Università degli Studi di Firenze
Scuola di Dottorato in Scienze

DIPARTIMENTO DI FISICA

Tesi di Dottorato di Ricerca in Fisica

XXII ciclo

(Settore Disciplinare FIS/04)

di

Giuseppe Broccolo

Study of prompt J/Ψ production in LHC proton-proton collisions at $\sqrt{s} = 10$ TeV in the J/$\Psi \rightarrow e^+e^-$ decay channel

Candidato: Giuseppe Broccolo .....

Relatore: Dr. Simone Paoletti .....

Coordinatore: Prof. Alessandro Cuccoli .....

Firenze, 31 Dicembre 2009

A chi mi ha voluto bene,
a chi ancora mi vuole bene,
a chi avrà il coraggio di volermene ancora...

...A Patrizio...

Contents

Introduction

In 2009 at CERN started operating the most powerful particle collider ever built before the Large Hadron Collider (LHC). Inside LHC high intensity proton beams are brought to collision with energies that will reach 14 TeV in the centre of mass when the machine will reach its design performance. The LHC will allow to investigate the symmetry breaking mechanism at the base of the *Standard Model* of fundamental interactions, allowing to search the yet undiscovered Higgs boson in all the mass range allowed by the theory, disclosing the origin of mass. New Physics might also be discovered by the $\sqrt{s} = 14$ TeV proton-proton interactions, like SUSY, Dark Matter, Extra Dimensions[1][2].

The LHC design energy and luminosity won't be available right from the start. The commissioning phase of the collider and the start-up program foresee to run at low luminosity and low energy ($\sqrt{s} = 7 - 10$ TeV) for a few months. The very first collisions produced by LHC at the end of 2009 reached energies of $\sqrt{s} = 900$ GeV and $\sqrt{s} = 2.36$ TeV. The most up to date program for 2010 is to run at $\sqrt{s} = 7$ TeV, with possible energy increases up to $\sqrt{s} = 10$ TeV in the second half of the year. While major discoveries will be unlikely, this period will be precious to calibrate and tune the detectors, and to perform some Standard Model physics, QCD studies and B-physics benefiting from the lowered trigger thresholds which will be available during the start-up phase and which won't be accessible at high energies and luminosity. To this respect, the LHC start-up phase can be used to bridge the studies performed at Tevatron (the $\sqrt{s} = 2$ TeV p-$\bar{\text{p}}$ collider operating at FNAL) to higher centre of mass energy.

One of the QCD studies which were performed at the Tevatron and need confirmation is the study of prompt J/Ψ production. $c\bar{c}$ bound states (called "charmonia") like the J/Ψ are produced in hadron collisions mainly either via decays of B mesons, featuring a displaced vertex due to the long b quark lifetime, either via the fragmentation of a quark or a gluon in QCD events, which is what is named "prompt" production.

Tevatron measurement of prompt charmonia did not match the theoretical predictions, based on the so called *Colour Singlet Model*, by orders of magnitude. Fits performed in the framework of an alternative model for "quarkonia" production, the NRQCD, allow to better describe the observed spectra and production rates.

At LHC will be possible to collect a higher statistics of data and to test the NRQCD at higher energy. In this work I will concentrate on the study of J/Ψ production in p-p interactions at LHC with the CMS detector. J/Ψ are reconstructed in the e^+e^- decay channel. CMS is particularly suited to this study, deploying the close distance between its Tracker and the high resolution electromagnetic calorimeter. This channel is, however, made challenging by the huge charged hadron combinatorial background present in p-p collisions at LHC.

The J/Ψ $\rightarrow e^+e^-$ channel is complementary to the J/Ψ $\rightarrow \mu^+\mu^-$ channel, the other signature viable in CMS to identify J/Ψ decays, and will allow a comparison of the results and a check for systematic errors induced by instrumental effects. The J/Ψ $\rightarrow e^+e^-$ and J/Ψ $\rightarrow \mu^+\mu^-$ channels will also complement each other in collecting statistics for B physics studies, like the search of the B_c meson in $B_c^{\pm} \rightarrow J/\Psi + \pi^{\pm}$ decay. The J/Ψ $\rightarrow e^+e^-$ sample will also be useful to study the CMS (ECAL and Tracker) detector performance and as a monitoring channel for the ECAL calibration.

In this work I will explain how it is possible to study Heavy Quarkonia production in the J/Ψ $\rightarrow e^+e^-$ channel, showing the achievable performance using Monte Carlo techniques.

This work is subdivided in 6 chapters: in the first one, I present some aspects of the strong interaction regarding heavy quark pairs production and theories underlying heavy quarkonia production, as well as some tools important to the simulation of events, which will be better described in chapter 3.

In chapter 2 I describe the CMS detector, in particular the two subdetector deployed for electron reconstruction, the Tracker and the ECAL.

In chapter 3 I show how is possible to simulate p-p collisions with Monte Carlo techniques, and all tools used for the simulation of signal and background, as well as used parameterization of the different models of prompt J/Ψ production simulation.

Chapter 4 is then dedicated to the description of the reconstruction of physical object like electrons four-momentum, decay vertexes, and of the performed identification algorithm of the electron pairs in signal events, extracting them from the huge combinatorial of background events.

In Chapter 5 I explain the problem of "triggering" events in LHC, and the techniques involved in CMS to trigger prompt J/Ψ $\longrightarrow e^+e^-$ events based on the reconstruction algorithm exposed in Chapter 4.

Finally, in Chapter 6 I show the used techniques and algorithms for the measurement of the differential cross section ($\frac{d\sigma}{dp_T}$) of prompt J/Ψ production and how to compare to

theory predictions, validating the quarkonia production models in pp collisions.

Chapter 1

Strong Interaction & Prompt J/Ψ Production at LHC

Basic concepts of particle physics are introduced, in particular the sector of the Strong Interactions of the *Standard Model*, and the direct (or "prompt") production of *charmonia* bound states in p-p collisions at LHC is described.
A non-relativistic description of the strong interactions is needed in order to account for the most recent measurements on prompt quarkonia production performed at the Tevatron.

In the following, the natural units are used and either in formulæ or in experimental results it is implicitly assumed, if it is not stated otherwise, $\hbar = c = 1$.

1.1 The *Standard Model* of the Elementary Particles

1.1.1 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 as it follows in (1.1):

$$\begin{pmatrix} \nu_{\mathrm{e}} \\ \mathrm{e} \end{pmatrix} \quad \begin{pmatrix} \nu_{\mu} \\ \mu \end{pmatrix} \quad \begin{pmatrix} \nu_{\tau} \\ \tau \end{pmatrix} \\ \begin{pmatrix} \mathrm{u} \\ \mathrm{d} \end{pmatrix} \quad \begin{pmatrix} \mathrm{c} \\ \mathrm{s} \end{pmatrix} \quad \begin{pmatrix} \mathrm{t} \\ \mathrm{b} \end{pmatrix} \tag{1.1}$$

Ordinary matter consists of leptons and quarks. Quarks exist only in bound states, which are called "hadrons". All the elements belonging to the three families have been directly

or indirectly observed and no experimental evidence of the existence of a 4^{th} generation is provided. In the Standard Model elementary particles interact with each other via three[1] fundamental interactions of Nature, which are mediated by **bosons**. The three forces of the Standard Model and the respective force carriers are:

Strong Interaction (8 massless gluons g) Only between colour-charged particles, i.e. between quarks and/or gluons.

Weak Interaction (W^+, W^- and Z^0 massive bosons) Between particle with weak hypercharge, i.e. between leptons and/or $W^{\pm}$, Z^0 bosons as well as quarks.

Electromagnetic Interaction (1 massless photon γ) Between all electrically charged particles.

1.1.2 The Standard Model Theory

The Standard Model[3] is a Quantum Field Theory based on local gauge symmetry to introduce the fundamental interactions (for this reason we usually refer to the force carriers as **gauge bosons**), substituting ordinary derivatives in fermionic Lagrangians with the so called "covariant derivatives", to keep the Lagrangians invariant under local gauge theory:

$$\partial_\mu \rightarrow \mathcal{D}_\mu = \partial_\mu + i\frac{g}{2}A_\mu \tag{1.2}$$

Where A_μ is the generator of the symmetry. Starting from the local gauge symmetry $SU(2)_L \otimes U(1)_Y$ S.L. Glashow[4], A. Salam[5] and S. Weinberg[6] defined in the theory the gauge bosons W^+, W^- and Z^0 for the weak interaction and γ for the electromagnetic interaction, so those two forces are actually two manifestations of the same fundamental interaction called **Electroweak** interaction. However, the obtained electroweak interaction lagrangians do not contain mass terms for the fermions and for the gauge bosons W^+, W^- and Z^0, in striking contrast with what is observed experimentally; a mass term artificially included in such lagrangian densities would break the $SU(2)_L \otimes U(1)_Y$ gauge symmetry. 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 $SU(2)_L \otimes U(1)_Y$ local gauge transformations and with a degenerate fundamental state, combining gauge theory together with spontaneous symmetry

[1] Gravity is not included in the Standard Model.

breaking ("Higgs Mechanism"[7]). This brings to mass terms for the weak gauge bosons[2] naturally included in theory, but with the introduction of a new particle in the Standard Model, the **Higgs boson**. This is the last particle of the Standard Model not yet experimentally observed.

The gauge symmetry of the Standard Model can be extended to $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ including the strong interactions related to the $SU(3)_c$ *colour* symmetry and mediated by eight coloured massless "gauge bosons", the gluons. The presence of the SU(3) symmetry group in the strong interaction brings to have some peculiarities which are not present in the others fundamental interaction. In the following section, strong interaction is presented with the most relevant peculiarities which explain how hadrons are produced.

1.2 The Strong Interaction

The quark model was proposed by Gell-Mann[8] in 1964. The idea of the "colour" quantum number was introduced by Han and Nambu[9] in 1965 to avoid the apparent paradox that the quark model seemed to require a violation of the Pauli exclusion principle in order to describe the hadron spectroscopy. Quantum Chromo Dynamics (QCD, the strong interaction sector of the standard model) was then quantized as a gauge theory with $SU(3)_c$ symmetry in 1973 by Fritzsch[10], Gross and Wilczec[11], Weinberg[12].

1.2.1 The QCD Lagrangian

Quantum Chromodynamics is introduced in the Standard Model as a SU(3) gauge theory and it's defined by its Lagrange density (1.3)

$$\begin{aligned}\mathcal{L}^{QCD} &= \mathcal{L}_{inv} + \mathcal{L}_{gaugefix} + \mathcal{L}_{ghosts} = \\ &\sum_f \bar{\Psi}_f (i\gamma_\mu D^\mu - m_f)\Psi_f - \frac{1}{4} F^a_{\mu\nu} F^{a,\mu\nu} - \\ &- \frac{1}{2}\sum_a (\partial_\mu A^\mu_a)^2 + \\ &+ (\partial_\mu \bar{c}_a)(\partial_\mu \delta_{ad} - g C_{abd} A^\mu_b) c_d\end{aligned} \tag{1.3}$$

the $\mathcal{L}_{inv}$ is the invariant term (where f runs over the six quarks fields - "flavours") which is obtained by the introduction of the covariant derivative $D^\mu = \partial^\mu + igA^\mu_a T_a$ and of the kinetic term $F^{\mu\nu}_a = \partial_\mu A_{a,\nu} - \partial_\nu A_{a,\mu} - g C_{abc} A^\mu_b A^\nu_c$. Here, A^μ_a are the 8 gluon fields, T_a the

[2]Spontaneous symmetry breaking is made in such a way that only the $SU(2)_L$ symmetry is broken, leaving $U(1)_Y$ symmetry unchanged, so photon remains massless as known.

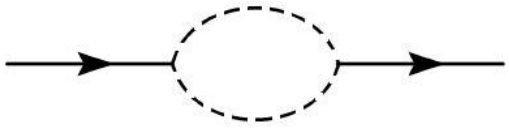

Figure 1.1: *A simple example of a Feynman diagram with a "loop". Lines describe fermions and dashed lines describes bosons.*

8 SU(3) generators, g is the "QCD" strong coupling[3] and C_{abc} are the structure constants of SU(3) which define its Lie Algebra[4]. Being SU(3) a non-abelian group, terms of self-interaction of A^{μ}_{a} fields appear in the lagrangian, so gluons can be coupled (interaction between gluons), in spite of the case of photons.

$\mathcal{L}_{gaugefix}$ and $\mathcal{L}_{ghosts}$ are needed for technical reasons connected to how the quantization of the QCD lagrangian is performed. The first term come from the requirement of the covariance of the theory, and the presence of *ghost* fields c_a (which are not physical fields, because they anticommute, despite their spin) in the second term ensure that the gauge fix term doesn't spoil the unitarity of the theory.

One of the main property of QCD is that $\mathcal{L}^{QCD}$ is *Poincaré*, *parity*, *time reversal* and (hence) *charge conjugation* (also called *C-parity*) invariant. It is in addition invariant under $U(1)^6$ which implies individual flavour conservation.

As it was said above, the $SU(3)_c$ colour symmetry is strongly bound to the requirement that Pauli principle has not to be violated: this brings to the existence of only observable "colour singlet" bound states. This means that QCD non-abelian gauge theory allows a physical spectrum which consists of colour singlet states only. The simplest of these states have the quantum numbers of quark-antiquark pairs (*mesons*) or of three quarks (*baryons*), although other possibilities are not excluded.

1.2.2 Renormalization Scheme in QCD: Running of α_s, Asymptotic Freedom and Confinement

As many other 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 why a theory with an arbitrary parameter (the regularization parameter) can be predictive.

When treating Feynman diagrams with "loops" (Fig.1.1), integrals of the following

[3]Usually $\alpha_s = g^2/(4\pi)$ is referred to be the strong coupling.

[4]For SU(3) are valid the commutation rules $[T_a, T_b] = iC_{abc}T_c$, with $a, b, c = 1...8$

form are involved:

$$\Gamma_{un}(p) = \int \frac{d^4k}{(2\pi)^4} \frac{1}{(k^2 - M^2(p))^2} \tag{1.4}$$

where $M(p)$ is a function of the incoming momentum. The 1.5 is obviously divergent because it goes as $\sim \int \frac{dx}{x}$ (so it is said *logarithmic divergence*). The divergent integral is replaced by a finite one by limiting the momentum in the loop to a cutoff value μ. In this way the 1.5 becomes:

$$\Gamma_{un}(p) \longrightarrow \Gamma_{ren}(p,\mu) = -\frac{i}{(4\pi)^2} ln\Big(\frac{M^2(p)}{\mu^2}\Big) \tag{1.5}$$

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 $\mathcal{O}(p,\mu)$ (such as the production *cross section* of a process), where p is a set of momentum and μ is the parameter introduced in the renormalization approach used above. All we have to do is to measure $\mathcal{O}$, fix μ to whatever value we like. At this point, at the price of one measurement, we can predict $\mathcal{O}$ for every other value of p.

Let's consider a physical observable such as the cross section σ, whose value depends on a scale Q, e.g. the momentum transfer in a scattering process, and on α_s ($\sigma = \sigma(Q, \alpha_s)$). As pointed above, when we calculate the cross section using Feynman diagrams technique, we need to introduce a renormalization scale μ in integrals computation, to make theme calculable. In general, σ is a function of Q^2/μ^2. The requirement that σ does not depend on our choice of μ is equivalent to the following equation 1.6:

$$\mu \frac{d\sigma(Q^2/\mu^2, \alpha_s)}{d\mu^2} = 0 \tag{1.6}$$

It can be shown that Eq. 1.6 can be satisfied if the whole scale dependence of the observable σ is embedded in the running value of the coupling constant, i.e. $\alpha_s = \alpha_s(Q^2)$, and its expression is of the form[13]

$$\frac{\alpha_s(Q^2)}{4\pi} = \frac{1}{\beta_1 ln(Q^2/\Lambda_{QCD})} - \frac{\beta_2 ln(ln(Q^2/\Lambda_{QCD}))}{\beta_1^3 ln^2(ln(Q^2/\Lambda_{QCD}))} + \mathcal{O}\Big(\frac{1}{ln^3(ln(Q^2/\Lambda_{QCD}))}\Big) \tag{1.7}$$

with β_i are the so called "beta function" of the QCD, as reported in [13], and Λ_{QCD} ($\sim$200 MeV) is often regarded as the parameter at which the interaction becomes "strong" (i.e. it becomes $\alpha_s(\Lambda_{QCD}) \simeq 1$). Indeed, from the 1.7 it can be shown that

$$\frac{d\alpha_s(Q^2)}{dQ^2} < 0 \tag{1.8}$$

which means that the value of the coupling constant decreases at higher scales, as shown also in Fig.1.2.

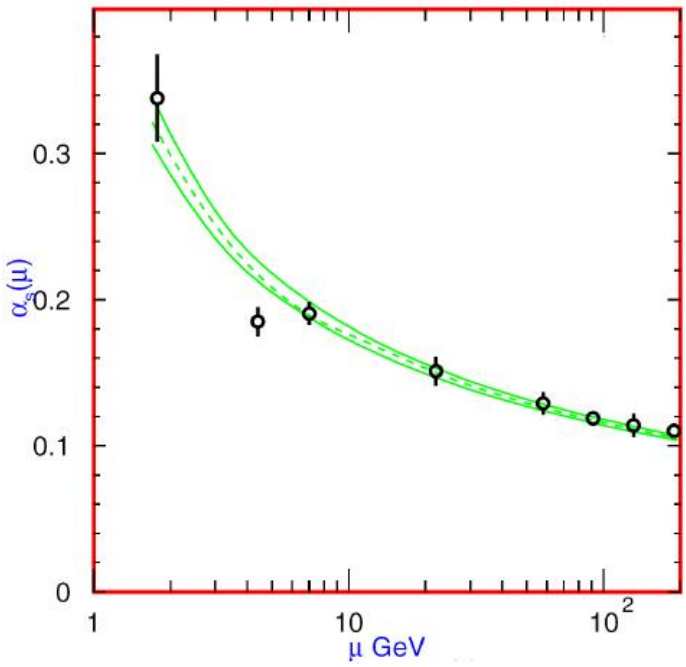

Figure 1.2: *Running of α_s[14]. α_s values are plotted as a function of the scale at which they were measures. Dotted line is the best fit. $\pm 1\sigma$ solid lines are also shown.*

This is a consequence of the non-abelian nature of the theory and it is one of the most peculiar characteristics of QCD with respect to Quantum ElectroDynamic (QED, the electromagnetic interaction sector of the standard model), where (being this an abelian theory) 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 fermionic 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. Indeed, the cross section of a process $\sigma(Q, \alpha_s))$ can be expanded in series of $\mathcal{O}(\alpha_s)$ as follows

$$\sigma = A_1\alpha_s + A_2\alpha_s^2 + \mathcal{O}(\alpha_s^3) \tag{1.9}$$

The coefficients A_i come from calculating the appropriate Feynman diagrams as referred in Fig.1.3. The renormalization scheme followed above ensures that observables such as cross section computed to all orders in perturbation theory do not depend on the renormalization scheme, i.e. by the μ parameter (called *renormalization parameter*). It follows that a truncated series *does* exhibit a μ dependence. In practice, QCD cross

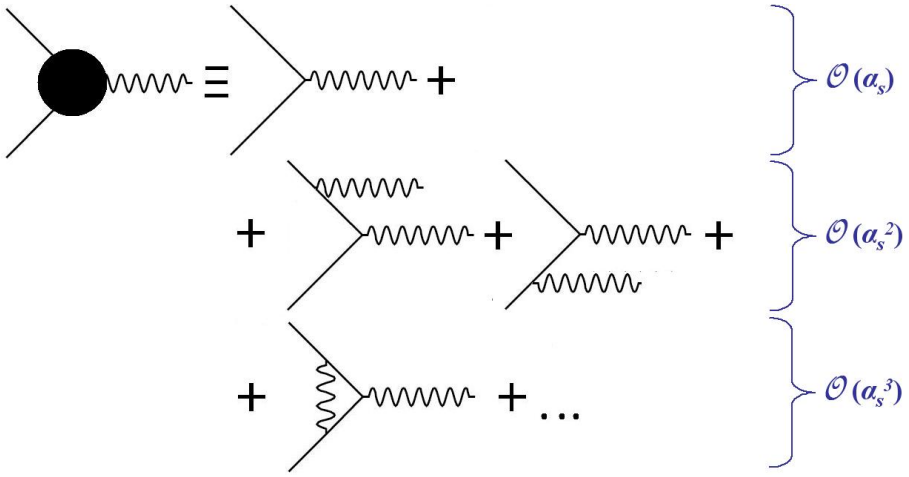

Figure 1.3: *Computation of A_i coefficients in the perturbation series using Feynman diagrams.*

sections are known to **leading order** (LO) of the series, or to **next to leading order** (NLO), or in same cases to **next to next to leading order** (NNLO), and it is only in the latter two cases, which have reduced renormalization scheme dependence, that are useful for precision tests.

