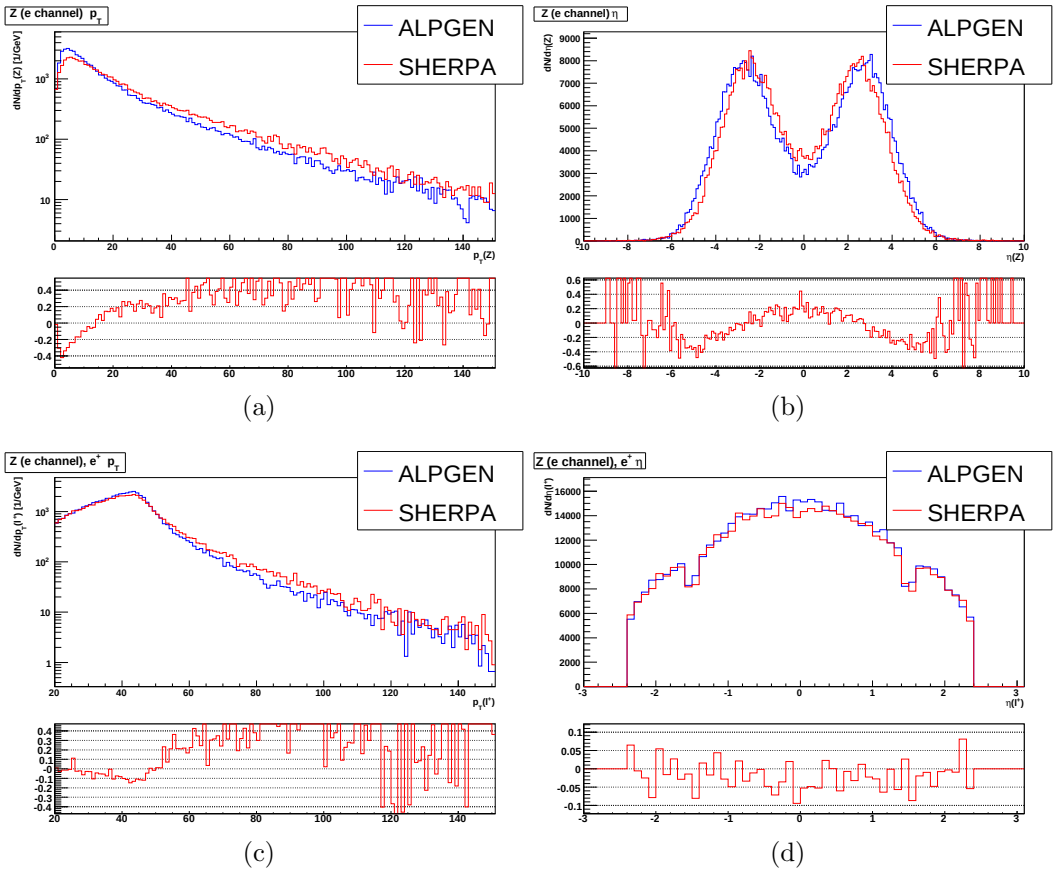
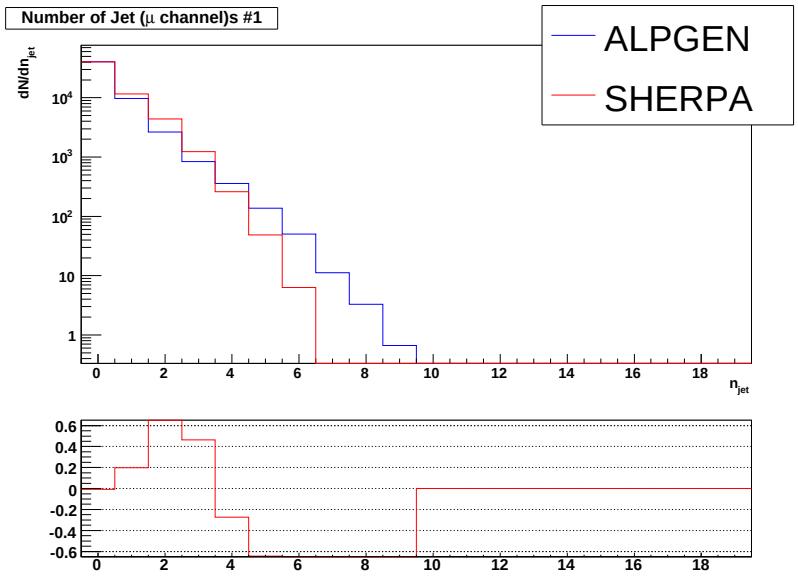
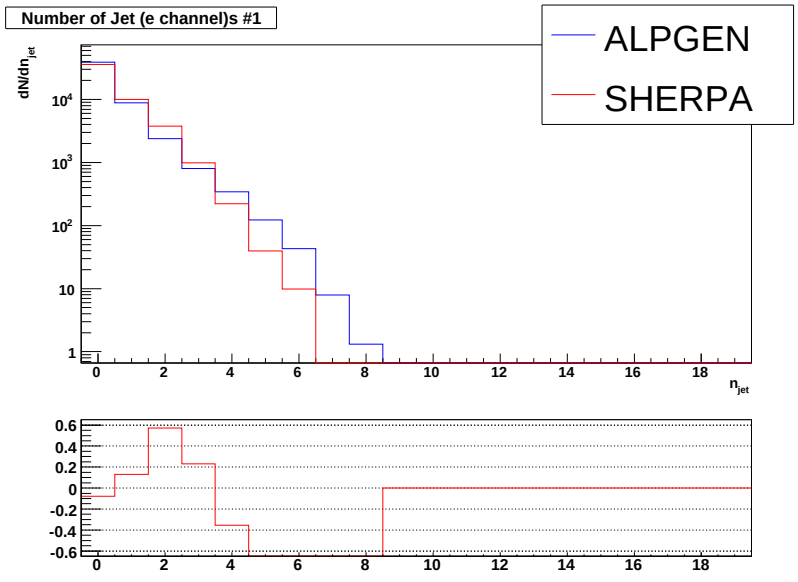