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 only quark bound states can be observed, no 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. This is bound to a non-perturbative effect of the theory which cannot well predicted from pure theory and need instead experimental inputs; it will be better discussed in the following chapters.

1.2.3 Factorization theorems: Parton density functions and DGLAP equations

Consider a $2 \longrightarrow N$ process at an hadronic collider. In the collision, interactions take place between only two *partons* of the colliding hadrons (the **valence quarks** constituting the hadrons and **quarks and gluon Dirac sea** presents inside the hadron). It can be shown that the hard perturbative interaction between two interacting partons can be factorized out from the non-perturbative long-distance terms which describe the partons evolutions, i.e. that this latter terms does not depend on the scale Q of the interaction, but only on the fraction x (called *Bjorken factor*) of Q carried by the parton.

So the cross section of this process (i.e. in pp collisions such as in LHC) can be expressed

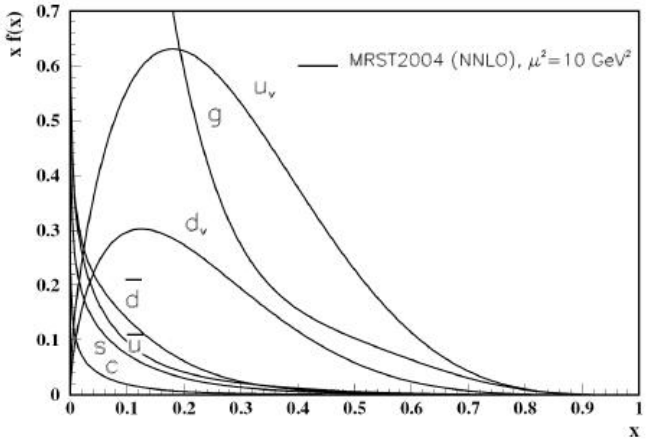

Figure 1.4: *Distribution of xf(x) in a proton as a function of the momentum fraction x at* $\mu_F^2 = 10$ *GeV*^{2} *for different partons[13].*

as the convolution of two factorized terms as follows:

$$d\sigma_{pp\to N}(Q^2,\mu_F^2) = \int_0^1 dx_1 \int_0^1 dx_2 f_1(x_1,\mu_F^2) f_2(x_2,\mu_F^2) d\hat{\sigma}_{pp\to N}(Q^2,\mu_F^2) \tag{1.10}$$

where $\hat{\sigma}$ 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, f_2 are the *parton density functions* (**PDF**s), that describe the probability that a parton carries momentum fraction x_1, x_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 partons takes place. The PDFs for quarks and gluons in a proton at a scale $\mu_F^2 = 10$ GeV^2 scale as calculated by the MRST[14] collaboration in 2004 are shown in Fig.1.4.

Everything in the example just described was considered for one fixed Q^2. A remarkable consequence of factorization theorem is that measuring parton distributions for one scale $Q = \mu$ allows their prediction for any other scale μ', as long as both μ and μ' are large enough that both $\alpha_s(\mu)$ and $\alpha_s(\mu')$ are small. This result, called the **evolution** of structure functions, increases the power of pQCD enormously. Thus, for instance, measuring the pdf of a parton i (where i can be a quark q or a gluon g) at a scale μ ($f_i(x,\mu)$) is enough to predict it at a new scale μ' ($f_i(x,\mu')$). The evolution of parton distributions is described by the DGLAP equations[15]:

$$\mu^2 \frac{d}{d\mu^2} f_i(x,\mu,\mu_F^2) = \int_x^1 \frac{d\xi}{\xi} P_{ij}\Big(\frac{x}{\xi}, \alpha_s(\mu_F^2)\Big) f_j(\xi,\mu,\mu_F^2) \tag{1.11}$$

This is an integral-differential equation, so it needs of the boundary condition to be solved, i.e. $f_i(x,Q)$ at a fixed scale $Q = \mu$ which can be measured experimentally. The sum is

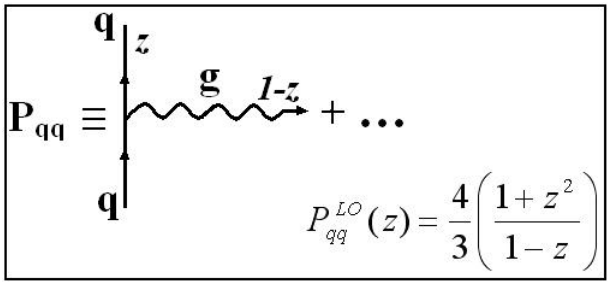

Figure 1.5: *Scheme of evolution kernel P_{qq} and its expression obtained at the LO in pQCD.*

implied when subscripts are the same, so the P_{ij} in the 1.11 is a matrix where i, j =q,g. These matrix elements are known as **splitting functions** or **evolution kernels**, and physically they are the probability that parton i with energy scale μ evolves to the energy scale μ', maintaining a fraction ξ of its momentum, emitting a parton j with a fraction $1-\xi$ of the parton i momentum. In Fig. 1.5 there is a scheme of the meaning of the evolution kernel Pqq.

As a consequence of factorization theorem, pdf evolution does not depend from perturbative effects, once the perturbative order is fixed when the pdf at a scale μ is evaluated. P_{ij} can be expressed as a perturbative series

$$P_{ij}(z, \alpha_s) = \frac{\alpha_s}{2\pi} P_{ij}^{LO}(z) + \left(\frac{\alpha_s}{2\pi}\right)^2 P_{ij}^{NLO}(z) + ... \tag{1.12}$$

and it can be computed once the order in α_s is fixed from the DGLAP equation. E.g. at LO it can be shown that[16]

$$\mu^2 \frac{d}{d\mu^2} f_i(x, \alpha_s(\mu^2)) = P_{ij}(x) + \mathcal{O}(\alpha_s^2) \tag{1.13}$$

1.3 Heavy Quarkonia & Prompt J/Ψ Production at LHC

1.3.1 The Colour Singlet Model (CSM)

A *quarkonium* is a quark-antiquark bound state (since the high mass, top quarks do not form bound states). Quarks with a mass higher than Λ_{QCD}, nominally the *charm* (c, ∼1.4 GeV), *bottom* (b, ∼4.5 GeV) and *top* (t, ∼175 GeV) quarks, are called "heavy". *Heavy Quarkonia* are the bound states $Q\bar{Q}$, where Q is either a c or b quark ($t\bar{t}$ pairs do not form bound states).

Heavy quarkonia can be produced in QCD interactions mainly via quark-pairs annihilation, gluon-gluon fusion, or c quark fragmentation diagrams (Fig.1.6).

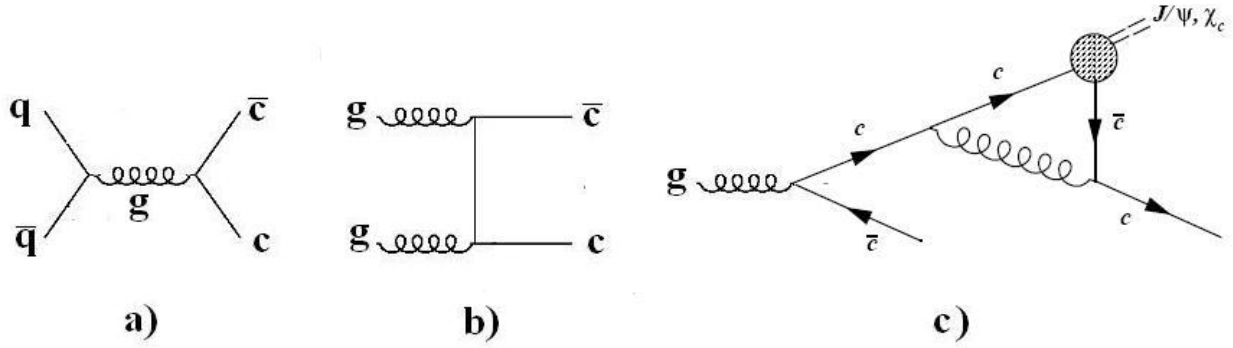

Figure 1.6: *a) quarks pairs annihilation, b) gluon-gluon fusion in pp→J/Ψ+g (additional g have to be emitted in the non-perturbative evolution into J/Ψ as a consequence of C-parity conservation, see Fig. 1.8) events and c) c quark fragmentation in pp→J/Ψ+cc̄ events.*

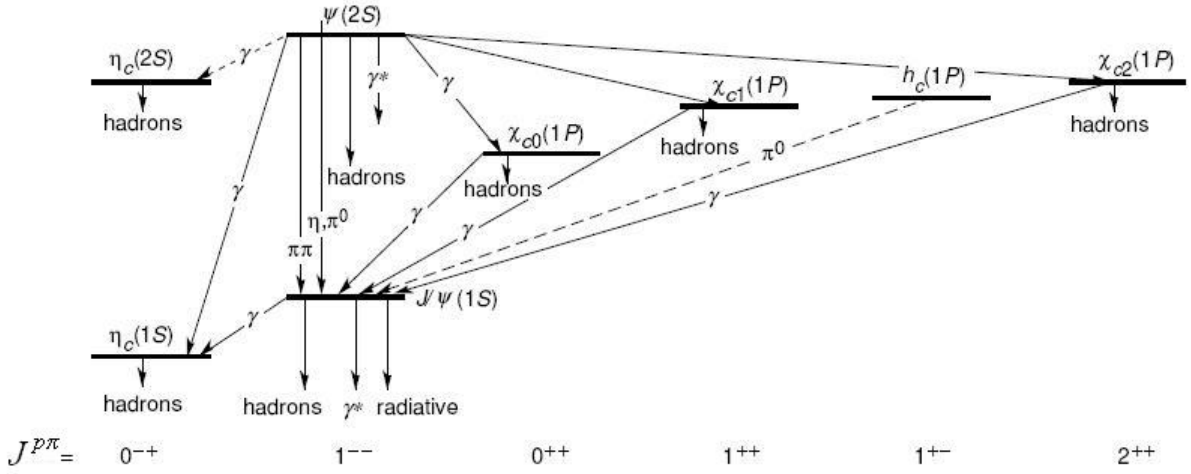

Figure 1.7: *Spectrum and transitions of the charmonium family.*

At LHC collisions will be between protons, and the second diagram will be more probable than the others because at the LHC energies the gluon PDF is the highest, making the gluons dominant partons in a proton. In order to form a bound state, the $Q\bar{Q}$ pairs have to be in colour singlet state and, due to the conservation of quantum numbers in QCD, produced with the same Spin (S), Orbital Angular Momentum (L) and Total Angular Momentum (J) of the final observable state. In the case of J/Ψ, this implies that only a subset of diagrams for the $c\bar{c}$ production can contribute at LO to the final state: no "charmonia" state can be directly produced from a s-channel gluon because of colour neutrality without extra legs (see c) diagram in Fig. 1.6). Moreover, because of C-parity (π operator) conservation in QCD, the creation of a J/Ψ ($J^{p\pi} = 1^{--}$, see Fig. 1.7) from simple gluon-gluon fusion at LO is forbidden. Therefore, the LO diagrams are $\mathcal{O}(\alpha_s^3)$ for

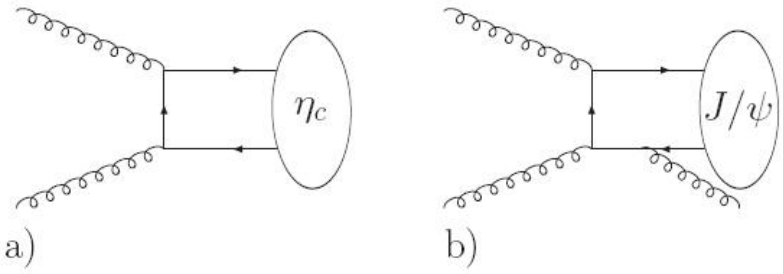

Figure 1.8: *a) LO diagrams in CSM for the production of η_c and b) J/Ψ. In b) case, an additional gluon emission ($\mathcal{O}(\alpha_s^3)$) is needed as a consequence of C-parity conservation.*

the production of the J/Ψ, and $\mathcal{O}(\alpha_s^2)$ for the other $c\bar{c}$(1S) bound state, the η_c (Fig.1.8). In this way, Q-pairs are created at partonic level with the same quantum numbers $^{2S+1}L_J$ (using the spectroscopy notation) of the hadron H and cross section ($\sigma(Q\bar{Q}(^{2S+1}L_J))$) can be computed at a fixed order of pQCD (perturbative approach can be used, given that $\alpha_s(m_c) \simeq 0.25$), using Feynman diagrams rules for the Matrix Element (ME) computation.

Once the pair was created with the right quantum numbers, it can evolve non-perturbatively into a bound state. Assuming that those two steps take place at two different scales (short distance and long distance respectively), the production cross section of H ($\sigma(H(^{2S+1}L_J))$) can be factorized in the product of the cross section perturbatively computed and the probability that $Q\bar{Q}$ pair with correct quantum numbers will bind to form the H bound state. This probability is given by the square of the l-th derivative of the radial Q wave-function at the origin (where l is the orbital angular momentum of the $Q\bar{Q}$), so

$$\sigma(H(^{2S+1}L_J)) = \sigma(Q\bar{Q}(^{2S+1}L_J)) \cdot |\frac{d^l}{dr^l} R_{nl}(0)|^2 \tag{1.14}$$

The wave-function R_{nl} can be obtained from the Schrödinger equation, of course assuming a known form for the potential between quarks. There are many models which allow to theoretically obtain those values, and the most likely one (basing on spectroscopy studies of mesons) is the *Cornell potential*[17], which assumes a potential of the form

$$V(r) = -\frac{4}{3}\frac{\alpha_s(1/r^2)}{r} + K^2 r \tag{1.15}$$

It can be also measured experimentally by the leptonic decay width of Quarkonia states[18].

In conclusion, Quarkonia states like J/Ψ are produced **directly** via strong interaction. The QCD can be used to obtain the final partons states by perturbative computation, then the formation of the hadrons (bound states) is a non-perturbative process which has to be factorized apart.

Heavy quarks pairs have to be produced with the same quantum number of the hadronized final state (since the strong interaction preserves the quantum numbers) and, in *singlet colour states*. For this reason, usually physicist refers to the phenomenology of this approach as **Colour Singlet Model** (CSM).
The non-perturbative process is not predictable by pQCD. The bound state wave-function and their first derivatives can be considered as *free parameters* of the CSM, which can be experimentally measured.

1.3.2 A non-relativistic treatment of QCD: the NRQCD

The CSM was considered successfully for many years, until 1995, when data from the Tevatron collider showed that CSM under-estimates the cross section for prompt J/Ψ production in p-$\bar{\text{p}}$ collisions by more than an order of magnitude[19]. Several other approaches have been developed[20][21][22] starting already from 80's, and have been tested with Tevatron data. Probably, the most compatible has been the **NRQCD** approach[21].
The NRQCD represents a non-relativistic treatment of QCD. In both heavy-quarkonium decays and hard-scattering production, large energy-momentum scales appear. The heavy-quark mass m_Q is much larger than Λ_{QCD}, and, in the case of production, the transverse momentum p_T can be much larger than Λ_{QCD} as well. Thus, the associated values of α_s are much less than one ($\alpha_s(m_c) \sim 0.25$ and $\alpha_s(m_b) \sim 0.18$). Therefore, one might hope that it would be possible to calculate the rates for heavy quarkonium production and decay accurately in perturbation theory. However, there are clearly low-momentum, nonperturbative effects associated with the dynamics of the quarkonium bound state that invalidate the direct application of perturbation theory. So, in order to make use of perturbative methods, one must first separate the short-distance/high-momentum perturbative effects from the long-distance/low-momentum nonperturbative effects, involving a procedure similar to usual factorization in full QCD. NRQCD reproduces full QCD accurately at momentum scales of order $m_Q v$ and smaller, where v is heavy-quark velocity in the bound state in the center-of-mass (CM) frame ($v^2 \sim 0.3$ for charmonium and $v^2 \sim 0.1$ for bottomonium). The low velocities of heavy-quarks in bound states production suggest a non-relativistic approach of the model. Virtual processes involving momentum scales of order m_Q and larger can affect the lower-momentum processes. Their effects are taken into account through the short-distance coefficients of the operators that appear in the NRQCD action.

The non-relativistic Lagrangian of QCD

Starting from the full QCD Lagrangian (Eq. 1.3), it can be expanded non-relativisticly integrating out modes of energy m_Q. On the other hand, NRQCD is designed to describe $Q\bar{Q}$ at energies scale ($\sim \mathcal{O}(m_Q v)$) much smaller than their mass (since $v << 1$), so new quark-antiquark pairs cannot be created and it is enough to use Pauli spinors instead of Dirac fields for both Q and $\bar{Q}$, simplifying the computation. The QCD Lagrangian (Eq. 1.3) can be expanded as series of $\mathcal{O}(1/m_Q)$. The high-energy modes that have been integrated out have a relevant effect on the low-energy physics. This effect is not lost, but encoded into the so called *matching coefficients* (c) in the NRQCD Lagrangian. In principle, there are infinite such terms to be included, in practice only few of them are needed. Each operator can be counted in v. The velocity v and α_s are the two small expansion parameters of NRQCD. If we aim at an accuracy of order $\alpha_s^k v^n$ we have to keep in the Lagrangian only terms and matching coefficients that contribute up to that order to the physical observable under study. The NRQCD couplings m, g, c are determined by the requirement that NRQCD reproduces the results of QCD up to order $\alpha_s^k v^n$[23].

Up to field redefinitions, the NRQCD Lagrangian for one heavy flavour of mass m at $\mathcal{O}(1/m^2)$, including the kinetic energy term $p^4/(8m^3)$ in the non-relativistic expansion, reads[24]:

$$
\begin{aligned}
\mathcal{L}^{NRQCD} &= \mathcal{L}_g + \mathcal{L}_\Psi + \mathcal{L}_\chi + \mathcal{L}_{\Psi\chi},\\
\mathcal{L}_g &= -\frac{1}{4}F^{\mu\nu a}F^a_{\mu\nu} + c_1^g\frac{1}{4m^2}g f_{abc}F^a_{\mu\nu}F^{\mu b}_{\alpha}F^{\nu\alpha c},\\
\mathcal{L}_\Psi &= \Psi^\dagger\Bigg[iD_0 + c_2\frac{|\vec{D}|^2}{2m} + c_3\frac{|\vec{D}|^4}{8m^3} + c_F g\frac{\vec{\sigma}\cdot\vec{B}}{2m} +\\
&\quad + c_D g\frac{\vec{D}\cdot\vec{E} - \vec{E}\cdot\vec{D}}{8m^2} + ic_s g\frac{\vec{\sigma}\cdot(\vec{D}\times\vec{E} - \vec{E}\times\vec{D})}{8m^2}\Bigg]\Psi,\\
\mathcal{L}_\chi &= charge \quad coniugate \quad of \quad \mathcal{L}_\Psi,\\
\mathcal{L}_{\Psi\chi} &= \frac{c_{f1}}{m^2}\Psi^\dagger\chi\chi^\dagger\Psi + \frac{c_{f1}}{m^2}\Psi^\dagger\sigma\chi\chi^\dagger\sigma\Psi + \frac{c_{f8}}{m^2}\Psi^\dagger T^a\chi\chi^\dagger T^a\Psi + \frac{c_{f8}}{m^2}\Psi^\dagger T^a\sigma\chi\chi^\dagger T^a\sigma\Psi \equiv\\
&\equiv \frac{c_{f1}}{m^2}O_1(^1S_0) + \frac{c_{f1}}{m^2}O_1(^3S_1) + \frac{c_{f8}}{m^2}O_8(^1S_0) + \frac{c_{f8}}{m^2}O_8(^3S_1)
\end{aligned}
\tag{1.16}
$$

where Ψ is the Pauli spinor that annihilates the quark, χ is the Pauli spinor that creates the antiquark and $\vec{\sigma}$ are the Pauli spin matrices. In 1.16, $\mathcal{L}_g$ describes gluon kinematic terms, $\mathcal{L}_\Psi$ and $\mathcal{L}_\chi$ describes $Q\bar{Q}$ production/annihilation and $\mathcal{L}_{\Psi\chi}$ describes the quantum numbers (spin, orbital and total angular momentum, and colour) of $Q\bar{Q}$ pairs in non-perturbative regime. The covariant derivative D^μ is splitted in a scalar term ($iD_0 = i\partial_0 - gA_0^aT^a$) and a three-vector term ($i\vec{D} = i\vec{\nabla} + g\vec{A}^aT^a$). $\vec{E}$ (defined as $E^i = F^{i0a}Ta$) and $\vec{B}$ (defined as $B^i = -\epsilon_{ijk}F^{jka}T^a/2$) are the chromoelectric and chromomagnetic fields,

and they are so called due to the analogies in the non-relativistic approach of QED[25]. $O_1(^1S_0)$, $O_1(^3S_1)$, $O_8(^1S_0)$ and $O_8(^3S_1)$ are the so called **four-fermions operators**; the indeces $_1$ and $_8$ indicate that pairs are in colour singlet or colour octet states (due to presence of T^a generators). The matching coefficients c_i may be calculated in perturbation theory requiring match of results with full QCD.

NRQCD as a Quantum Effective Field Theory

Since several scales remain dynamical in NRQCD, it is not possible to give a homogeneous power counting for each operator without extra assumptions, i.e. the power counting in v is not unambiguous. The NRQCD Lagrangian is organized in operators of increasing dimension, hence, it is in general non-renormalizable in the usual sense. In spite of this, NRQCD could be thought as an *Effective Field Theory* (EFT), which provide a realization of Wilson renormalization group[26]. An EFT is a quantum field theory with the following properties: a) it contains the relevant degrees of freedom to describe phenomenology in certain limited range of energies and b) it contains an intrinsic energy scale Λ (e.g., m in NRQCD) that sets the limit of applicability of the EFT. In this sense, NRQCD Lagrangian can be made finite to any finite order in $1/\Lambda$ by renormalizing the matching coefficients in front of the operators in the Lagrangian until that order. This means that one needs more renormalization conditions when the order in $1/\Lambda$ is increased.

Factorization in NRQCD: COM and CSM

The formalism in Eq. 1.16 shows that NRQCD is based upon a double power series expansion in the strong interaction fine structure constant α_s and the velocity parameter v. Heavy quarkonia are described within the NRQCD framework in terms of Fock state decomposition of the ortho-quarkonium vector meson wave-function as described in 1.17

$$\begin{aligned}|\Psi_Q>=&\mathcal{O}(1)|Q\bar{Q}_{(1)}[^3S_1]>+\\&+\mathcal{O}(v)|Q\bar{Q}_{(8)}[^3P_J]g>+\\&+\mathcal{O}(v^2)|Q\bar{Q}_{(8)}[^1S_0]g>+\mathcal{O}(v^2)|Q\bar{Q}_{(1,8)}[^3S_1]gg>+\mathcal{O}(v^2)|Q\bar{Q}_{(1,8)}[^3D_J]gg>+\\&+...\end{aligned} \tag{1.17}$$

The spin, orbital and total angular momentum quantum numbers of the $Q\bar{Q}$ pairs in each Fock component are indicated in square brackets in spectroscopic notation, while the colour assignment is specified by singlet or octet indeces.

If the relative importance of various production channels depend solely upon the order in v at which pairs hadronize into physical bound states, those modes which proceed through

the leading Fock component in quarkonia wave-functions would generally be dominant. So we retrieve (first row in the 1.17) the basic assumption of the previously quoted CSM, where heavy quarkonium production is presumed to be mediated by parton interactions that generate colourless $Q\bar{Q}$ pairs with the same quantum numbers as the mesons into which they non-perturbatively evolve[27]. Anyway, the breakdown of this kind of approach stems from its neglect of all high energy processes that create $Q\bar{Q}$ pairs with quantum numbers different from those of all final state mesons. So, as considering higher order in v, also colour octet bound states can be produced in association with hard gluons (g, gg, etc.) which then evolve non-perturbatively in observable colour singlet states. In this sense, NRQCD includes also production in the so called *Colour Octet Model* (COM).

Starting from the formalism in Eq. 1.16, the general expression for a production cross section of a heavy quarkonium state H can be written as

$$d\sigma(1+2\rightarrow H)=\sum d\hat{\sigma}(1+2\rightarrow Q\bar{Q}[^{2S+1}L_J(1,8)])\langle O_{1,8}(^{2S+1}L_J)\rangle \tag{1.18}$$

Here the sum is made over all possible quantum states (spin, angular momentum and colour) of the produced $Q\bar{Q}$ pair. $Q\bar{Q}$ states with quantum numbers other than H arise from the expansion of the H Fock state in powers of the heavy quark velocity v. The first important thing is that the 1.18 includes itself the "factorization" between the perturbative short-distance term ($d\hat{\sigma}(1+2\rightarrow Q\bar{Q}[^{2S+1}L_J(1,8)])$) and the non-perturbative long-distance evolution of the quarkonium state ($\langle O_{1,8}(^{2S+1}L_J)\rangle$).

The short-distance terms describes the production of the $Q\bar{Q}$ pairs in the interaction between parton 1 and parton 2 in a certain quantum state and they are, in practice, the cross-sections which can be computed as in pQCD using α_s as a perturbative parameter, basing on Feynmann diagrams in Fig. 1.6 if production at LHC is considered. They include also informations of the incoming partons PDFs also described in 1.10. The long-distance terms are the *vacuum expectation values of the NRQCD four-fermions operators* and they represent the probability that the formed $Q\bar{Q}$ pair evolves non-perturbatively in the final H meson. Except for pairs produced in CSM (where long-distance terms can be computed basing upon radial wave-functions and their firsts derivatives), the computation of the vacuum expectation values of the four-fermions operators is very difficult, so they constitute free parameters of the theory and have to be tuned from experimental data. They are also known as **NRQCD matrix elements**, which will be better discussed in the following chapters. Even if they represent free parameters of the theory, there are simplifying relations between matrix elements (obtained from the heavy quark spin symmetry[28]), that reduce the number of independent matrix elements. For example,

$c\bar{c}$	$b\bar{b}$
$\langle O_1^{J/\Psi}(^3S_1)\rangle$	$\langle O_1^{\Upsilon}(^3S_1)\rangle$
$\langle O_8^{J/\Psi}(^3S_1)\rangle$	$\langle O_8^{\Upsilon}(^3S_1)\rangle$
$\langle O_8^{J/\Psi}(^1S_0)\rangle$	$\langle O_8^{\Upsilon}(^1S_0)\rangle$
$\frac{\langle O_8^{J/\Psi}(^3P_0)\rangle}{m_c^2}$	$\frac{\langle O_8^{\Upsilon}(^3P_0)\rangle}{m_b^2}$
$\frac{\langle O_1^{\chi c0}(^3P_0)\rangle}{m_c^2}$	$\frac{\langle O_1^{\chi b0}(^3P_0)\rangle}{m_b^2}$

Table 1.1: *The 10 NRQCD matrix elements.*

matrix elements of the NRQCD CSM satisfy the 1.19

$$\begin{aligned}\langle O_1^{J/\Psi}(^3S_1)\rangle &= \langle O_1^{\eta_c}(^1S_0)\rangle,\\ \langle O_1^{\chi_{cj}}(^3P_j)\rangle &= \frac{1}{3}(2j+1)\langle O_1^{h_c}(^1P_1)\rangle\end{aligned} \tag{1.19}$$

And for the COM the following relations are valid:

$$\begin{aligned}\langle O_8^{J/\Psi}(^3S_1)\rangle &= 3\langle O_8^{\eta_c}(^1S_0)\rangle,\\ \langle O_8^{J/\Psi}(^1S_0)\rangle &= 3\langle O_8^{\eta_c}(^3S_1)\rangle,\\ \langle O_8^{J/\Psi}(^3P_j)\rangle &= \frac{1}{3}(2j+1)\langle O_8^{h_c}(^1P_1)\rangle,\\ \langle O_8^{\chi_{cj}}(^3S_1)\rangle &= \frac{1}{3}(2j+1)\langle O_8^{h_c}(^1S_0)\rangle\end{aligned} \tag{1.20}$$

In this way, NRQCD can be reduced to an EFT with **10 free parameters**, 5 referred to $c\bar{c}$ states and 5 refereed to $b\bar{b}$ states (see Tab. 1.1). The NRQCD CSM matrix elements are related with the non-perturbative coefficients of the CSM in the full QCD (Eq. 1.14) by the following relations

$$\begin{aligned}\langle O_1^{J/\Psi}(^3S_1)\rangle &= \frac{3N_c}{2\pi}|R(0)|^2,\\ \langle O_1^{\chi_{c0}}(^3P_0)\rangle &= \frac{3N_c}{2\pi}|R'(0)|^2\end{aligned} \tag{1.21}$$

where $N_c = 3$ is the *number of colour* in QCD.

The cornerstone of the predictive power of NRQCD is the "universality" of the NRQCD matrix elements: indeed, their values do not depend on the details of the hard process thanks to the factorization included in the NRQCD theory.

Infrared singularities which appear in some of the short-distance terms of P-wave states (where soft-collinear gluons can be emitted) can be absorbed into the long-distance part of colour octet S-wave terms (where soft gluons can be seen as emitted during non-perturbative hadronization), thereby ensuring a well defined overall result[24]. So a striking consequence of the NRQCD approach is that the effect of colour octet contributions

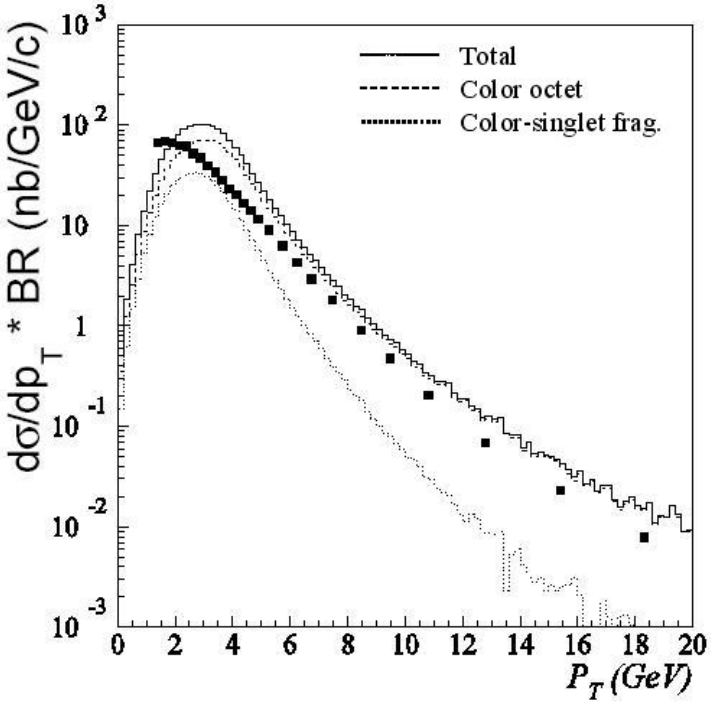

Figure 1.9: *Comparison between experimental data from the CDF experiment[19] (black dots) and theoretical predictions (continues lines).*

can be extremely important even in the case of S-wave production, even if their effects are predicted to be suppressed by powers of v with respect to the leading colour singlet ones.

1.3.3 Observables of the NRQCD Theory

For the experimental validation of the theoretical model physical observables (which can be experimentally detectable) have to be found and studied. There are three possibilities: **inclusive and p_T differential cross section of prompt quarkonium (J/Ψ) production** (enhanced for NRQCD), **prompt quarkonium polarization** and **hadronic activity** around the prompt quarkonium direction.

Differential Cross Section & p_T spectrum

Relation in Eq. 1.17 shows the possible quantum numbers of $Q\bar{Q}$ pairs during the quarkonium production via NRQCD, weighted by the order in v. In the second and in the third raw of Eq. 1.17 it can be seen that in COM production cases there is the associated production with gluons, which are usually low p_T[28]. Gluons emitted in CSM are expected to be less than the previous case, but have usually higher p_T, in order to change

the quantum numbers of the $Q\bar{Q}$ pair into the ones expected for the final hadron (i.e. the prompt quarkonium). Since these gluons carry the most part of the p_T, it is expected that prompt quarkonia produced via CSM are emitted with a softer p_T than the cases where the $Q\bar{Q}$ pair is produced in a colour-octet state.

According to the NRQCD factorization approach, the charmonium production cross-section contains not only the CSM terms, which are absolutely normalized, but also colour-octet terms, whose normalizations are determined by colour-octet matrix elements. In the case of J/Ψ production, the most important colour-octet matrix elements are $\langle O_8^{J/\Psi}(^3S_1)\rangle$, $\langle O_8^{J/\Psi}(^3P_0)\rangle$, and $\langle O_8^{J/\Psi}(^1S_0)$. In the only large p_T region, the J/Ψ cross sections are dominated by gluon fragmentation into colour-octet $^3S_1$ $c\bar{c}$ pairs, which falls as $\frac{d\sigma}{dp_T^2} \sim \frac{1}{p_T^4}$. The colour-octet $^1S_0$ and ^3P_J channels are significant in the region $p_T \leq 10GeV$, but fall as $\frac{d\sigma}{dp_T^2} \sim \frac{1}{p_T^6}$ and become negligible at large p_T. Because the $^1S_0^{(8)}$ and $^3P_J^{(8)}$ short-distance cross sections have a similar p_T dependence, the transverse momentum distribution is sensitive only to the linear combination $\sim \langle O_8^{J/\Psi}(^1S_0)\rangle + \frac{\langle O_8^{J/\Psi}(^3P_0)\rangle}{m_c^2}$.

Moreover, there is the contribution $\chi_c \rightarrow$J/Ψ γ, not-negligible in the colour-octet production case.

Indeed, this is what it can be found by computation via Monte Carlo techniques: in Fig. 1.9 it can be seen the p_T differential cross section ($\frac{d\sigma}{dp_T}$) of the production of prompt J/Ψ at the Tevatron theoretically expected for the only CSM and for production via NRQCD (CSM+COM), compared with data from the CDF experiment[29]. The theoretical cross section computation is made with `pythia` Monte Carlo (MC) generator[30]; this is the same Monte Carlo tool used in this work, so further informations about this program will be given in the next chapters (Chapter 3). The Fig. 1.9 shows that COM and CSM can be discriminated starting from prompt J/Ψ $p_T = 3GeV/c$. Indeed, there is at least one order of magnitude of difference between the two models expected cross sections. Experimental data from Tevatron seems to be better fitted to NRQCD COM+CSM than the only CSM from full QCD.

J/Ψ polarization spectrum

Another observable is the J/Ψ polarization. It's experimental determination is difficult, but systematics due to rate predictions and absolute trigger and detector efficiencies cancel out in the measurement of cross section ratios for different polarization states.

The angular distribution of the decay products of the quarkonium depends on the spin state of the quarkonium. The polarization of a $J^{p\pi} = 1^{--}$ state, such as the J/Ψ, can be measured from the angular distribution of its decays into lepton pairs. Let θ be the angle in the J/Ψ rest frame between the positive lepton momentum and the chosen polarization axis. The most convenient choice of polarization axis depends on the process (usually, the

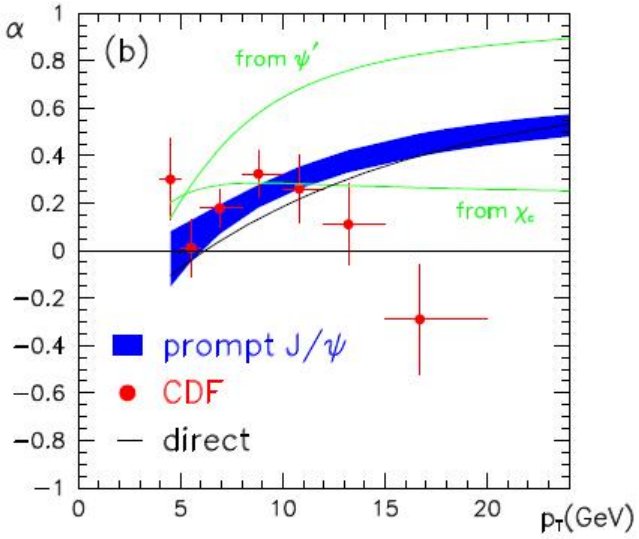

Figure 1.10: *Comparison between experimental data (red dots) and theoretical MC predictions (blue area). Green lines are diluting feed down effects due to* $\Psi(2S)$ *and* χ_{cJ}*.*

direction of the boost vector from the prompt J/Ψ rest frame to the centre-of-momentum frame of the colliding hadrons). The differential cross-section has the form

$$\frac{d\sigma}{d(\cos\theta)} \propto 1 + \alpha\cos^2\theta \tag{1.22}$$

which defines a polarization variable whose range is $-1 \leq \alpha \leq +1$. We can define not polarized, longitudinally and transversely polarized J/Ψ's to be ones whose spin components along the polarization axis are 0 and ± 1, respectively. The NRQCD factorization approach gives a simple prediction for the polarization variable α at very large transverse momentum. The production of a quarkonium with p_T that is much larger than the quarkonium mass is dominated by gluon fragmentation, where quarkonium is formed in the hadronization of a gluon that is created with even larger transverse momentum. From the Eq. 1.17 it can be seen that the dominant gluon-fragmentation process is gluon fragmentation into a $Q\bar{Q}$ pair in a colour-octet $^3S_1$ state. The fragmentation probability for this process is of order α_s, while the fragmentation probabilities for all other processes are of order α_s^2 or higher. The NRQCD matrix element for this fragmentation process is $\langle O_8^{J/\Psi}(^3S_1)\rangle$. At large p_T, the velocity scaling rules predict that the colour-octet $Q\bar{Q}$ state retains the transverse polarization as it evolves into an S-wave quarkonium state, up to corrections of relative order v^2 (see Eq. 1.17). Radiative corrections and colour-singlet production can dilute the quarkonium polarization somewhat. Feed down from χ_{cJ} states is about 30% of the J/Ψ sample and dilutes the polarization; feed down from the $\Psi(2S)$ is about 10% of the J/Ψ sample and is largely transversely polarized. So, despite these various diluting effects, a substantial polarization is expected at large p_T, and its detec-

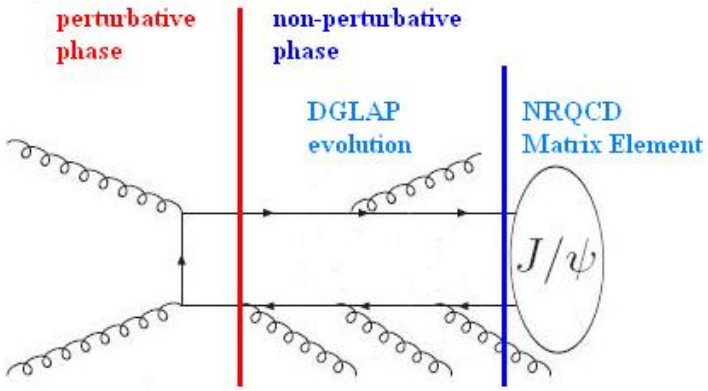

Figure 1.11: *NRQCD production of prompt J/Ψ in COM. After perturbative phase, DGLAP evolution of coloured cc̄ states allows a bigger emission of gluons than CSM case.*

tion would be a valid physical observable for the study of the colour-octet production mechanism.