Figure 6.10: (a) p_T distribution and (b) η distribution for the Z boson; (c) p_T distribution and (d) η distribution for the positron. Reconstructed quantities for `AlpGen` and `SHERPA` are shown. The overall normalization is the number of events expected for $100\,\mathrm{pb}^{-1}$ by the two generators respectively. Relative difference plot with respect to `AlpGen` is shown.

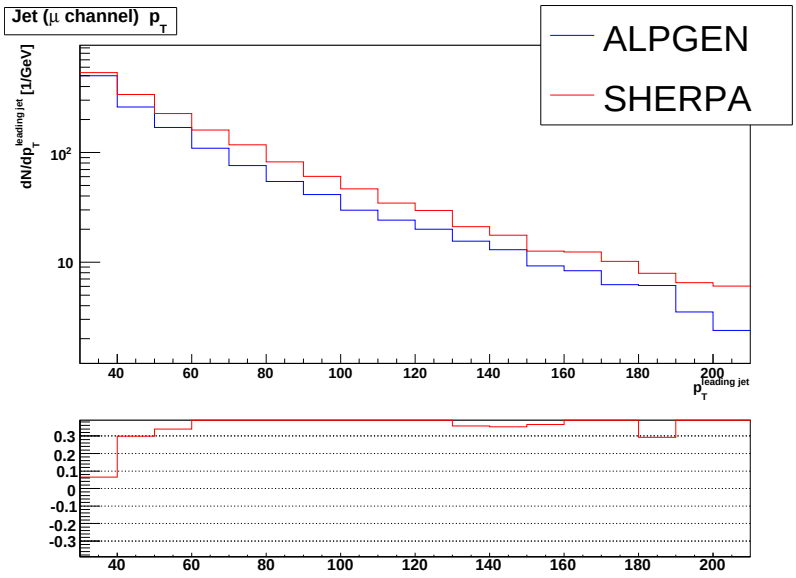


(a)