The CDF measurement of the J/Ψ polarization[31] as a function of p_T is shown in Fig. 1.10, along with the NRQCD factorization prediction. The observed J/Ψ polarization is in agreement with the prediction, except for the highest p_T bin, even if the *prediction of increasing polarization with increasing p_T is not in evidence* (blue solid band in Fig. 1.10). On the other hand, HERA-B experiment at DESY measurements[28] in the J/Ψ p_T range 2÷5 GeV/c seems to be in agreement with theoretical prediction, consistent with $\alpha = 0$, and in contradiction with the CDF results in Fig. 1.10, which state $\alpha \simeq 0.3$ at J/Ψ $p_T = 5$ GeV/c.

hadronic activity around J/Ψ

Recently[32][33] another physical observable was proposed to study NRQCD. It predicts production in COM for the $Q\bar{Q}$ pairs. Colour-octet bound states are coloured states, so they evolve non-perturbatively until cc̄ states became J/Ψ states following the NRQCD matrix elements (Fig. 1.11). Technical issues on evolution of the octet-coloured state will be better explained in Chapter 3: the splitting kernels of the DGLAP equation are used to evaluate the probability of gluons emissions until the hadronization in a J/Ψ. This formalism is not related to the DGLAP equations in reality, but since their splitting kernels are used, this phase of the non-perturbative process is known as *DGLAP evolution*. What is important for now is that in the evolution of coloured states it is expected a certain amount of gluons emission. Theory is not able to say what is the correct approach to describe those emissions, anyway it is expected there is a bigger amount of emission in COM cases than in CSM cases; the measure of the amount of hadronic activity around the direction of the boost vector of the prompt J/Ψ can be a further observable to dis-

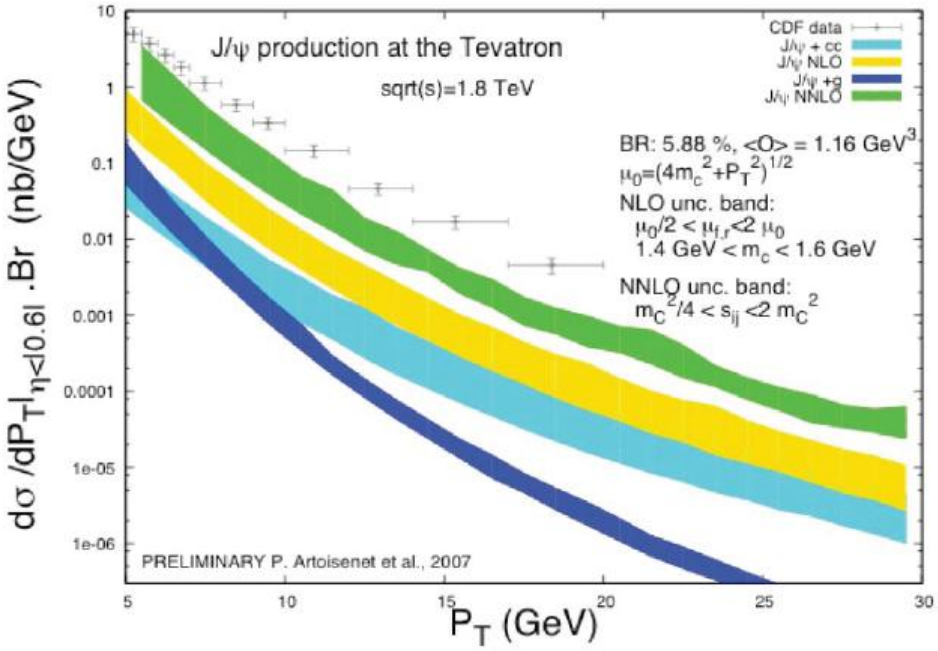

Figure 1.12: *Differential Cross Section as a function of p_T for prompt J/Ψ production at CDF ($\sqrt{s} = 1.8$ GeV) with overlaid results of CSM calculations (and theoretical error bands due to the μ_R and μ_F scales) up to NNLO. It result still have little gap between theoretical predictions and data[37].*

criminate CSM and COM and so to validate NRQCD theory. Further information on this observable will be given in chapter 3.

1.4 Summary on Prompt J/Ψ Physics

Analysis of CDF data at Tevatron on different observables brought to different results. Data on differential cross section measurements seems to be in agreement with theoretical predictions of the NRQCD with p_T enhanced COM $Q\bar{Q}$ pairs. On the other side, the expectation of increasing polarization with increasing p_T (as predicted in NRQCD) is not confirmed. Those results are in contrast. Other possibilities have been suggested: the colour-evaporation model (CEM)[34][35] predicts zero quarkonium polarization and, at the same time, an enhanced p_T spectrum for the prompt J/Ψ. Also computations at NNLO[36] for $c\bar{c}$ pairs in the CSM using Full QCD (see Fig. 1.12) seems to be more compatible with CDF data, without expecting any kind of polarization.

However, the experimental statistics are not as good as it is needed, so CDF data are not enough to discriminate the validity of the NRQCD. LHC at CERN will run at

higher energies and luminosities of Tevatron, so higher statistics can be reached starting yet from the firsts runs of the collider. This work is involved to the measure of one of those observables ($d\sigma/dp_T$) at the LHC with the CMS detector. All studies on quarkonium until today are based on the decay channel $J/\Psi \rightarrow \mu^+\mu^-$. An increment of the statistic and an independent channel for systematics studies can be given by the study in the decay channel $J/\Psi \rightarrow e^+e^-$. In the next chapters I will show how it is possible to make cross section measurements in this decay channel at LHC.

Chapter 2

The CMS experiment at LHC

The Large Hadron Collider (LHC)[38] is the most powerful hadron collider running in the next two decades. Its installation was completed in summer 2008. LHC will investigate processes with really tiny cross sections, down to the femtobarn.

There are two reasons that drove the choice of a hadron collider instead of an electron collider (like the "old" LEP[39]): first of all, 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, which is proportional to the fourth power of E/m. Compared to electrons, the energy loss is reduced by a factor of $\mathcal{O}(10^{12})$. Then, due to the composite nature of protons, hadron colliders provide access to a wider energy spectrum, which in addition can be explored simultaneously. This is a good issue for a machine involved in the discovery of "unknown" new physics.

In this chapter I will briefly review the main characteristics of the LHC and I will describe the LHC experiment this work is involved in, the Compact Muon Solenoid (CMS)[40].

2.1 The Large Hadron Collider at CERN

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[1]. The acceleration is performed in several stages. The proton injection starts at the duo-plasmatron, which is the proton source. A linear accelerator (LINAC) boosts the protons to energy of 750 KeV using Radio Frequency Quadruples. In case of the LHC a 30

[1]This is the "nominal" energy. At the start up energy will be lower.

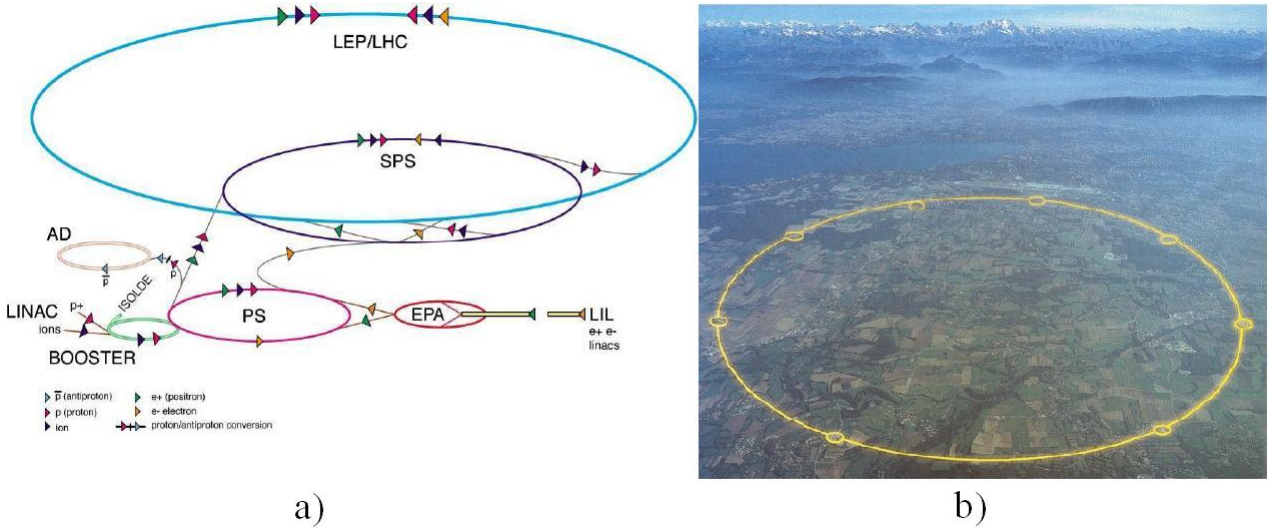

Figure 2.1: *a) A schematic view of the LHC accelerator complex and b) a panoramic picture of the LHC area near Geneva.*

μs pulse is ejected into the Proton Synchrotron Booster (PSB) which increases the energy to 1.4 GeV. The LHC bunch train with the 25 ns spacing starts in the proximate Proton Synchrotron (PS), in which the energy increases to 25 GeV. Then the protons will be accelerated up to 450 GeV by the Super Proton Synchrotron (SPS), and they will finally be injected into the LHC. In the LHC ring, the acceleration continues until the protons reach their nominal energy of 7 TeV. The two beams will collide in four interaction points; four experiments are built around the interaction points. Two general purpose experiments, called ATLAS[41] and CMS, will do general Standard Model measurements and will seek new physics; one experiment called LHCb[42] is dedicated to the B meson physics and it will carry out precise measurements of CP violation; one experiment called ALICE[43] will investigate heavy ion physics (Pb ions will be accelerated in a later phase of LHC operation). This collider will be able to investigate mass scales from order of few GeV, as 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 charged particle's momentum and the magnetic field is the 2.1:

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

where B is the magnetic field in Tesla, p the momentum in GeV and 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. The temperature of these dipoles is kept below 1.9 K by a superfluid helium cooling system.

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 (event/s) for a process with cross section σ is

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

There is an approximated formulæ (2.3) which allows to connect the luminosity with the beam properties, as it is shown in [44]:

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

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 transverse dimensions of the beam can be squeezed down to 15 μm.

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. Tab. 2.1 describes the main design characteristics of LHC. In Fig. 2.2 there are shown the production cross section of many Standard Model processes that may be produced at LHC energies.

In the hard proton proton collision, with high transferred momentum, the centre of mass energy $\sqrt{\hat{s}}$ of the fundamental interaction is connected to the total centre of mass energy $\sqrt{s}$ remembering the factorization theorem in Chapter 1 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 (*Bjorken factors*).

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 Lorentz boost invariant observables are 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

Circumference	26.659 km
Maximum dipole field	8.33 T
Magnetic Temperature	1.9 K
Beam energy at injection	450 GeV
Beam energy at collision	7 TeV
Beam energy at collision (start up)	3.5-5 TeV
Maximum Luminosity	10^{34} cm^{-2}s^{-1}
Luminosity at start up	$8{\cdot}10^{29} \div 1{\cdot}10^{31}$ cm^{-2}s^{-1}
Number of proton bunches	2808
Number of protons per bunches	$1.1{\cdot}10^{11}$
Number of protons per bunches at start up	$5{\cdot}10^{10}$
Bunch separation	24.95 ns
Bunch separation at start up	50 ns
Collisions frequency	40.08 MHz
Collisions frequency at start up	20 MHz
Crossing angle	300 μrad
Bunch Length (r.m.s.)	7.5 cm
Transverse beam size at impact point	15 μm
Luminosity lifetime	10 h
Filling time per ring	4.3 min
Energy loss per turn (at 14 TeV)	7 keV
Total radiated power per beam (at 14 TeV)	3.8 kW
Stored energy per beam (at 14 TeV)	350 MJ

Table 2.1: *Technical parameters of LHC for p-p collisions (values for* $^{82}_{208}Pb$-$^{82}_{208}Pb$ *collisions are available in [38]).*

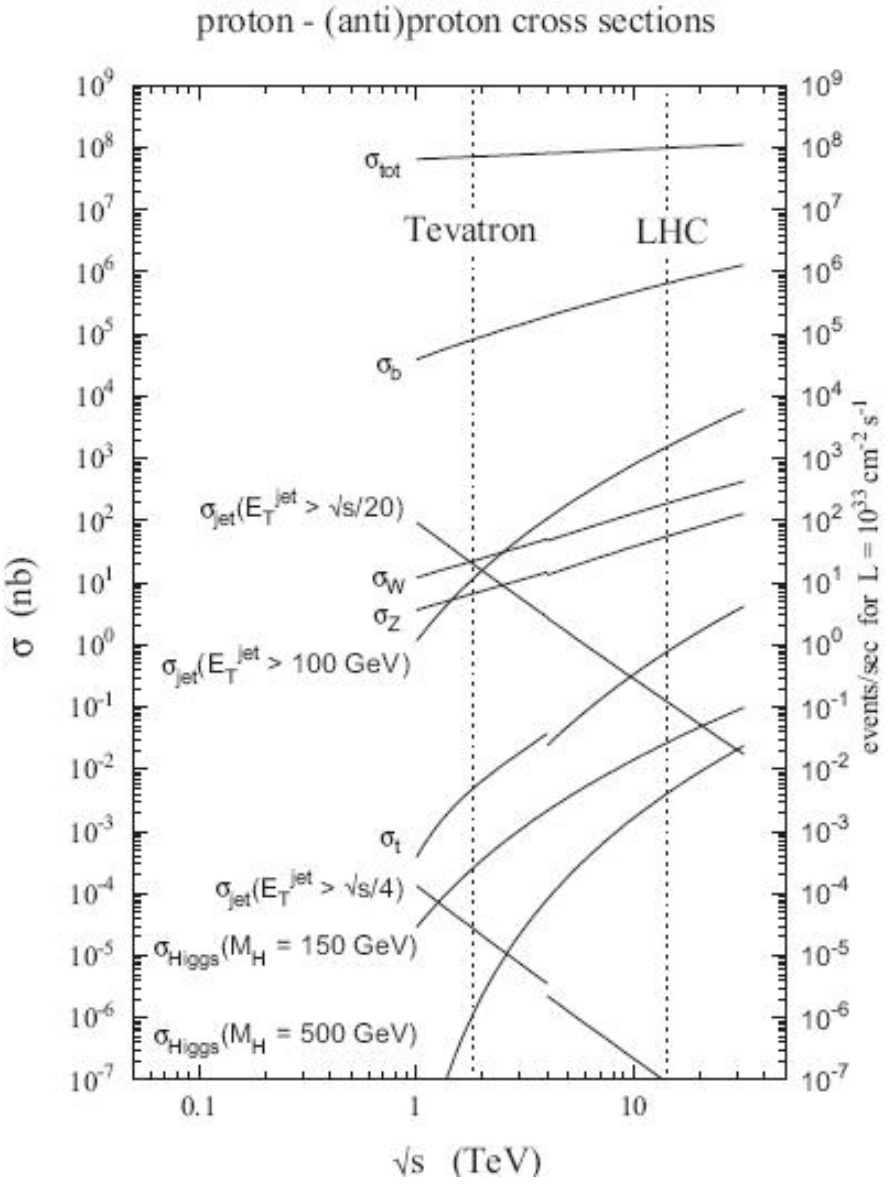

Figure 2.2: *Expected cross sections and event rates of several processes as a function of the centre of mass energy of hadron collisions[47]. Discontinuities are due to the difference between $p\bar{p}$ collision for Tevatron and pp for LHC.*

(Lorentz boost does not trasform transverse coordinates of a four-vector when the boost is along z). Another useful observable is the *rapidity* y defined in relativistic kinematic as

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

where E is the particle's energy, and p_z the projection of particle's momentum along the beam direction. Under a Lorentz boost along z with speed β, y undergoes the following transformation: $y \rightarrow y - \tanh^{-1}(\beta)$, hence rapidity differences are invariant, thus the shape of the rapidity distribution dN/dy is invariant.
In the ultra-relativistic approximation ($m \sim 0$, so $p_z \sim E$) it can be shown that the rapidity y is the same as the pseudo-rapidity η defined as

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

where θ is the polar angle between the direction of the three-vector and the boost direction (i.e. beam direction). It's useful (being often in the ultra-relativistic approximation) to refer to pseudorapidity η as it depends only on the direction of the three-vector.

2.1.1 The LHC Schedule

First LHC runs with single beams have been in September 2008. Unluckily, some problems arised during the tests[45], bringing to an to an accident which showed some weaknesses in the LHC project and forced a shutdown and the revision of the plans for the firsts years of running of the machine. First of all, beam energy and luminosity have been lowered. First single beams and collisions were at 900 GeV and 2.36 TeV at the end of the 2009. Energy will be increased at the start of 2010. In the start up period (the first months of operation) the luminosity will be $8 \cdot 10^{29}$ cm^{-2}s^{-1}. The luminosity will then be raised gradually to $1 \cdot 10^{31}$ cm^{-2}s^{-1}. In all this period, LHC will provide collisions at no more than 3.5-5 TeV per beam (7-10 TeV in the centre of mass frame). Only when the machine will be completely understood the energy will be raised to the nominal 7 TeV per beam (14 TeV in the centre of mass frame). Anyway, luminosity will be kept less than 10^{34}cm^{-2}s^{-1}. These values are expected to be reached only with the upgrade of LHC (SLHC)[46]. The *integrated luminosity*[2] in the first year of running is expected to reach 200-300 pb^{-1}. In Tab. 2.1 there are reported also the nominal values of working of the machine foreseen for the firsts runs. The reported values depend from the right behaviour of the machine and can be changed in the future.

[2]The integrated luminosity is defined as the time integral of the luminosity, i.e. $\int dt\mathcal{L}$, so it is measured in cm^{-2}. Anyway, in particle physics is often used the *inverse barn* (1 b^{-1}) and its submultiple to refer directly to the cross section of a process measured in *barn* (1 b = 10^{-24} cm^{-2}).

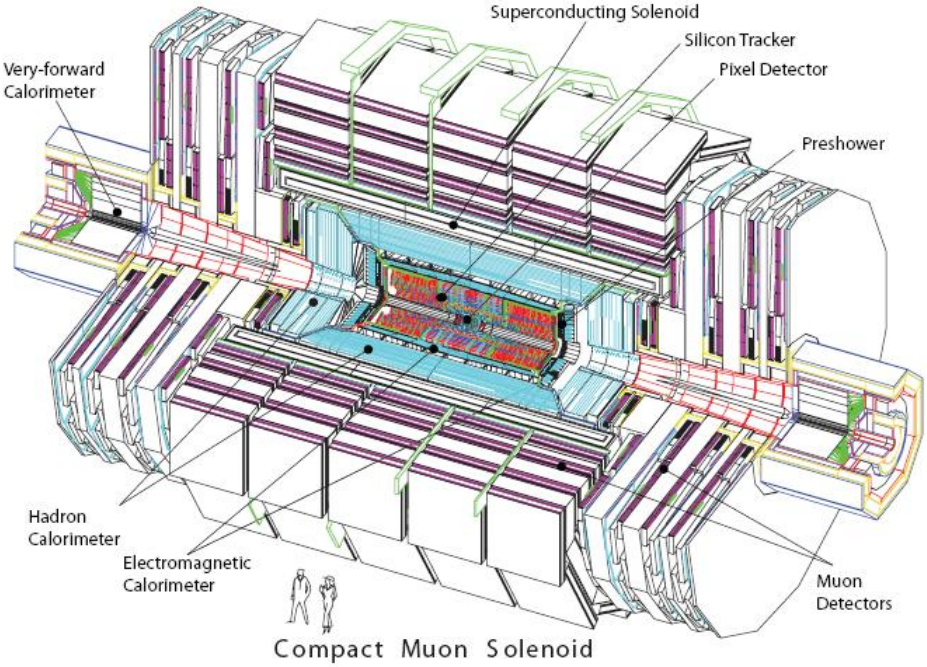

Figure 2.3: *A view of the CMS detector with its subdetectors labelled.*

2.2 The CMS Detector

The Compact Muon Solenoid experiment (CMS)[40] is a general purpose LHC experiment. Its main feature is the 3.8 T 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*). Fig. 2.3 and 2.4 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 12500 tons.

The coordinate frame used in CMS is a right-handed cartesian one, 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 coordinate frame. 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, and η is defined in Eq. 2.6.

CMS is made up of four main subdetectors:

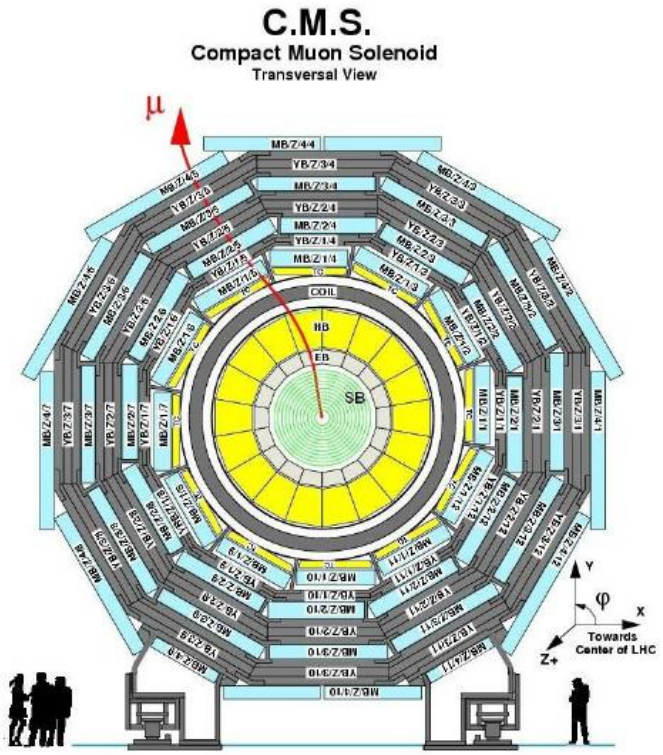

Figure 2.4: *A transverse view of the CMS detector.*

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

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.