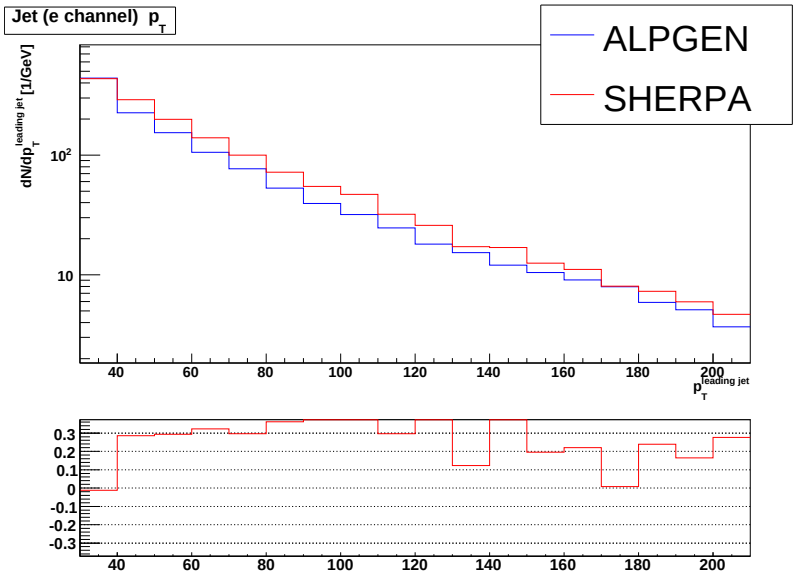


(b)

Figure 6.11: Jet multiplicity as reconstructed in the AlpGen and SHERPA samples for the Z muon (a) and electron (b) decay channels. The overall normalization is the number of events expected for 100 pb^{-1} by the two generators respectively. Relative difference plot with respect to AlpGen is shown.

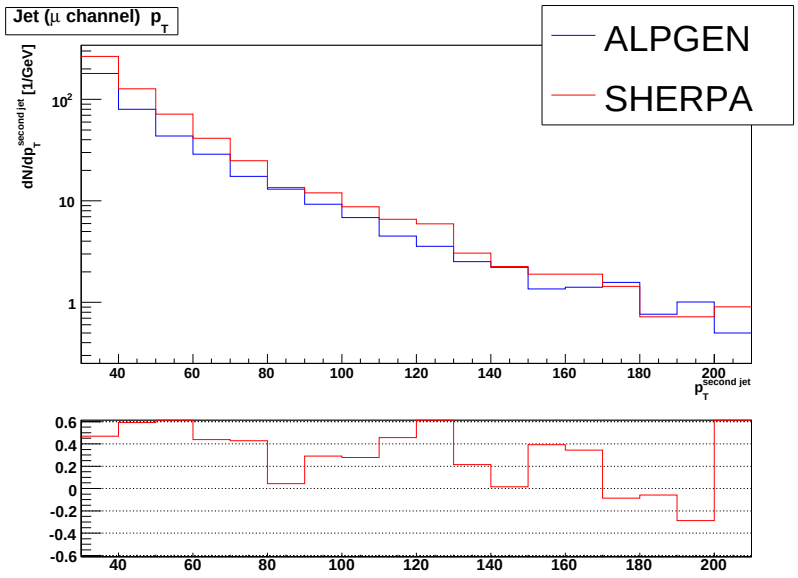


(a)