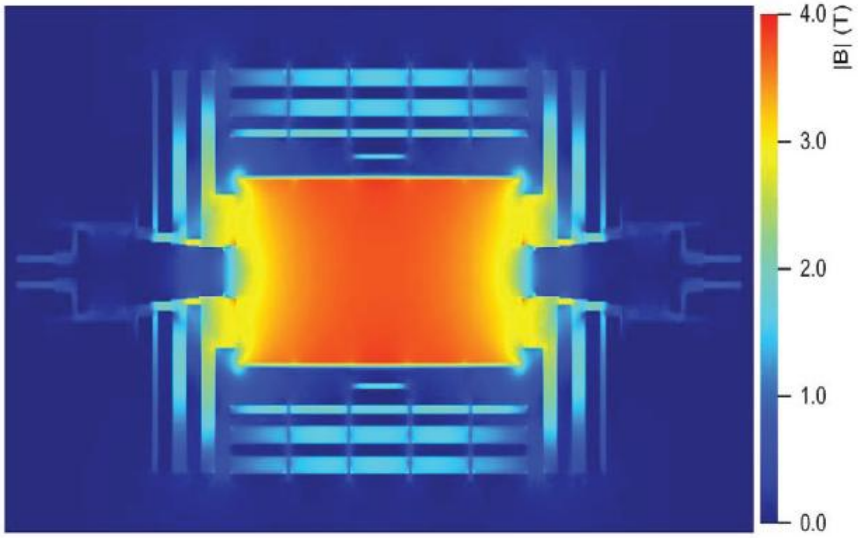

Figure 2.5: *The calculated magnetic field in CMS for a central field of 3.8 T. The field is plotted for a cross-section of the CMS detector in the horizontal plane at the level of the beam line [49].*

2.2.1 The Solenoid

The magnet in CMS, which houses the tracker, the electromagnetic and the hadronic calorimeters, is the biggest superconducting solenoid in the world. The calculated map of CMS magnetic field is shown in Fig. 2.5. The magnet achieves a magnetic field of 3.8 T in the free bore of 6 m in diameter and 12.5 m in length. The energy stored in the magnet is about 2.6 GJ at full current. The superconductor is made of four layers NbTi. In case of a quench, when the magnet looses its superconducting property, the energy is dumped to resistors within 200 ms. The magnet return yoke of the barrel has 12-fold rotational symmetry and is assembled of three sections along the z-axis; each is split into 4 layers (holding the muon chambers in the gaps). Most of the iron volume is saturated or nearly saturated, and the field in the yoke is around the half (1.8 T) of the field in the central volume. More detailed information can be found in reference [48].

2.2.2 The Tracker

The Silicon Tracker 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 trough the reconstruction of

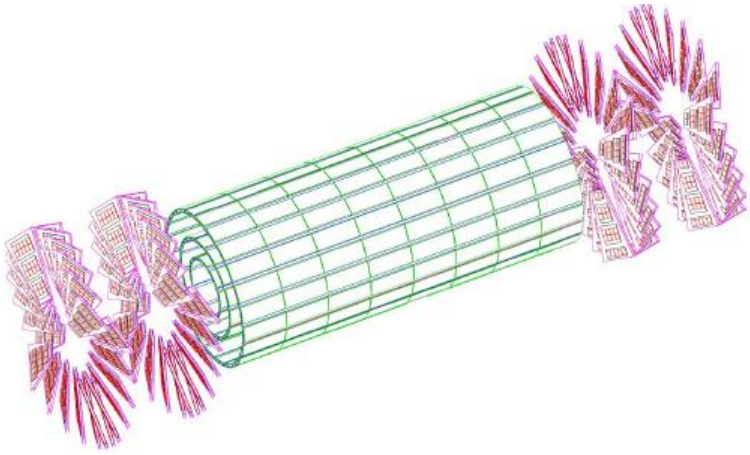

Figure 2.6: *A schematic view of the pixels vertex detector.*

the so called *hits*[3].

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 redudancy

The low hit occupancy is achieved with a highly granular detector, while the redundancy is achieved with a large number of detecting layers. In the following, the two tracker subdetectors are explained.

The Pixel vertex detector

The pixel detector[50] of Fig. 2.6 is a fundamental device for impact parameter measurements. It has also paramount importance as a starting point in reconstructing charged particle tracks. It covers the region $|\eta| < 2.5$ and it is organized into three 53 cm long barrel layers, positioned at $r = 4.4$, 7.3 and 10.2 cm, and two disks per each side, placed at $z = \pm 34.5$ cm and ± 46.5 cm covering radii between 6 and 15 cm to guarantee at least two crossed layers per track coming from the centre of the detector within the fiducial angle $|\eta| <$2.5. Each layer is composed with modular detector units, containing a 250 μm thin segmented sensor plate with highly integrated readout chips. Since both $r\phi$ and z coordinates are important for vertex finding and impact parameter resolution, a

[3]The *hit* definition is not trivial, but at this level it can be defined as the trajectory point intersecting the detector layer plane

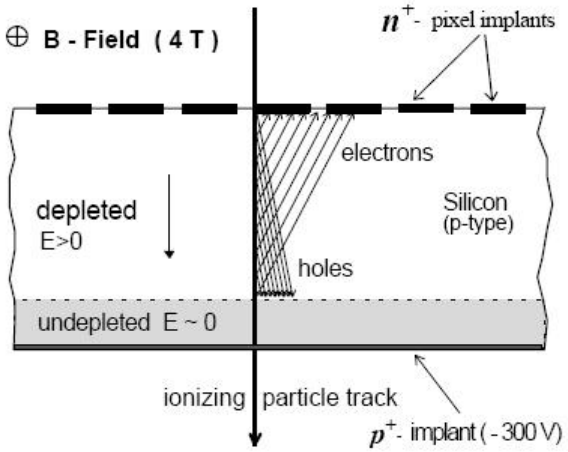

Figure 2.7: *Charge sharing induced by Lorentz drift.*

rectangular pixel shape has been chosen to optimize both measurements. The pixels have a size of 100×150 μm^2 and are combined with analog signal readout to profit of charge sharing effects among pixels and improve position resolution by interpolation. The charge sharing between pixels is enhanced by the Lorentz drift of charge carriers, which is about 25^0 in the barrel and 4^0 for electrons in a 4 T magnetic field at 100 V of bias voltage, three times wider than for the holes. Therefore initial n-type substrate sensors are chosen to collect electron signals on n^+ implants, which in turn are more radiation hard. In the barrel the pixels are tilted to induce significant charge sharing between neighbouring implants in rϕ plane improving the intrinsic hit resolution down to 10-15 μm, far below the 150 μm width of each n^+ implant, with the mechanism drawn in Fig. 2.7. Charge sharing is present also along z direction for inclined tracks leading to a similar resolution. The detectors placed on the disks are rotated with an angle of 20^0 around the central radial axis to benefit of charge sharing improved both in r and $r\phi$ directions by induced Lorentz effects. Despite a Lorentz angle reduced with respect to the barrel case, the resolution in r and $r\phi$ is expected to be 15 μm at CMS start and degrading to 20 μm when radiation damages arise.

The Silicon Microstrips

The outer part of the tracker is made with layers of silicon microstrip detectors[51]. The detector unit is the *module*[4], which is made with one or two sensors glued on a carbon fibre mechanical support together with the readout electronics. The sensor is a n-type phosphorus doped substrate with p^+ implant strips. The p^+-n junction is reversely biased

[4]There are more than 15000 modules in the microstrips detector.

by applying a positive voltage (hundreds of Volts) to the n side keeping the strips at ground. In such a way the region between the junction and the backplane is completely depleted of free charge carriers, with exception of the thermally created ones. When an ionizing particle passes through the silicon, it interacts in the bulk creating electron-hole pairs which drift in the electric field towards the backplane and the p^+ implants respectively. The mean energy required to create a electron-hole pair in silicon is 3.6 eV, therefore a minimum ionizing particle (*mip*) with an average energy loss per path length of 390 eV/μm should create 32500 electron-hole pairs passing through a 300 μm thick sensor. At the detector edges, n^+ implants are placed to limit charge injection from the regions damaged by the cut on the wafer. Insulating capacitor layers of dielectrics (SiO_2, Si_3N_4) are grown between the p^+ implants and the aluminium strip electrodes to decouple the readout electronics from the detector leakage current.

The analog signal of each strip is transmitted to ADCs located in the counting room via optical links. With these microstrip sensors it is possible to measure one coordinate, interpolating the crossing position by means of charge sharing between adjacent strips. A configuration to allow the bidimensional measurement of coordinates is realized gluing two detectors back-to-back with 100 mrad tilted strip directions. This configuration is referred to as *stereo* or *double-sided* and is preferred to the pixel segmentation, although the resolution is worse, because the number of readout channels is less. The hit ambiguities present in this detector configuration are resolved at track reconstruction level.

The silicon strip tracker covers a tracking volume up to $r = 1.1$ m with a length of 5.4 m and is divided in four parts, as it is shown in Fig. 2.8: TIB (*Tracker Inner Barrel*), TID (*Tracker Inner Disks*), TOB (*Tracker Outer Barrel*) and TEC (*Tracker EndCap*).

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^{0}C).

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[5] and in terms of

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

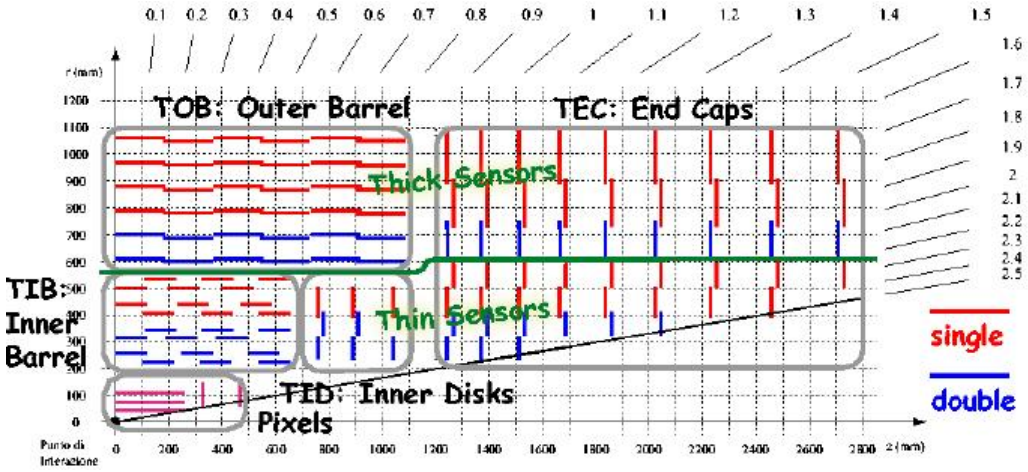

Figure 2.8: *An $r-z$ schematic view of a sector of the Silicon Strip Tracker. The location of single-sided and double-sided detectors is put into evidence.*

interaction length λ/λ_0[6] as obtained from the full simulation of the tracker is shown in Fig. 2.9 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.

The **alignment** of the tracker modules is very important to obtain high-precision spatial resolution. Deviations are caused by assembly inaccuracy, deformations due to cooling and stress from the magnetic field. Therefore, three methods are used for the tracker alignment. The geometry was determined during assembly to an accuracy of 80 to 150 μm. An infrared laser system is used for continuous monitoring of the position of selected tracker modules. The final alignment is done with tracks from well known physics processes, e.g. cosmic muons, or di-muons from J/Ψ, Υ and Z^0.

2.2.3 The Electromagnetic Calorimeter (ECAL)

The main goal of the ECAL[52][53] is the precise measurement of the energy of electrons and photons. Its design was led by the requests imposed by the H$\rightarrow \gamma\gamma$ channel, which is studied looking for a peak in the di-photon invariant mass distribution, and where the invariant mass width is dominated by the experimental resolution[56]; it is required to be order of 1%. High granularity is necessary, to improve the measurement of the angle between the two photons and to obtain a good $\pi^0 \rightarrow \gamma\gamma/\gamma$ separation.

An homogeneous calorimeter has been chosen. ECAL consists of almost 76000 Lead Tungstate ($PbWO_4$) scintillating crystals divided into a barrel and two endcaps; a 3D

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

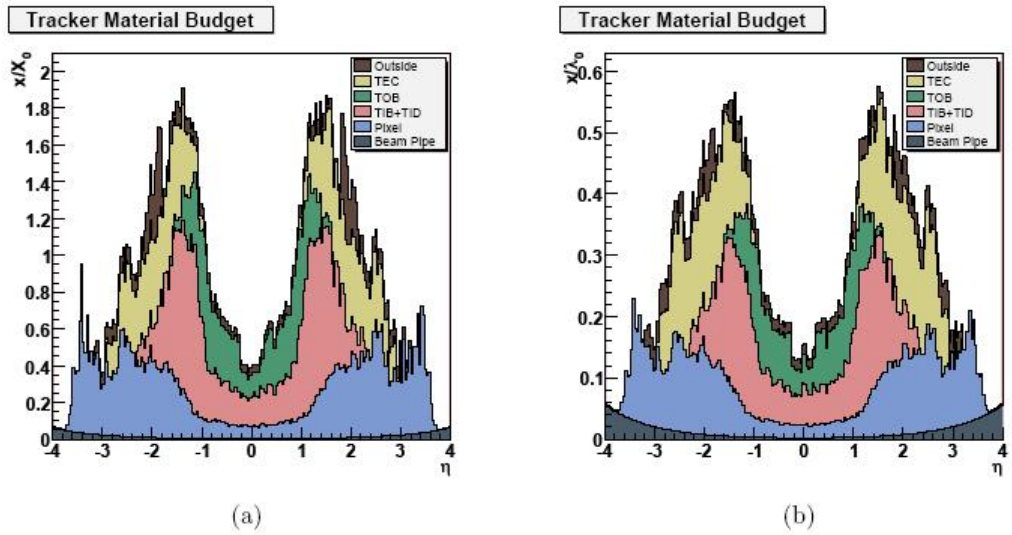

Figure 2.9: *a) Radiation length and b) interaction length of the tracker as a function of η. Contributions from different components are put into evidence.*

view of the calorimeter is given in Fig. 2.10. The barrel ($|\eta| < 1.479$) consists of 36 **supermodules**, each one containing a matrix of 20 crystals in ϕ × 85 crystals in η and covering an azimuthal angle of 20^0. The supermodules are divided along η in 4 modules with the **submodules** as basic units, which are the basic units of ECAL and which consist of 5×2 crystals each. The geometrical shape of the crystals slightly changes along η and there are 17 types of crystals, with length close to 230 mm and front face area of about 22×22 mm^2. The barrel granularity is $\Delta\phi \times \Delta\eta = 0.0175 \times 0.0175$, the crystals are grouped into 5×5 arrays called **trigger towers**, providing information to the trigger system (see Chapter 5). To avoid that cracks might align with the particles trajectories, the crystal axes are tilted by 3^o with respect to the direction from the interaction point, both in ϕ and in η.

Each endcap consists of two halves (*Dees*) and covers the pseudorapidity region between 1.48 and 3. All the crystals have the same shape (220×24.7×24.7 mm^3) and they are grouped in structures of 5×5 crystals called **super-crystals**. The granularity varies from $\Delta\phi \times \Delta\eta = 0.0175 \times 0.0175$ to 0.05×0.05. As for the barrel, the crystals have a non pointing geometry. To ensure good hermeticity, an overlap of half crystal between the endcaps and the barrel is obtained by orienting the crystals axis to point 1300 mm beyond the interaction point.

To improve the π^0/γ separation and the vertex identification, a preshower is designed to cover the region between $|\eta| = 1.6$ and $|\eta| = 2.6$. It consists of two lead converters (2 X^0 and 1 X^0 thick) followed by silicon strips with a pitch of less than 2 mm. The strips

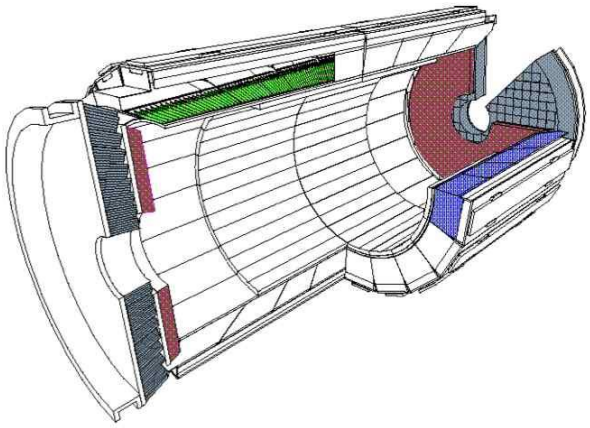

Figure 2.10: *Three dimensional view of the electromagnetic calorimeter.*

following the two absorbers are disposed in orthogonal way. The preshower will operate at the temperature of -5^{0}C.

Different reasons brought to the choice of the $PbWO_4$ as active medium for ECAL. Its short radiation length (X^0 = 0.89 cm) and Moliere Radius[7] (R_M = 2.19 cm) allow to build a compact and high granularity calorimeter. An important aspect is the fast response (∼80% of the light is collected within 25 ns), which is compatible with the high LHC rate. Finally, the $PbWO_4$ has a good intrinsic radiation hardness, which makes it suitable to work in the hard LHC environment. The main drawback of the $PbWO_4$ crystals is the low light yield (∼10 photo-electrons/MeV), which makes an internal amplification for the photodetectors necessary.

The photodetectors for ECAL have to be radiation hard, fast and able to operate in the strong CMS magnetic field. The devices which match these characteristics and that have been chosen for the electromagnetic calorimeter are the *Avalanche PhotoDiodes* (APDs) for the barrel and the *Vacuum PhotoTriodes* (VPTs) for the endcaps.

The APDs are silicon detectors. The scheme of functioning is shown in Fig. 2.11: a 5 μm thick p^+ layer acts as photoconverter, the photoelectrons are accelerated and multiplied through the $p-n$ junction reverse biased and then a n^{++} doped region provides the ohmic contact with the preamplifier. The active area is a 5×5mm^2 surface. Even if it is quite small, the APDs have an high quantum efficiency (∼75% at 430 nm) which well matches the emission spectrum of the $PbWO_4$; besides, each crystal is equipped with 2 APDs to increase the acceptance to the scintillation photons.

[7]The Moliere Radius (R_M) gives an estimate of the transverse development of an electromagnetic shower: on average 90% of the energy released by the shower lies in a cylinder with radius R_M.

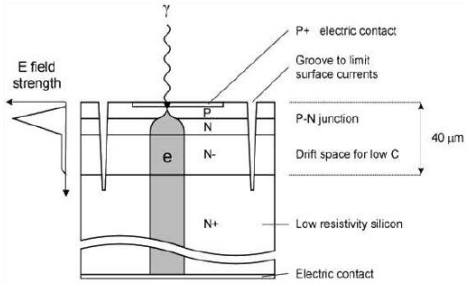

Figure 2.11: *Scheme of an APD.*

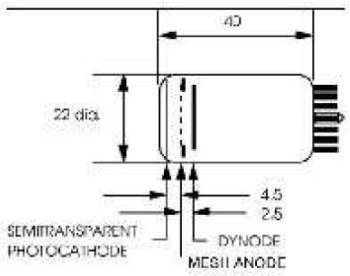

Figure 2.12: *Scheme of a VPT.*

The statistical fluctuations in the multiplication process influence the stochastic term of the energy resolution (see later). The calorimeter noise is affected by the APD capacitance and by the leakage currents flowing on the surface and in the bulk of the APD. Finally, the sensitivity of the APD gain to the biasing voltage and to the temperature directly influences the energy resolution with a constant term. The APDs are sensitive to radiation. They might get damaged by neutrons which create defects in the silicon increasing the leakage currents. Since the APDs can not survive the radiation doses of the endcaps, the technology of the Vacuum PhotoTriodes has been chosen for those regions.

The VPTs are phototubes with a bi-alkali photocathode deposited on a glass window (Fig. 2.12). The electrons emitted from the cathode are accelerated towards a 10 μm thick anode; the fraction of them passing the anode grid impacts on a reflective dynode with a planar geometry and then emits new electrons ($\sim$20 secondary electrons are emitted for each impinging electron), which are accelerated back toward the anode. The VPT quantum efficiency is about 15% at the peak of the $PbWO_4$ emission spectrum, but since the active area is almost 300 mm^2 the total light collection is at the same level as for the APDs. The VPTs operate at gain which is much lower with respect the one of the APDs but which presents very little dependence on the temperature fluctuations and on the biasing voltage. The capacitance is low (few pF), and the leakage current is < 2 nA. The VPTs use a radiation hard glass which has been appositely developed and whose

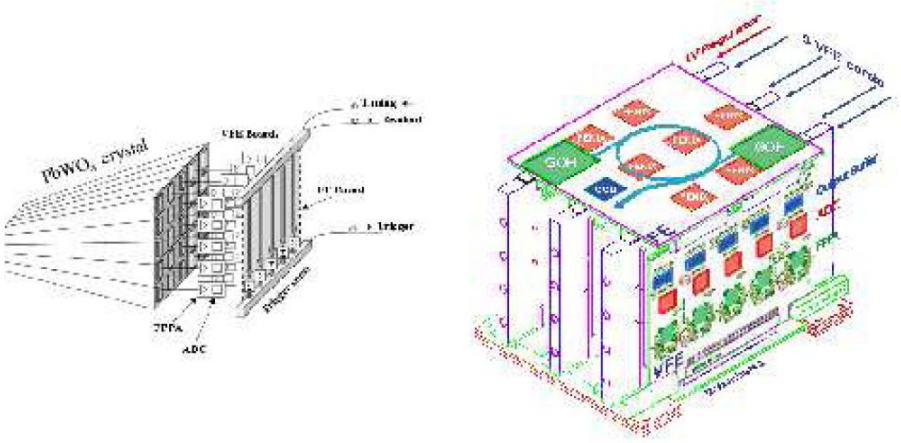

Figure 2.13: *Left: schematic view of the ECAL electronics for 5×5 crystals. Right: the arrangement of the front-end electronics into VFE and FE boards.*

transparency is not strongly affected by irradiating particles.