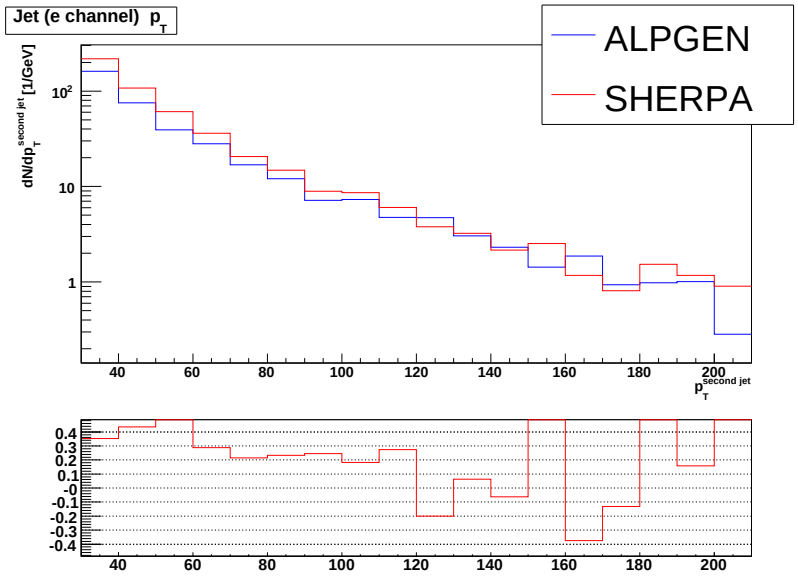


(b)

Figure 6.12: p_T of the leading jet as reconstructed in the `AlpGen` and `SHERPA` samples for the Z muon (a) and electron (b) decay channels. The overall normalization is the number of events expected for 100 pb^{-1} by the two generators respectively. Relative difference plot with respect to `AlpGen` is shown.



(a)



(b)

Figure 6.13: p_T of the second jet as reconstructed in the AlpGen and SHERPA samples for the Z muon (a) and electron (b) decay channels. The overall normalization is the number of events expected for 100 pb^{-1} by the two generators respectively. Relative difference plot with respect to AlpGen is shown.

Conclusion

The work presented in this thesis describes the differences among various event generators in the production of $Z/\gamma^* + jets$ events. We used mainly four event generators: PYTHIA, HERWIG, AlpGen and SHERPA.

PYTHIA and HERWIG are Leading Order event generators that implement QCD radiation pattern using the Parton Shower technique. Both generators implement a modified Parton Shower that can fully take into account the ME effect for one additional parton emission from the initial state.

AlpGen and SHERPA assert that they can describe the QCD radiation pattern using multi-parton ME calculations matched with Parton Shower. Those generators use a different matching prescription to combine the ME part and the PS part of the calculation.

We investigated the differences among these approaches. The comparison was carried out in various steps. First, we investigated the effect of ME corrections in PYTHIA and HERWIG. We verified that ME corrected distributions obtained with the two generators agree fairly well. We noticed that switching off the ME corrections has a different effect on the two generators. One of the most informative distributions in this respect is the Z boson transverse momentum. While in HERWIG the low p_T shape seems to be unaffected by ME corrections, in PYTHIA the low p_T shape changes considerably.

We compared ME-corrected PYTHIA with AlpGen and SHERPA configured so that at most one additional parton can emerge from the ME calculations. In such conditions results from the three generators should be similar, because the ME content should be the same. Actually we observe that SHERPA follows ME corrected PYTHIA quite well, in all the distributions that we considered. On the other hand AlpGen (used with the PYTHIA for Parton Shower) shows differences with respect to ME corrected PYTHIA. One of the distributions in which the discrepancy is more evident is the Z p_T spectrum. AlpGen follows uncorrected PYTHIA at low p_T, then it starts to follow corrected PYTHIA. The reason for this effect is actually not related to AlpGen itself but to the fact that when PYTHIA is used to shower AlpGen events ME corrections are switched off, because AlpGen itself is going to introduce its own corrections. The side effect is that the low p_T shape in the Z p_T spectrum changes as an effect of the deactivation of ME corrections; AlpGen cannot modify the shape in that region, that is fully determined by the PS alone.

We also compared AlpGen and SHERPA allowing both generators to emit up to three partons from the ME calculations. We observed that p_T distributions in AlpGen are significantly softer than SHERPA, and that the shape of both the lepton and jet observables are different. We also made comparisons to Tevatron data collected by the D0 collaboration and published in three papers. The first paper reported a measure of the inclusive Z p_T spectrum with data collected at $\sqrt{s}$=1.8 TeV. SHERPA appears to be the best at describing these data. The second paper considered concerned the ratio $\sigma(Z/\gamma^* + \geq n\ jet)/\sigma(Z/\gamma^*$ inclusive). Also in this

case SHERPA appears to reproduce data better than AlpGen and PYTHIA. Finally we made comparisons with a recent paper in which the p_T and rapidity spectra of the Z boson and of the leading jet in $Z/\gamma^* + \geq 1\ jet$ events were measured. In this case AlpGen is the best at describing spectra, with SHERPA showing too hard p_T spectra. Thus, this comparison still does not give a definitive answer to the question whether AlpGen or SHERPA is the best at describing the effect of higher order radiation.

Finally, we used the Fast Simulation of the CMS detector and we reconstructed events produced with AlpGen and SHERPA. The differences spotted at generator level are still clearly visible on the reconstructed quantities. Thus, it should be not difficult to discriminate between these different theoretical models once data will be available.

Acknowledgements

I would like to thank my supervisor Dr. M. Meschini for his guidance and for the constant dedication that he showed in following my work.

I would also like to thank Prof. J. Butterworth and Dr. J. Monk who supervised my work during my stay at University College of London. I'm very grateful to Prof. F. Krauss, Dr. F. Siegert, Dr. F. Piccinini, Prof. S. Catani for many useful discussions.

This work was supported in part by the Marie Curie Research Training Network "MCnet", contract number MRTN-CT-2006-035606.

Bibliography

[1] F. Halzen and A. Martin, *Quarks and leptons: an introductory course in modern Particle Physics*. Wiley, New York USA, 1984.

[2] S. L. Glashow, "Partial symmetries of weak interactions," *Phys. Rev.* **22** (1961) 579–588.

[3] A. Salam, *Elementary Particle Theory*. Almquist and Wiksells, N. Svartholm, Stockholm, 1968.

[4] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.* **19** (1967) 1264–1266.

[5] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, "EXPERIMENTAL TEST OF PARITY CONSERVATION IN BETA DECAY," *Phys. Rev.* **105** (1957) 1413–1414.

[6] R. L. Garwin, L. M. Lederman, and M. Weinrich, "Observations of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: The Magnetic Moment of the Free Muon," *Phys. Rev.* **105** (1957) 1415–1417.

[7] M. Goldhaber, L. Grodzins, and A. W. Sunyar, "Helicity of neutrinos," *Phys. Rev.* **109** (1958) 1015–1017.

[8] E. Fermi, "*Trends to a theory of Beta radiation.*," *Nuovo Cimento* **11** (1952) 1–19.

[9] Particle Data Group, "Review of particle physics," *Phys. Rev.* **D66** (2002) no. 010001, .

[10] N. Cabibbo, "Unitary Symmetry and Leptonic Decays," *Phys. Rev. Lett.* **10** (1963) 531–533.

[11] M. Gell-Mann, "A Schematic Model of Baryons and Mesons," *Phys. Lett.* **8** (1964) 214–215.

[12] M. Y. Han and Y. Nambu, "Three-triplet model with double SU(3) symmetry," *Phys. Rev.* **139** (1965) B1006–B1010.

[13] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the Color Octet Gluon Picture," *Phys. Lett.* **B47** (1973) 365–368.

[14] D. J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories. 1," *Phys. Rev.* **D8** (1973) 3633–3652.