VPT cannot be used in the barrel region, where the magnetic field is perpendicular to the axis of the photo-detector.

Important requirements are imposed on the front-end (FE) electronics. It has to be fast enough to match the 25 ns LHC crossing rate and it must keep the noise level below 50MeV per crystal over a dynamic range of about 95dB; since it is placed on the detector it has to be radiation hard. The electronics mirrors the trigger structure and its basic element is a group of 5×5 crystals called **trigger tower**. The signals coming from each photodetector are sent to a motherboard which hosts 5 *Very Front End cards* (VFE) and a *Low Voltage Regulator card* (LVR) providing the power for the VFE cards. The output of the 5 VFEs is fed into a FE card which processes the digitised data of one trigger tower.

In addition, each VFE card has a *Detector Control Unit* which measures the crystal temperature and the APD leakage currents. One chip for the L1 Trigger primitive generation (the *Fenix Asics*) is put on each trigger tower within the detector to store the data until the positive response of L1.

The radiation induces the creation of "colour centres" in the crystals, which make them partially opaque to their emitted light. Thanks to the radiation resistant properties of the $PbWO_4$ crystals, this effect is reversable and limited to $\sim$5% of the crystal light yield. Nevertheless, it has to be monitored. That's why a monitoring system was created, based on the injection of laser light into each crystal to measure its transparency and perform fast connections to its calibration. During the LHC data taking, the monitoring system will regularly provide light pulses during the 3 μs gap the beam will have every 89

μs; dedicated runs to follow the crystal recovery are also foreseen during the LHC refills. The laser distribution system proceeds in three steps. First the light emitted by the laser source is split by means of an optical switch into each of the 80 channels the calorimeter is organized in (72 half supermodules in the barrel and 8 groups of supercrystals in the endcaps), then a 2 steps distribution system at the level of the single unit further splits the pulses to reach each crystal. A level 2 *fan-out* splits the light into 5 or 6 (according to which part of the supermodule is fed) quartz fibers; 4 or 5 fibers go to the level 1 fan-out and the last one is sent to a reference $p-n$ diode for the monitoring of the pulse amplitude. The level 1 *fan-out* then splits the light into 200 fibers going to the crystals. Two output fibers finally bring the light to two monitoring PNs. In the barrel the light is injected from the front face of the crystal and it directly reaches the APD, which is on the rear face; in the endcaps, the light is injected from the rear face and it reaches the photodetector after a reflection on the front face of the crystal. Each pulse energy power is 1 mJ at the monitoring wavelength.

The energy resolution of an homogeneous calorimeter is usually written as

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.7}$$

where a, b and c represent respectively the *stochastic*, *noise* and *constant term* of the energy resolution.

As it was mentioned above, The *stochastic term* a includes the contribution of the fluctuations in the number of electrons which are produced and collected. Since the fluctuations are poissonian, the stochastic term is $1/\sqrt{n_{pe}}$, where n_{pe} is the number of photo-electrons which are emitted per energy unit. Contributions come from the light yield of the crystals, from the efficiency in the light collection and from the quantum efficiency of the photo-detectors. Important contributions also come from the fluctuations in the multiplication process inside the photodetectors (*excess noise factor*). The target value for ECAL is $a = 0.027$ for the barrel and 0.057 for the endcaps, where the main contribution ($5\%/\sqrt{E}$) comes from the preshower sampling term.

The noise term b includes contributions from the electronic noise, both due to the photodetector and to the preamplifier, and from pile-up events. The contributions change at the different pseudorapidities and with the luminosity of the machine. The target values in the low ($\sim 10^{30}$ cm^{-2}s^{-1}) and in the high ($\geq 10^{33}$ cm^{-2}s^{-1}) luminosity phases are respectively 155 MeV and 210 MeV at $|\eta| = 0$, 205 MeV and 245 MeV at $|\eta| = 2$.

The constant term c is the dominating term at high energies and it includes many different contributions. Among them, the most important are:

- the stability of the operating conditions, such as the temperature and the high

voltage.
Both the scintillation mechanism and the APD gain are affected by the temperature and the response for a given energy deposit varies with the temperature of the calorimeter with a slope which is around -4%/^{o}C for the barrel. The stability of the temperature within 0.05^{o}C is required to keep the contribution to the constant term below 0.1%.

- the presence of dead materials in front of the crystals and the rear leakage of the electromagnetic shower. Anyway, the 25 radiation length featured by the crystals keeps this effect negligible up to energies in the TeV range.

- the longitudinal non uniformity of the crystal light yield. A strong focusing effect of the light takes place due to the tronco-pyramidal shape of the crystals and to the high refractive index, so the light collection is not uniform. The light collection was corrected by mechanically abrading one lateral face of each crystal to an extent able to keep non-uniformity at the level of 0.35%/X^0, keeping the contribution to the constant term c below 0.3%.

- the intercalibration errors, which is going to be the dominant contribution.

- the radiation damage of the crystals, which changes their response to a certain amount of deposited energy when exposed to high radiation dose rates.

The target value for the constant term c of the CMS ECAL is 0.5%.

2.2.4 The Hadronic Calorimeter (HCAL)

The CMS hadron calorimeter (HCAL)[54] is used together with ECAL to measure energy and direction of jets, and the energy imbalance in the transverse plane E_T^{miss} . It provides good segmentation, moderate energy resolution and angular coverage up to $|\eta| < 5$.
HCAL is made of four subdetectors (Fig. 2.14):

- 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 nuclear interaction length λ

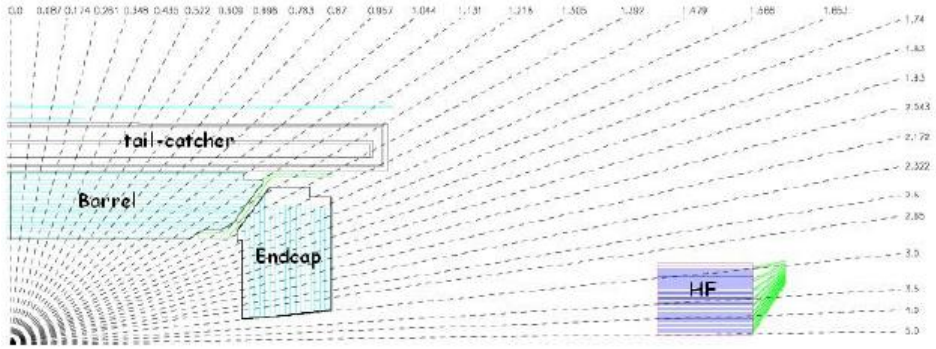

Figure 2.14: *A schematic representation of a quadrant of HCAL.*

- 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 $|\eta| = 5$.

HB and HE are made with layers of 4÷7.5 cm thick brass stainless steel absorber plates interleaved with 3.7 mm thick plastic scintillators. The signal is readout through wavelength-shift fibres and hybrid photodiodes (HPD). The granularity[8] 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 λ. 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 λ, with an improvement in both linearity and energy resolution. HE has a minimum depth of 10 λ. 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 0.17×0.1745.

2.2.5 Muon System

The CMS muon system [55] 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 1.8 T return flux.

The system consists of three independent subsystems (Fig. 2.15):

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

[8]Defined as $\Delta\phi \times \Delta\eta$

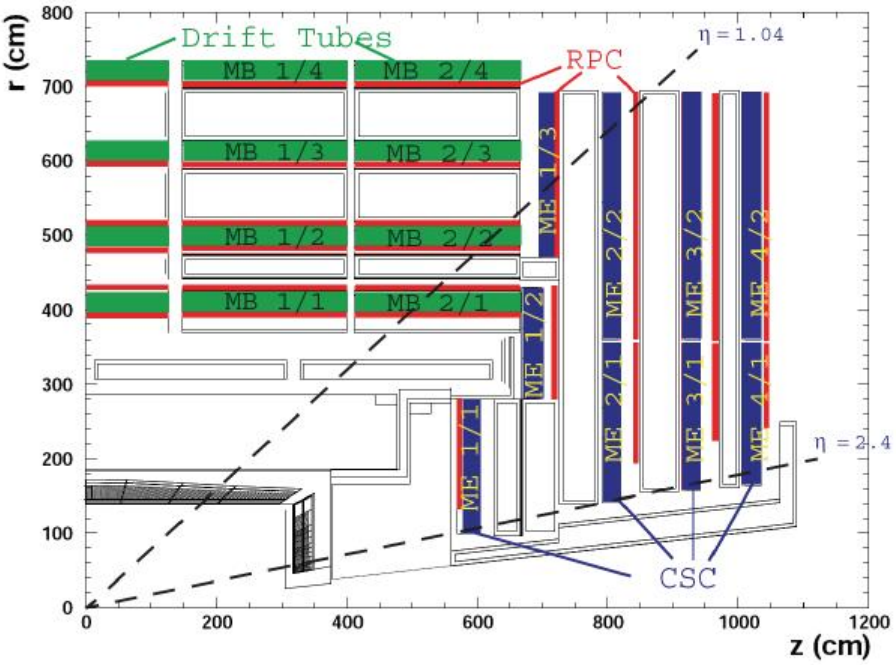

Figure 2.15: *A schematic view of a quadrant of the CMS muon system.*

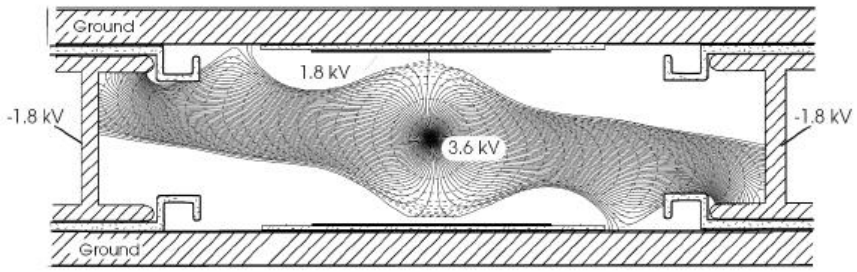

Figure 2.16: *A schematic representation of a drift tube chamber. Drift lines in presence of magnetic field are also shown.*

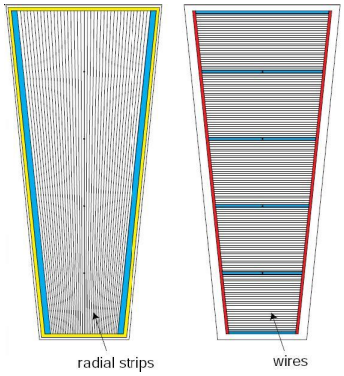

Figure 2.17: *A schematic representation of CSC cathode panel (left) and anode panel (right).*

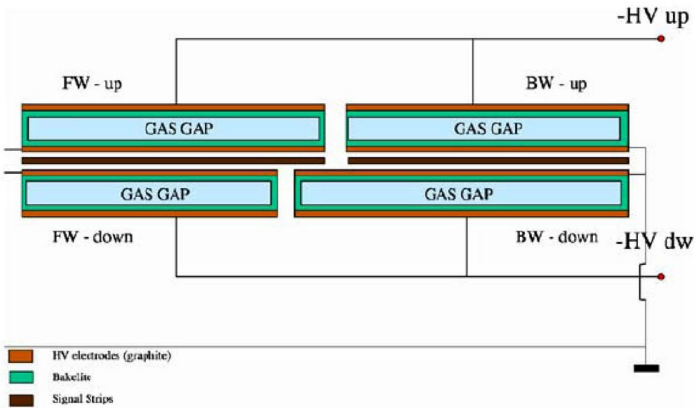

Figure 2.18: *A schematic representation of RPCs in the magnetic field return yoke.*

- **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*, for a total of four chambers with three super-layers per chamber. In each chamber two super-layers have anode wires parallel to the beam axis, and one has perpendicular wires. Thus, each chamber can provide two measurements of the $r-\phi$ coordinate and one measurement of the z coordinate of the track hit positions. Each chamber (Fig. 2.16) is made of two parallel aluminium plates jointed with "I" shaped spacer cathodes. 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 $r-z$.

Cathode Strip Chambers are multi-wire proportional chambers with segmented cathodes (Fig. 2.17). 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 (Fig. 2.18) are made of parallel bakelite planes, with a bulk resistivity of $10^{10} \div 10^{11}$ Ωcm. The gap between the plates is filled with a mixture of $C_2H_2F_4$ (96.2%), iso-C_4H_{10} (3.5%) and SF_6 plus water vapour (0.3%). They operate in avalanche mode. Those chambers have limited spatial resolution, but they have excellent timing performances; they are used for bunch crossing identification and for trigger purposes.

Chapter 3

Generation of $J/\Psi \rightarrow e^+e^-$ events at LHC with the `Pythia` Monte Carlo generator

In this work reconstruction, triggers and off-line analysis studies of prompt $J/\Psi \rightarrow e^+e^-$ events are made using Monte Carlo technique, i.e. via the simulation of LHC collisions and consequent formation of prompt J/Ψ's assuming various theoretical models. In this sense, the simulation step of physics processes via Monte Carlo is said **generation**.

The generation step is based on Monte Carlo numerical algorithms: the structure of events produced at high energy colliders is extremely complex. Computer programs are needed for this scope, which are called **generators**. Monte Carlo generators 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.

A Monte Carlo Generator of hadron-hadron collision events has to implement the following steps[57]:

- The first one is the calculation of cross sections for the selected processes. 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 (according to the knowledge of the PDFs) 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 electric charges are present, thus bremsstrahlung can occur. This effect is called Initial State Radiation (ISR). To simulate ISR knowledge of the PDFs is needed.
- Also the final state partons can produce further radiation, called Final State Radiation (FSR).
- 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 need 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 generator takes care of decaying unstable particles such as τ leptons or B-hadrons; in general particles with very short lifetime are decayed by the generator itself. Those that live enough to reach detectors (such as μ leptons) 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 (LO in perturbation theory) generators. One of them is the `Pythia` generator.

`Pythia`, such as other generators, is mainly subdivided into two sections: in the first the *Matrix Element Computation* (ME) is done for the calculation of the hard processes

cross sections. In Pythia differential cross sections are hard-coded and 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. Then, the main problems with tree-level matrix elements are the soft and collinear divergences[57], which arise when partons have low p_T exchange or are closely emitted, so the phase space has to be carefully tailored to avoid the problematic regions in the integrals of the differential cross sections. This means that the matrix element cross section calculations are performed away from soft and collinear divergences and, 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* computations.

The *Parton Shower* (PS) is the second section of a Monte Carlo generator, and it offers 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. This behaviour is modelled 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. This effect is modelled by Final State parton showers. Soft and collinear divergences are removed by *resummation* technique. Indeed, when calculating a QCD observable perturbatively, the expansion in terms of α_s contains terms with form $\alpha_s^n L^k$ (with $k \leq 2n$), where $L = \ln q_{cut}$, *qcut* 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*; it 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; etc.) 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 [57][58][59] 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 this work, prompt J/Ψ physics simulation is based on Pythia simulation. In the following, it is described how NRQCD model is implemented in *Pythia* and which settings are chosen in events generation.

3.1 Heavy Quarkonia sector in Pythia

The original code for the inclusion of NRQCD matrix elements in Pythia has been developed by S. Wolf and implemented officially by T. Sjostrand[60] in 2005. In 2006 it was tuned by M. Bargiotti[61].

3.1.1 The Sub-processes

In Pythia are hard-coded (at LO in perturbation theory) the sub-processes which brings to a quarkonium bound state (see Tab.3.1). Imposing the Pythia flag MSEL(61)=1 all sub-processes are switched on. In the simulation, the octet $c\bar{c}$-states have a mass of 3.1 GeV/c^2.

The non-perturbative part of the generation (see Chapter 1) is encoded in the *NRQCD matrix element* (Tab. 3.2), and they represent free parameters of the theory. Their values are taken from [61], where they are tuned based on Tevatron data. It is important to note that all ${}^3P_j$ octet states are simulated using a single NRQCD matrix element $< \mathcal{O}({}^3P_j^{(8)}) >$ with the relation of $< \mathcal{O}({}^3P_j^{(8)}) >= (2j+1) < \mathcal{O}({}^3P_0^{(8)}) >$ (see Chapter 1).

The default branching ratios for the radiative decays of the χ_c-states into J/Ψ's have been modified to correspond to the recent PDG values[63].

In this study we forced the J/Ψ to always decay into e^+e^- pairs via electromagnetic interaction; its branching ratio is 6.02%.

3.1.2 Cross Section Regularization at low p_T

All the cross-sections of the CSM and COM processes considered in Pythia are divergent at LO for p_T tending to zero. In fact QCD perturbation theory breaks down at low p_T values because confinement is not taken into account. From a phenomenological point of view, the story is that the gg→gg (which is the process with the higher contribution

ISUB	Charmonium subprocesses
421	$g + g \to c\bar{c}[^3S_1^{(1)}] + g$
422	$g + g \to c\bar{c}[^3S_1^{(8)}] + g$
423	$g + g \to c\bar{c}[^1S_0^{(8)}] + g$
424	$g + g \to c\bar{c}[^3P_j^{(8)}] + g$
425	$g + q \to q + c\bar{c}[^3S_1^{(8)}]$
426	$g + q \to q + c\bar{c}[^1S_0^{(8)}]$
427	$g + q \to q + c\bar{c}[^3P_j^{(8)}]$
428	$q + \bar{q} \to g + c\bar{c}[^3S_1^{(8)}]$
429	$q + \bar{q} \to g + c\bar{c}[^1S_0^{(8)}]$
430	$q + \bar{q} \to g + c\bar{c}[^3P_j^{(8)}]$
431	$g + g \to c\bar{c}[^3P_0^{(1)}] + g$
432	$g + g \to c\bar{c}[^3P_1^{(1)}] + g$
433	$g + g \to c\bar{c}[^3P_2^{(1)}] + g$
434	$g + q \to q + c\bar{c}[^3P_0^{(1)}]$
435	$g + q \to q + c\bar{c}[^3P_1^{(1)}]$
436	$g + q \to q + c\bar{c}[^3P_2^{(1)}]$
437	$q + \bar{q} \to c\bar{c}[^3P_0^{(1)}] + g$
438	$q + \bar{q} \to c\bar{c}[^3P_1^{(1)}] + g$
439	$q + \bar{q} \to c\bar{c}[^3P_2^{(1)}] + g$

Table 3.1: *Colour-singlet/octet production sub-processes and their corresponding* ***ISUB*** *indices in* ***Pythia****.*

Pythia flag	Matrix El.	Value
PARP(141)	$< \mathcal{O}_{(J/\Psi)}(^3S_1^{(1)}) >$	1.16
PARP(142)	$< \mathcal{O}_{(J/\Psi)}(^3S_1^{(8)}) >$	0.0119
PARP(143)	$< \mathcal{O}_{(J/\Psi)}(^1S_0^{(8)}) >$	0.01
PARP(144)	$< \mathcal{O}_{(J/\Psi)}(^3P_0^{(8)}) > /m_c^2$	0.01
PARP(145)	$< \mathcal{O}_{(\chi_{c0})}(^3P_0^{(1)}) > /m_c^2$	0.05
PARP(146)	$< \mathcal{O}_{(\Upsilon)}(^3S_1^{(1)}) >$	9.28
PARP(147)	$< \mathcal{O}_{(\Upsilon)}(^3S_1^{(8)}) >$	0.15
PARP(148)	$< \mathcal{O}_{(\Upsilon)}(^1S_0^{(8)}) >$	0.02
PARP(149)	$< \mathcal{O}_{(\Upsilon)}(^3P_0^{(8)}) > /m_b^2$	0.02
PARP(150)	$< \mathcal{O}_{(\chi_{b0})}(^3P_0^{(1)}) > /m_b^2$	0.085

Table 3.2: *The 10 J/Ψ and Υ NRQCD matrix elements and corresponding parameters in Pythia.*

in charmonia production at hadrons colliders) cross section should be suppressed at small transverse momentum of the exchanged gluon, because this gluon has such a large associated transverse wavelength that it cannot resolve the individual colours of the partons inside the proton. So the simulation has to take in account this effect which reduces the cross section for low p_T values. It was made use of an event reweighting function provided by Pythia called PYEVWT[64], allowing the quarkonia cross section to be rescaled. Omitting this procedure results in a predicted cross section which is several orders of magnitude above the differential cross section measured at CDF at small J/Ψ p_T, as can be seen in Fig. 3.1. PYEVWT adopt a formalism where the strong coupling constant is further decreased at low p_T, re-defining an "effective" α_s; so the differential cross section gg→gg is regularized as:

$$\frac{d\sigma}{dp_T^2} \propto \frac{[\alpha_s(p_T^2)]^2}{p_T^4} \Longrightarrow \frac{[\alpha_s(p_{T0}^2 + p_T^2)]^2}{(p_{T0}^2 + p_T^2)^2} \tag{3.1}$$

where p_T is the transverse momentum of the exchanged gluon, and p_{T0} is a free parameter of the model, empirically of order 2 GeV/c at the Tevatron Run 2 ($\sqrt{s} = 1.96$ TeV)[61].