[15] S. Weinberg, "Nonabelian Gauge Theories of the Strong Interactions," *Phys. Rev. Lett.* **31** (1973) 494–497.

[16] **CTEQ** Collaboration, R. Brock *et al.*, "Handbook of perturbative QCD: Version 1.0," *Rev. Mod. Phys.* **67** (1995) 157–248.

[17] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*. Cambridge University Press, N. Svartholm, Stockholm, 1996.

[18] **Particle Data Group** Collaboration, W. M. Yao *et al.*, "Review of particle physics," *J. Phys.* **G33** (2006) 1–1232.

[19] **JADE** Collaboration, S. Bethke *et al.*, "Experimental Investigation of the Energy Dependence of the Strong Coupling Strength," *Phys. Lett.* **B213** (1988) 235.

[20] S. Catani, Y. L. Dokshitzer, and B. R. Webber, "The K-perpendicular clustering algorithm for jets in deep inelastic scattering and hadron collisions," *Phys. Lett.* **B285** (1992) 291–299.

[21] G. C. Blazey *et al.*, "Run II jet physics," `arXiv:hep-ex/0005012`.

[22] G. P. Salam and G. Soyez, "A practical Seedless Infrared-Safe Cone jet algorithm," *JHEP* **05** (2007) 086, `arXiv:0704.0292 [hep-ph]`.

[23] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "MRST2001: Partons and alpha(s) from precise deep inelastic scattering and Tevatron jet data," *Eur. Phys. J.* **C23** (2002) 73–87, `arXiv:hep-ph/0110215`.

[24] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys.* **B126** (1977) 298.

[25] H. L. Lai *et al.*, "Global QCD analysis and the CTEQ parton distributions," *Phys. Rev.* **D51** (1995) 4763–4782, `arXiv:hep-ph/9410404`.

[26] The LHC Study Group, "The large hadron collider conceptual design report." Cern/ac 95-05, 1995.

[27] Lep Injector Study Group, "Lep design report vol. 1: The lep injector chain." Cern-lep-th-83-29, cern-ps-dl-83-81, cern-sps-83-26, lal-rt-83-09, 1983.

[28] Lep Injector Study Group, "Lep design report vol. 2: The lep main ring." Cern-lep/8-01, 1984.

[29] Lep Injector Study Group, "Lep design report vol. 3: Lep2." Cern-ac-96-01-lep2, 1996.

[30] The CMS Collaboration, "The compact muon solenoid technical proposal." Cern/lhcc 94-38, lhcc/p1, December, 1994.

[31] The ATLAS Collaboration, "Atlas: A toroidal lhc apparatus technical proposal." Cern/lhcc 94-18, lhcc/p1, December, 1994.

[32] The LHCb Collaboration, "A large hadron collider beauty experiment for precision measurements of cp-violation and rare decays lhcb technical proposal." Cern/lhcc 98-04, lhcc/p4, 1998.

[33] The ALICE Collaboration, "A large ion collider experiment technical proposal design." Cern/lhcc 95-71, lhcc/p3, 1995.

[34] *The CMS magnet project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[35] V. Karimki, *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[36] *The CMS tracker: addendum to the Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2000.

[37] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[38] P. Bloch, R. Brown, P. Lecoq, and H. Rykaczewski, *Changes to CMS ECAL electronics: addendum to the Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2002.

[39] G. Dissertori, "Electromagnetic calorimetry and $e\gamma$ performance in cms. oai:cds.cern.ch:687894," Tech. Rep. CMS-CR-2003-024, CERN, Geneva, Jul, 2003.

[40] *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS, CERN-LHCC-97-031. CERN, Geneva, 1997.

[41] *The CMS muon project: Technical Design Report*. Technical Design Report CMS, CERN-LHCC-97-032. CERN, Geneva, 1997.

[42] *CMS TriDAS project: Technical Design Report; 1, the trigger systems*. Technical Design Report CMS, CERN-LHCC-2000-038. CERN, Geneva, 2000.

[43] *CMS trigger and data-acquisition project: Technical Design Report.* Technical Design Report CMS, CERN-LHCC-2002-026. CERN, Geneva, 2002.

[44] **CMS** Collaboration, "CMSSW, CMS reconstruction software," 2008. `http://cms.cern.ch/iCMS/jsp/page.jsp?mode=cms&action=url&urlkey=CMS_OFFLINE`.

[45] B. Stroustrup, *The C++ Programming Language (Third Edition).* Addison-Wesley, 1997.

[46] M. Lutz, *Programming Python.* Sebastopol, CA: O'Reilly, 1996.

[47] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl. Instrum. Meth.* **A389** (1997) 81–86.

[48] M. Dobbs and J. B. Hansen, "The HepMC C++ Monte Carlo event record for High Energy Physics," *Comput. Phys. Commun.* **134** (2001) 41–46.

[49] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4: A simulation toolkit," *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[50] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *JHEP* **05** (2006) 026, `arXiv:hep-ph/0603175`.

[51] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions," *JHEP* **07** (2003) 001, `arXiv:hep-ph/0206293`.

[52] T. Gleisberg *et al.*, "SHERPA 1.alpha, a proof-of-concept version," *JHEP* **02** (2004) 056, `arXiv:hep-ph/0311263`.

[53] G. Corcella *et al.*, "HERWIG 6.5 release note," `arXiv:hep-ph/0210213`.

[54] T. Stelzer and W. F. Long, "Automatic generation of tree level helicity amplitudes," *Comput. Phys. Commun.* **81** (1994) 357–371, `arXiv:hep-ph/9401258`.

[55] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph," *JHEP* **02** (2003) 027, `arXiv:hep-ph/0208156`.

[56] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," *JHEP* **06** (2002) 029, `arXiv:hep-ph/0204244`.

[57] F. Caravaglios and M. Moretti, "An algorithm to compute Born scattering amplitudes without Feynman graphs," *Phys. Lett.* **B358** (1995) 332–338, `arXiv:hep-ph/9507237`.

[58] R. Kleiss and W. J. Stirling, "Spinor Techniques for Calculating p anti-p → W+- / Z0 + Jets," *Nucl. Phys.* **B262** (1985) 235–262.

[59] F. Krauss, R. Kuhn, and G. Soff, "AMEGIC++ 1.0: A matrix element generator in C++," *JHEP* **02** (2002) 044, arXiv:hep-ph/0109036.

[60] T. Sjostrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, ch. 10.3.2. 2006. arXiv:hep-ph/0603175.

[61] T. Sjostrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, ch. 10.3.3. 2006. arXiv:hep-ph/0603175.

[62] M. Dasgupta and G. P. Salam, "Resummation of non-global QCD observables," *Phys. Lett.* **B512** (2001) 323–330, arXiv:hep-ph/0104277.

[63] M. Dasgupta, "Parton shower Monte Carlos vs resummed calculations for interjet energy flow observables," arXiv:0706.2630 [hep-ph].

[64] M. Bengtsson and T. Sjostrand, "Coherent Parton Showers Versus Matrix Elements: Implications of PETRA - PEP Data," *Phys. Lett.* **B185** (1987) 435.

[65] M. H. Seymour, "Matrix element corrections to parton shower algorithms," *Comp. Phys. Commun.* **90** (1995) 95–101, arXiv:hep-ph/9410414.

[66] G. Miu and T. Sjostrand, "W production in an improved parton shower approach," *Phys. Lett.* **B449** (1999) 313–320, arXiv:hep-ph/9812455.

[67] M. Mangano, "The so-called MLM prescription for ME/PS matching," October, 4, 2002. Talk presented at the Fermilab ME/MC Tuning Workshop.

[68] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, "QCD matrix elements + parton showers," *JHEP* **11** (2001) 063, arXiv:hep-ph/0109231.