It is not a priori clear what p_{T0} would be at LHC energies, but is generally believed to grow with s. This is because larger $\sqrt{s}$ values imply that smaller bjorken factor x values of the probing gluon are available, which means there is a denser packing of the gluons and thus more screening. Following Heisemberg's uncertainty relation the value of p_{T0} thus increases. The default energy dependence in Pythia is inspired on Regge Pomerons phenomenology[62], and is parameterized as $p_{T0} \propto E_{CM}^{\texttt{PARP(90)}} \equiv \sqrt{s}^{\texttt{PARP(90)}}$, where PARP(90) is a Pythia flag set to 0.16, being this value inspired by the rise of the total cross section

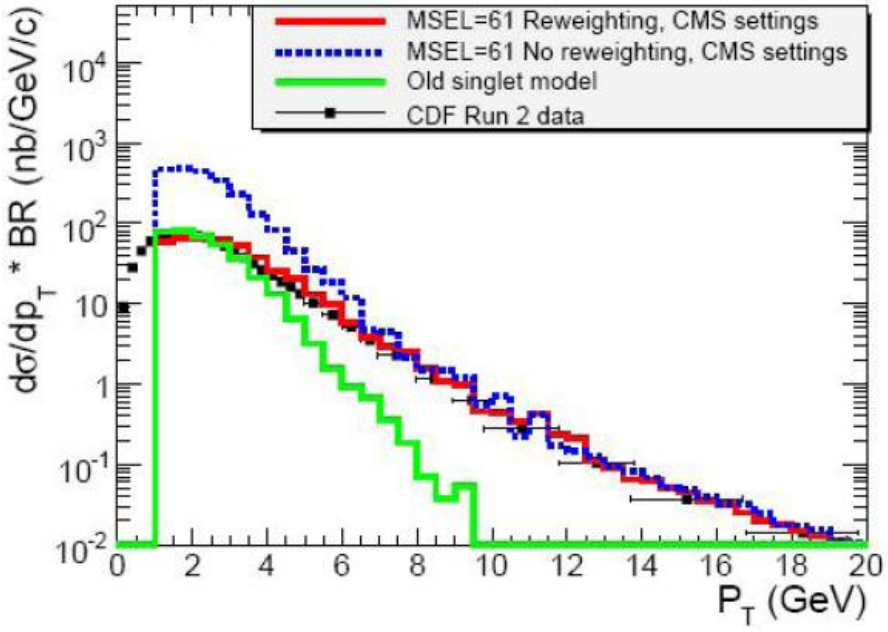

Figure 3.1: *Prompt J/Ψ p_T spectrum obtained at CDF, and by simulation (with or without reweighting)[65].*

with the Pomeron term ($s^{0.08} = E_{CM}^{2\cdot 0.08} = E_{CM}^{0.16}$).

A normalization of the differential cross section along these lines applies naturally also to quarkonium production, since a large part of the production takes place via the Feynman diagram in Fig. 3.2. The normalization of the cross section is not done by default in `Pythia`, but it can be done using the routine `PYEVWT`, giving each event an extra weight w_i while it is being generated, and multiplying the normal cross section. As is illustrated in Fig. 3.2 three rather than two factors of α_s are present in the production cross section, once coming from g→c$\bar{\text{c}}$ and so, from Eq. 3.1, it thus becomes apparent

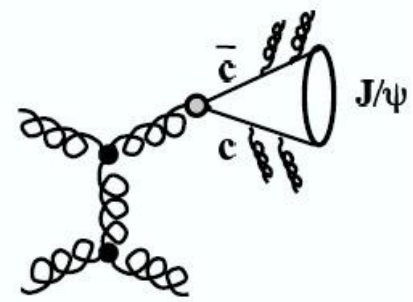

Figure 3.2: *Feynman diagram of c$\bar{\text{c}}$ production via the process gg→gg. Different fillings of the higher vertex represents a value of α_s valuated at a lower energy scale than the other two vertices.*

that each generated event should obtain a weight w_i according to

$$w_i = \frac{\sigma_{reweighted}}{\sigma_{no-reweighted}} = \Big(\frac{\hat{p}_T^2}{p_{T0}^2 + \hat{p}_T^2}\Big)^2 \Big(\frac{\alpha_s(p_{T0}^2 + Q^2)}{\alpha s(Q^2)}\Big)^3 \tag{3.2}$$

where $\hat{p}_T$ is the transverse momentum evaluated in the rest frame of the scattering, Q^2 the momentum transfer scale, and p_{T0} the regularization scale introduced above. As can be seen from Eq. 3.2, at small J/Ψ p_T a suppression is obtained while at larger values the cross section is unmodified.

Finally, to extrapolate the p_{T0} value determined at 1960 GeV to the LHC E_{CM}[65], the energy dependence based on Regge theory is used, as explained above:

$$\begin{aligned} p_{T0}^{Tevatron} &= const \cdot (1.96[TeV])^{0.16}, \\ p_{T0}^{LHC} &= const \cdot (10[TeV])^{0.16} \end{aligned} \tag{3.3}$$

$$\Longrightarrow p_{T0}^{LHC} = p_{T0}^{Tevatron} \cdot \Big(\frac{10[TeV]}{1.96[TeV]}\Big)$$

The latter parameterization in Pythia is

$$p_{T0} = \texttt{PARP(82)} \Big(\frac{E_{CM}}{\texttt{PARP(89)}}\Big)^{\texttt{PARP(90)}} = 1.94[GeV/c] \cdot \Big(\frac{10[TeV]}{1.96[TeV]}\Big)^{0.16} \simeq 2.66[GeV/c] \tag{3.4}$$

where PARP(82), PARP(89) and PARP(90) are Pythia flags, and the LHC E_{CM} was considered to be 10 TeV.

3.1.3 The Shower Activity associated with Quarkonia Production

In the case the J/Ψ is produced via the singlet mechanism, the $c\bar{c}$ does not radiate gluons and the quarkonia is thus produced in isolation, up to normal underlying event (i.e. hadronic radiation from the remnant partons of protons in the collisions) activity. In the case that it is produced via the octet mechanism, the situation is not a priori clear, but it is reasonable to assume that a shower can evolve and that the quarkonia is embedded in a certain amount of jet activity. The NRQCD formalism does not contain any description for perturbative parton cascades, and the approach used in Pythia is the following.

Shower activity around a charmonium state produced via the octet mechanism can result from two contributions. The first contribution is parton showers produced in the hard process gg→gg (*step 1*), where repeated g→gg branches can occur, and eventually a g→$c\bar{c}^{(8)}$ branching (*step 2*). The second contribution is from the radiation of the colour octet $c\bar{c}^{(8)}$ once formed, via $c\bar{c}^{(8)}$ →$c\bar{c}^{(8)}$+g. Although the production of a colour-octet

state is assumed to take place via the two separate *steps 1* and *2*, technically when these two steps are merged together it is as factorization is used, and the colour-octet $c\bar{c}^{(8)}$ states are produced as if in the hard interaction. For the simulation of the shower activity, this means that technically the shower is generated as if the $c\bar{c}^{(8)}$ were produced already at the hard interaction, but where the question at which stage of the overall shower the g→$c\bar{c}^{(8)}$ branching occurs is left open. Thus, both the shower contribution from repeated gluon branchings and that from $c\bar{c}^{(8)}$ radiation are incorporated! `Pythia` parameterize the shower activity with the two flags `MSTP(148)` and `MSTP(149)`, which describe the DGLAP colour evolution until the $c\bar{c}^{(8)}$ hadronizes in the hadron state (see Chapter 1): When `MSTP(148)=0` and `MSTP(149)=0` a small amount of shower evolution of the octet $c\bar{c}^{(8)}$ does occur. As "splitting function" (i.e. function which describe the probability that splitting occurs) for the branching $c\bar{c}^{(8)}$ →$c\bar{c}^{(8)}$+g, the q→qg splitting kernel of DFLAP equations is used, with a modification for the c$\bar{c}$-mass. This splitting function strongly favours the c$\bar{c}$-state to obtain the largest energy in each branching. Thus only a small amount of activity is expected around the J/Ψ, yet the activity here is expected to be much larger than in the colour singlet case. After the shower evolution the octet c$\bar{c}$-state hadronizes into a singlet c$\bar{c}$-state by soft and collinear gluon emission, described with the NRQCD matrix elements. Since the amount of shower evolution is small the shower activity around the J/Ψ here might be underestimated. This production case is thus referred to as **octet low radiation scenario**.

With `MSTP(148)=1` and `MSTP(149)=0` in for the branching $c\bar{c}^{(8)}$ →$c\bar{c}^{(8)}$+g, the g→gg splitting function is used (modified for the c$\bar{c}$-mass). Here, there is a strong suppression of the $z < 0.5$ values, where z is defined as in Eq. 3.5[65]:

$$z = \frac{p_T^{J/\Psi}}{p_T^{J/\Psi} + p_T^{jet_around}} \tag{3.5}$$

(where $p_T^{jet_around}$ is the total p_T of the hadronic jet activity around the direction of the J/Ψ, within a cone with $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.7$) analogous to the q→qg splitting function, so that in each branching $c\bar{c}^{(8)}$ →$c\bar{c}^{(8)}$+g, the c$\bar{c}$-state is likely to get the largest energy. This implementation probably represents a more realistic physics scenario. Because of the different splitting functions used the amount of radiation is larger than in the previous case. A considerable amount of jet activity is expected around the J/Ψ in this case, however because of the suppression of $z < 0.5$ only a small amount of hard gluons are expected to be emitted during the parton cascade. We will refer to this scenario as **octet medium radiation**.

Finally, with `MSTP(148)=1` and `MSTP(149)=1` the g→gg splitting function is used again, but in the branching $c\bar{c}^{(8)}$ →$c\bar{c}^{(8)}$+g the c$\bar{c}$ and the g's have equal probability to be the hardest, i.e. $z < 0.5$ and $z > 0.5$ up to non-negligible phase space corrections for the

$c\bar{c}^{(8)}$ mass. It implies that a significant amount of hard gluons can be emitted before the hadronization process. This production case almost certainly overestimates showering activity since, in a g→gg branching, the phase space for the gluon with lower z later to branch into a $c\bar{c}^8$ ought to be kinematically lower than that for the gluon with higher z. Also, after the g→c$\bar{c}$ branching one would expect subsequent emissions to reflect the shape of the c→cg splitting kernel, which strongly favours $z > 0.5$. Thus having equal probabilities for $z < 0.5$ and $z > 0.5$ may result in an exaggarated activity around the J/Ψ. We will refer to this model as a **octet high radiation scenario**. Since this production case is the most challenging case from an experimental point of view, `MSTP(148)=1` and `MSTP(149)=1` are used in this work to keep conservative conditions. None of these cases is expected to produce the data by itself. The reality might be a combination of these models, and only experimental observation should clarify the situation.

3.1.4 The J/Ψ Polarization

The most recent measurements from Tevatron data are not in agreement with theoretical predictions and with what Hera-B found (which is instead in agreement with theory, at least for low J/Ψ p_T). Anyway, in `Pythia` it is possible to simulate polarization states. Kinematics acceptance of the detector is a function of $p_T^{J/\Psi}$ and so it depends upon the J/Ψ polarization, which influences heavily the leptons p_T spectrum. So in the simulation it can be useful to generate a sample with polarized J/Ψ's to understand systematic effects in reconstruction efficiencies determination (see next chapters).

The polarization switches in `Pythia` are `MSTP(145-147)`. With `MSTP(145)=0` non-polarized quarkonium is produced. If `MSTP(145)=1` and `MSTP(147)=0` (i.e. helicity is equal to 0) is used to generated longitudinally polarized quarkonium and if `MSTP(147)=1` (helicity= ±1) is used to generated transversely polarized quarkonium. In this work, `MSTP(145-147)` have been set to *zero*, i.e. no polarization, as Hera-B data seem to confirm. In the generation, the polarization can be adapted at a later stage by event-reweighting similar to the reweighting in Fig. 3.1.

Other general `Pythia` switches, not specifically related to quarkonia, such as PDFs, running of α_s, etc., are treated as follows: the used NRQCD elements are determined from CDF data with the parton distribution function of **CTEQ5L**, so the same PDFs are used here. We made use of the default Underlying Event `Pythia` settings that are used for most of the Monte Carlo productions in CMS[66].

3.2 Used Monte Carlo Samples

3.2.1 Prompt J/Ψ events

Following the issues explained above, about 170000 prompt J/Ψ events were generated with `Pythia6.416v` at 10 TeV as centre of mass energy, in NRQCD theory. J/Ψ's are forced to decay in e^+e^- pairs switching the `Pythia` flag `MDME(858,1)` on. Due to the restrictive requests of the detector (as it will be explained in the next chapters), at generator level only events that have e^+e^- pairs with the following kinematic criteria were retained for the detector simulation:

$$p_T(\mathrm{e}) > 3.5\mathrm{GeV/c}, \quad |\eta(\mathrm{e})| < 2.5 \tag{3.6}$$

where $p_T(\mathrm{e})$ is the transverse momentum of a single electrons of the pair and $|\eta(\mathrm{e})|$ is its pseudorapidity.

The effective cross section for prompt J/Ψ of this sample can be estimated by Eq. 3.7:

$$\sigma_{prompt} = \sigma_{pp\to c\bar{c}\to J/\Psi+X} \cdot BR_{J/\Psi\to e^+e^-} \cdot \xi_{kine} \tag{3.7}$$

where $\xi_{kine} = 0.71\%$ is the probability of an event to pass the kinematic cuts of Eq. 3.6. Taking into account the experimentally measured value[63]

$$BR_{J/\Psi\to e^+e^-} = (6.02 \pm 0.06)\% \tag{3.8}$$

Being the total cross section $\sigma_{pp\to c\bar{c}\to J/\Psi+X}$ computed by `Pythia` equal to 78.50 μb[1] (including also the feed-down from $\chi_c \to$J/Ψ γ decays), the total effective cross section is estimated as:

$$\sigma_{prompt} \simeq 33.40\mathrm{nb} \tag{3.9}$$

corresponding to a "visible"[2] sample of ∼33400 events/pb^{-1}.

In a similar way, also prompt J/Ψ events in CSM are generated as a check sample in the analysis, to show how it is possible to discriminate various model using the differential cross section of prompt J/Ψ events as a function of J/Ψ's p_T as observable, and also events in NRQCD with "low radiation scenario" to study systematics in the analysis due to different models used in the Monte Carlo generator. The same kinematic preselections on e^+e^- pairs are used. CSM prompt J/Ψ's are obtained switching on only sub-process with colour singlet charmonia produced, i.e. `ISUB=421`, `ISUB=431-439`.

[1] The total cross section for general charmonia production computed with `Pythia` is $\sigma_{pp\to c\bar{c}} = 286.3$ μb.

[2] It mean events that CMS detector can measure, i.e. in the kinematic acceptance explained in Eq. 3.6.

For the last two samples, the cross section computed by Pythia is the same of above, because the ME computation is the same. For CSM sample, the effective cross section it is estimated always by Eq. 3.7, remembering that in this case it is $\sigma_{pp\to c\bar{c}\to J/\Psi+X} = 44.26$ μb and $\xi_{kine} = 0.23\%$, so

$$\sigma^{CSM}_{prompt} \simeq 6.128\text{nb} \tag{3.10}$$

corresponding to a visible sample of ∼6100 events/pb^{-1}.

In Fig. 3.3 there are compared the differential cross section of (NRQCD) prompt J/Ψ events as a function of J/Ψ's p_T and JΨ's η spectrum, comparing the cases of samples with or without kinematic preselections on e$^+$e$^-$ pairs. η distribution is not peaked in 0 but present a double symmetric peak because the produced J/Ψ's are not relativistic (i.e. $p_T \sim m_{J/\Psi}$) and so η cannot be considered as rapidity y (which is always peaked in 0) in first approximation. In Fig. 3.4 there is the comparison between the differential cross section distribution of prompt J/Ψ's in the case of "high radiation scenario" and "low radiation scenario" in NRQCD and of CSM case, and the "visible" comparisons (i.e. requiring kinematic preselections on e$^+$e$^-$ pairs). Even if the distributions in the case "high radiation scenario" and "low radiation scenario" seem to be very similar, the number of particles per event changes in the two cases, from ∼200 in the second case to ∼600 in the first case. This has the consequence that detectors have different occupancy in the two cases, and so this is a systematic effect to take into account in the analysis (see next chapters).

In Fig. 3.5 there are shown the hardest electron p_T spectrum Vs. the softest electron p_T spectrum and the electrons η spectrum, comparing the NRQCD case and the CSM case.

3.2.2 B→J/Ψ+X events

With Pythia are generated also non-prompt J/Ψ's events, i.e. J/Ψ's from weak decays of a B-meson, as shown in Fig. 3.6. Here, b quarks are generated in b$\bar{\text{b}}$ pairs[3] in pp collisions. When b$\bar{\text{b}}$ pairs are generated, the Pythia parton shower hadronizes the pairs into B-mesons (B$^+$, B^0, B$_s$, B$_c$) or baryon (Λ^0_{b}). In the simulation, B-mesons (baryon) are then forced to decay in $J/\Psi + X$. Events where χ_c states are directly produced during parton shower via the diagram in Fig. 3.6 are also included, forcing then the decay $\chi_c \to J\Psi + \gamma$. J/Ψ's are so forced to decay in e$^+$e$^-$ pairs, with the same kinematic preselections used in the prompt J/Ψ's samples.

The effective cross section for non-prompt J/Ψ's events can be computed (remembering

[3] b quarks are not valence quarks in protons, so they can be produced only in pairs.