[69] F. Krauss, "Matrix elements and parton showers in hadronic interactions," *JHEP* **08** (2002) 015, arXiv:hep-ph/0205283.

[70] F. Krauss, R. Kuhn, and G. Soff, "AMEGIC++ 1.0: A matrix element generator in C++," *JHEP* **02** (2002) 044, arXiv:hep-ph/0109036.

[71] F. Krauss, A. Schalicke, and G. Soff, "APACIC++ 2.0: A parton cascade in C++," *Comput. Phys. Commun.* **174** (2006) 876–902, arXiv:hep-ph/0503087.

[72] A. Buckley, "CEDAR: tools for event generator tuning," arXiv:0708.2655 [hep-ph].

[73] "Design report Tevatron 1 project,". FERMILAB-DESIGN-1983-01.

[74] **CDF IIb** Collaboration, P. T. Lukens, "The CDF IIb detector: Technical design report,". FERMILAB-TM-2198.

[75] **D0** Collaboration, V. M. Abazov *et al.*, "Run IIb upgrade technical design report,". FERMILAB-PUB-02-327-E.

[76] T. Sjostrand and S. Mrenna. Private communication.

[77] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K(t) clustering algorithms for hadron hadron collisions," *Nucl. Phys.* **B406** (1993) 187–224.

[78] G. Corcella and M. H. Seymour, "Initial state radiation in simulations of vector boson production at hadron colliders," *Nucl. Phys.* **B565** (2000) 227–244, arXiv:hep-ph/9908388.

[79] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, "QCD matrix elements + parton showers," *JHEP* **11** (2001) 063, arXiv:hep-ph/0109231.

[80] J. Pumplin *et al.*, "New generation of parton distributions with uncertainties from global QCD analysis," *JHEP* **07** (2002) 012, arXiv:hep-ph/0201195.

[81] M. Cacciari and G. P. Salam, "Dispelling the N**3 myth for the k(t) jet-finder," *Phys. Lett.* **B641** (2006) 57–61, arXiv:hep-ph/0512210.

[82] J. Campbell, R. K. Ellis, and D. L. Rainwater, "Next-to-leading order QCD predictions for W + 2jet and Z + 2jet production at the CERN LHC," *Phys. Rev.* **D68** (2003) 094021, arXiv:hep-ph/0308195.

[83] A. Buckley *et al.*, "HepData and JetWeb: HEP data archiving and model validation," arXiv:hep-ph/0605048.

[84] V. V. Serbo, "Status of AIDA and JAS 3," *Nucl. Instrum. Meth.* **A502** (2003) 663–665.

[85] **D0** Collaboration, B. Abbott *et al.*, "Differential production cross section of Z bosons as a function of transverse momentum at $\sqrt{s} = 1.8$ TeV," *Phys. Rev. Lett.* **84** (2000) 2792–2797, arXiv:hep-ex/9909020.

[86] **D0** Collaboration, V. M. Abazov *et al.*, "Measurement of the ratios of the Z/G* + ¿= n jet production cross sections to the total inclusive Z/G* cross section in ppbar collisions at sqrt(s) = 1.96 TeV," *Phys. Lett.* **B658** (2008) 112–119, arXiv:hep-ex/0608052.

[87] **D0** Collaboration, V. M. Abazov *et al.*, "Measurement of differential Z/gamma*+jet+X cross sections in proton anti-proton collisions at sqrts=1.96 TeV," arXiv:0808.1296 [hep-ex].

[88] R. Field, "Pythia UE Tunes," May, 14, 2007. Talk presented at the CMS MC Tuning Mini Workshop.

[89] **CMS** Collaboration, "Physics Analysis Toolkit (PAT)," 2008. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuidePAT.

[90] **CMS** Collaboration, "Plans for jet energy corrections at cms." CMS Physics Analysis Summary JME-07-002, 2008. http://cms-physics.web.cern.ch/cms-physics/public/JME-07-002-pas.pdf.