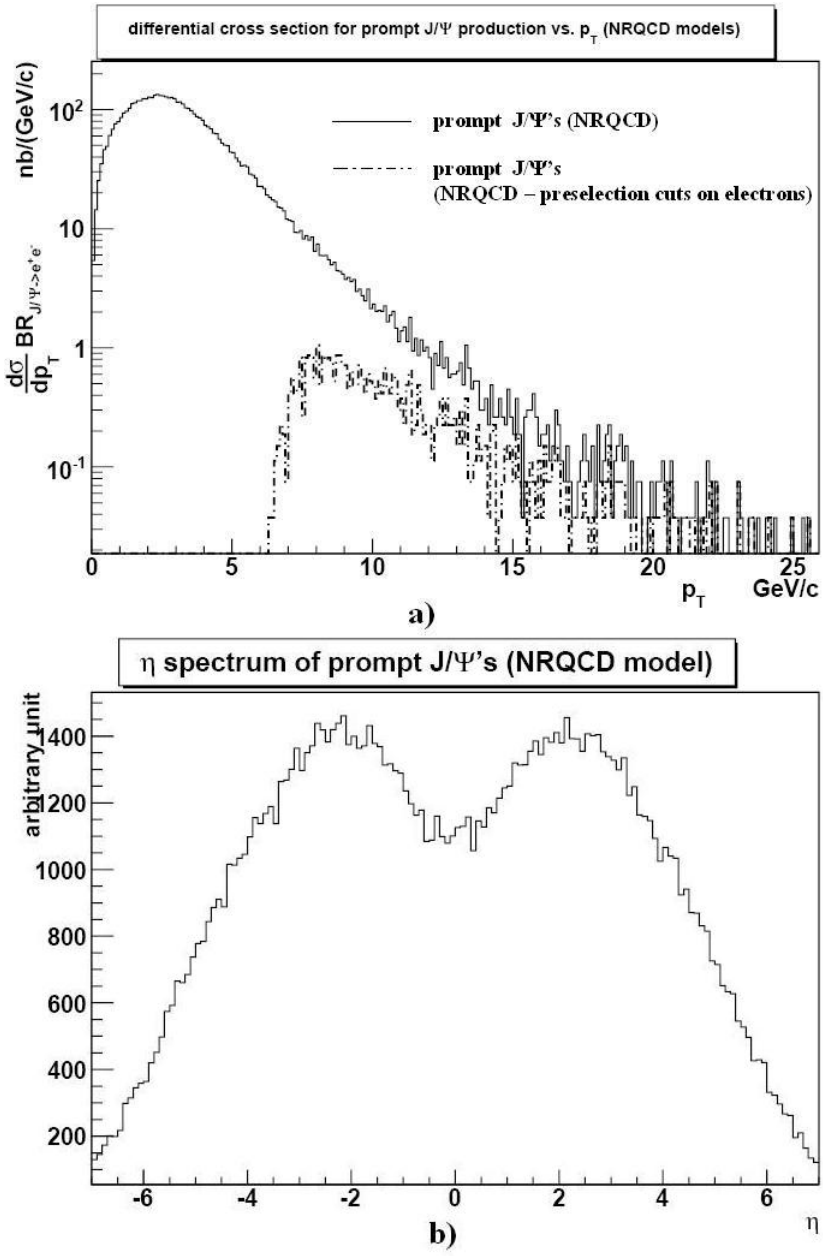

Figure 3.3: *a) Differential Cross Section for prompt J/Ψ's as a function of J/Ψ's p_T (compared to the "visible" distribution obtained requiring kinematic preselections on electrons). b) η distribution of prompt J/Ψ's.*

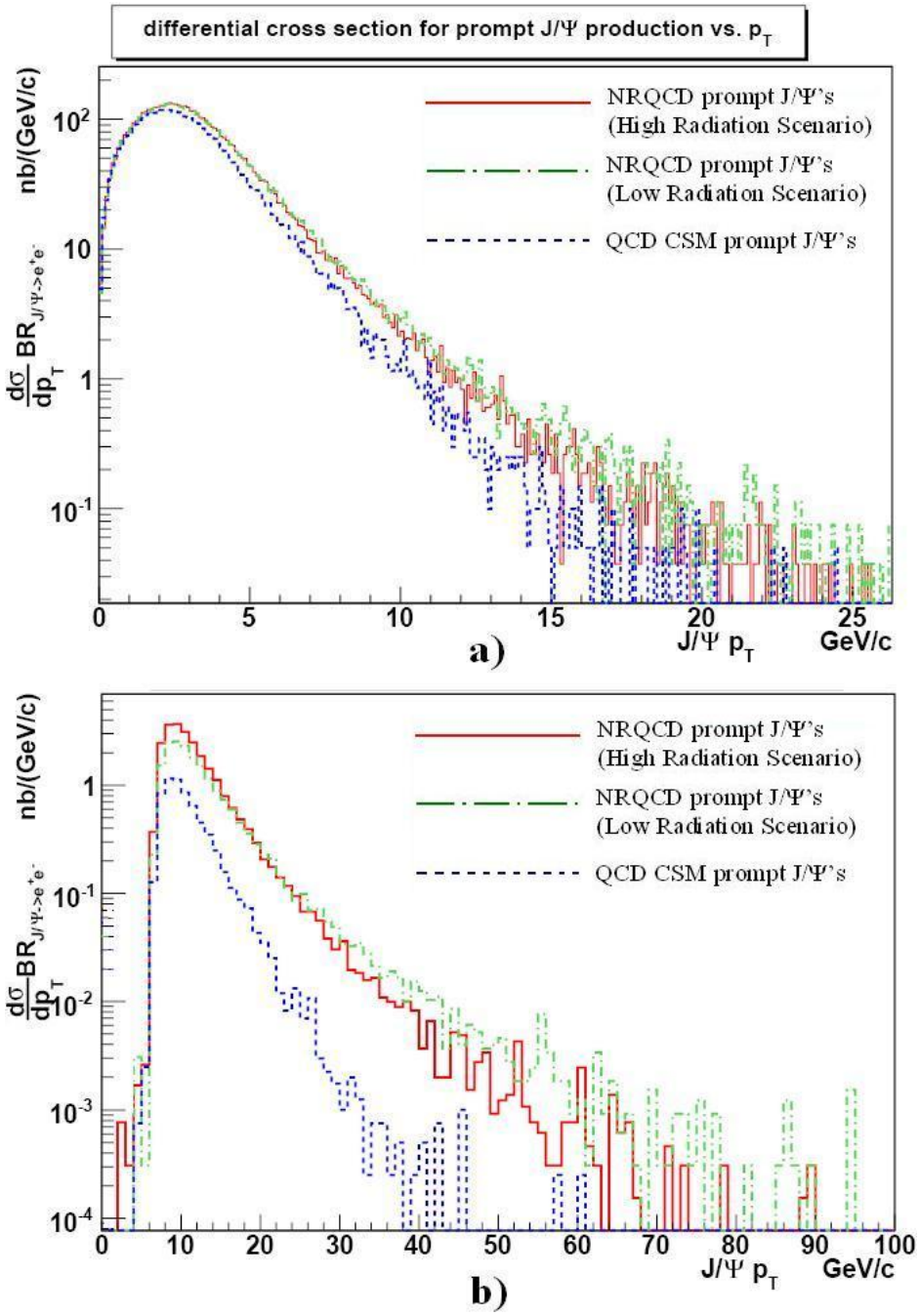

Figure 3.4: *a) Differential cross section for prompt J/Ψ production in the cases of different models used in the generation with* `Pythia`*. b) The "visible" distributions obtained requiring preselection cuts on e^+e^- pairs.*

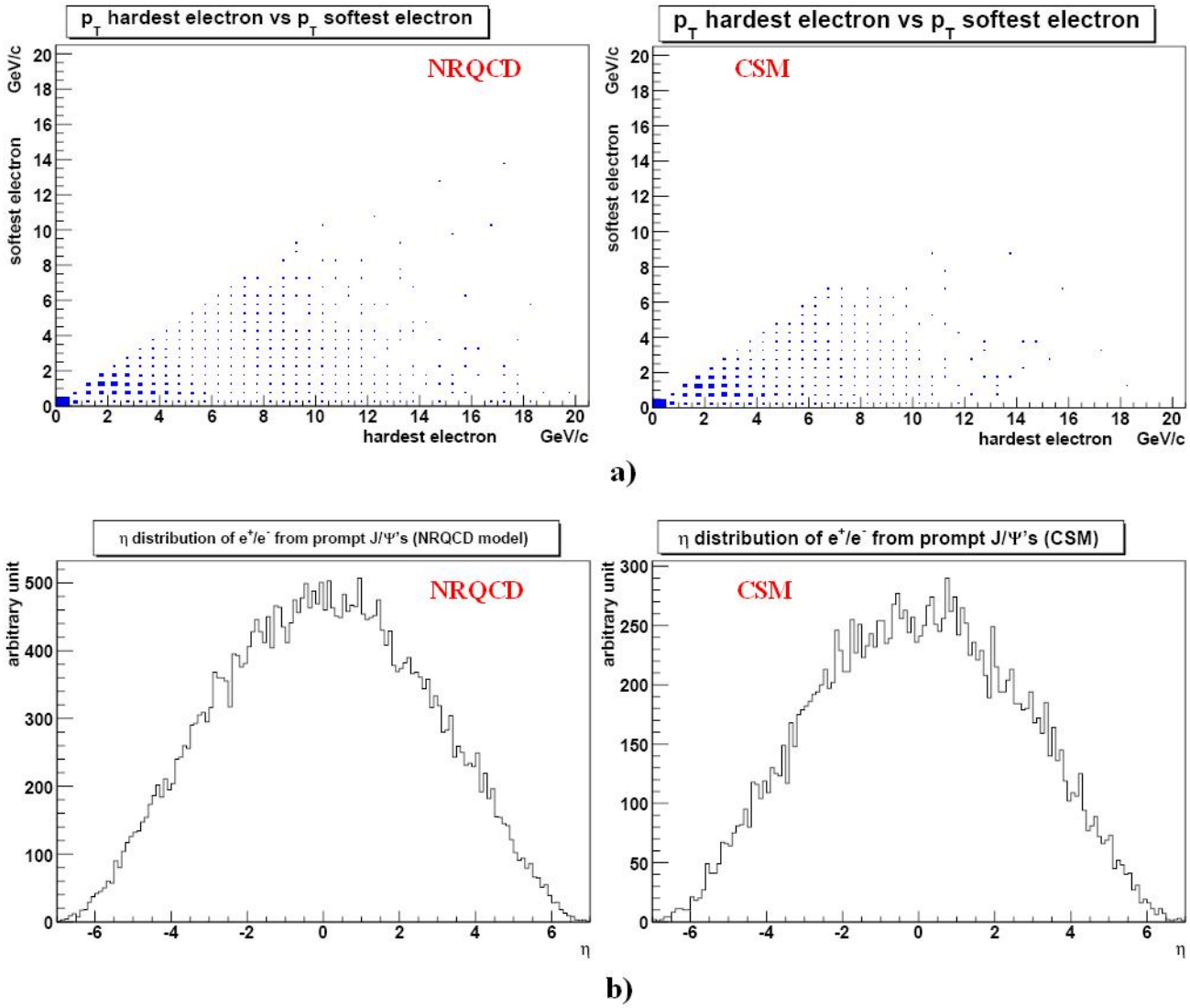

Figure 3.5: *a) p_T distribution of the e^+e^- pairs (hardest vs. softest). b) η distribution. The comparison is between the samples obtained using NRQCD and CSM simulations. The two Monte Carlo Sample are referred to the same luminosity.*

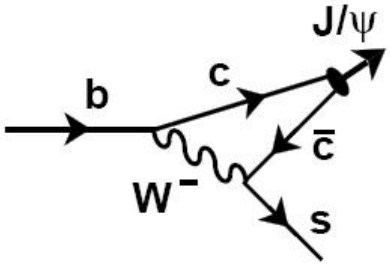

Figure 3.6: *Feynman diagram of a weak-decay of a B-meson.*

the branching ratios of the forced decays) as

$$\sigma_{nonprompt} = 2 \cdot \sigma_{pp \to b\bar{b}} \cdot [BR_{b \to directJ/\Psi} + (BR_{b \to \chi_c} \cdot BR_{\chi_c \to J/\Psi\gamma})] \cdot BR_{J/\Psi \to e^+e^-} \cdot \xi_{kine} \quad (3.11)$$

where $\sigma_{pp \to b\bar{b}}$ is computed by Pythia (330.7 μb), and BR's[63] are

$$BR_{b \to directJ/\Psi} \simeq 0.79\%, \quad BR_{b \to \chi_c} \simeq 1.32\%, \quad BR_{\chi_c \to J/\Psi\gamma} \simeq 27.3\% \quad (3.12)$$

The kinematic preselection filter efficiency $\xi_{kine} \simeq 4.32\%$. So the visible cross section at CMS will be

$$\sigma_{nonprompt} \simeq 26.2\text{nb} \quad (3.13)$$

corresponding to a visible sample of ∼26200 events/pb^{-1}.

In Fig. 3.7 there are compared the differential cross section of non-prompt J/Ψ events as a function of J/Ψ's p_T and JΨ's η spectrum, comparing the cases of samples with or without kinematic preselections on e$^+$e$^-$ pairs. In Fig. 3.7 there are shown the hardest electron p_T spectrum Vs. the softest electron p_T spectrum and the electrons η spectrum. Compared with the cases of e$^+$e$^-$ pairs from prompt J/Ψ's in Fig. 3.5, electrons from non-prompt J/Ψ's seem to be with harder p_T spectrum.

3.2.3 Minimum Bias and QCD events

This sample represents the big amount of background in all physics analysis in the CMS experiment (and, generally, in all hadronic collisions studies), specially in analysis with electrons in the final state, as it will be explained better in the next chapters. Compared to those events, other events topologies[4] are negligible.

Minimum Bias events are events where protons don't have hard interactions, but partons exchange soft gluons to re-combine together into others hadrons (elastic scattering and diffractive interactions). Due to the relatively low p_T exchange, Minimum Bias events

[4]In particle physics, the *topology* term is used meaning the kind of the interaction and what are the particles in the final states and in which region of the phase space they are.

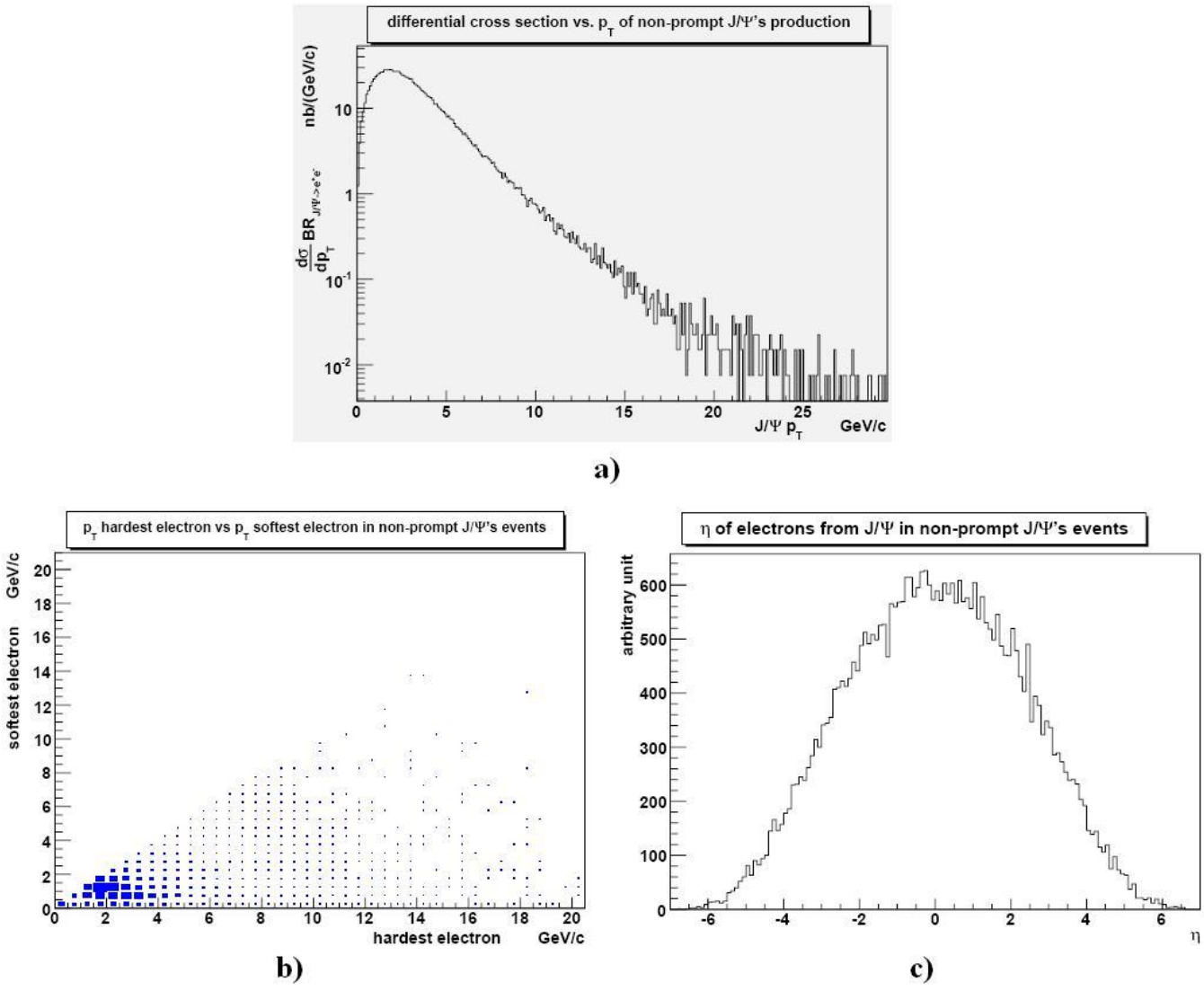

Figure 3.7: *a) Differential Cross Section as a function of p_T for non-prompt J/Ψ's production. b) p_T distribution of the e^+e^- pairs (hardest vs. softest) in non-prompt J/Ψ events and c) their η distribution.*

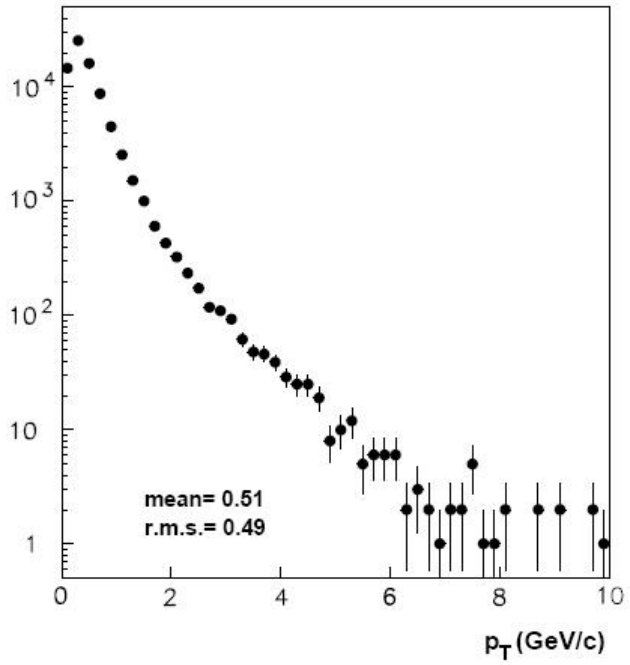

Figure 3.8: *p_T distribution of particles in Minimum Bias events.*

presents particles with low p_T in the final state, as it is shown in Fig. 3.8. Anyway, the cross section of low p_T production events is really huge ($\sigma \sim \mathcal{O}(50$ mb) at 10 TeV) and $\sim$50 billions of events are expected per pb^{-1}, so Minimum Bias events are very important from a statistical point of view. *QCD events* sample instead include events where partons interact via strong interaction. In `Pythia` those two samples are obtained in the same way. The flag `MSEL=1` pass to the generator the hard-encoded cross sections of the 2→2 processes shown in Tab. 3.3. So `Pythia` produces hard 2 jets QCD processes at the LO. Anyway, 2→2 LO diagrams brings to divergences in the ME computation at low $\hat{p}_T$[5]. `Pythia` avoids this problem introducing a cut-off in the ME computation: if the $\hat{p}_T$ is less than a flag (`PARP(82)`, set to $\sim$1.9 GeV/c), then low p_T production events (i.e. Minimum Bias) are generated.

In `Pythia` there are two flags, `CKIN(3)` and `CKIN(4)`, which are the limits of the $\hat{p}_T$ range used in the generation and that can be customized by the user. If one set `CKIN(3)=0` and `CKIN(4)=-1` the generation with $\hat{p}_T \in [0, +\infty)$ is suited, so one can obtain a background sample with Minimum Bias and QCD events correctly weighted with the cross sections computed by `Pythia`.

Usually, in the CMS collaboration, Minimum Bias and background sample are generated

[5]The definition of $\hat{p}_T$ has given in Chapter 1. In 2→2 processes generally it coincides to the p_T exchanged in the interaction, e.g. in $q_i + \bar{q}_i \to q_k + \bar{q}_k$ it is the p_T of the exchanged gluon

ISUB	2→2 QCD subprocesses
11	$q_i + q_j \rightarrow q_i + q_j$
12	$q_i + \bar{q}_i \rightarrow q_k + \bar{q}_k$
13	$q_i + \bar{q}_i \rightarrow g + g$
28	$q_i + g \rightarrow q_i + g$
53	$g + g \rightarrow q_k + \bar{q}_k$
68	$g + g \rightarrow g + g$

Table 3.3: *2→2 events production sub-processes and their corresponding ISUB indices in Pythia.*

$\hat{p}_T$ bins (GeV/c)	$\sigma_{pp\rightarrow X}$ @ 10 TeV (μb)	weights
$0-15$	51000	95.00%
$15-20$	1500	2.60%
$20-30$	630	1.10%
$30-50$	160	0.26%
$50-80$	22	0.04%
>80	3.1	<0.01%

Table 3.4: *2→2 events production sub-processes and their corresponding ISUB indices in Pythia.*

officially by a proper generation team in big quantity and in differents bins in $\hat{p}_T$, due to the high cross sections of this sample. In Tab. 3.4 there are reported the cross section values of the different $\hat{p}_T$ bins samples computed by Pythia, and the corrispective weight in the global Minimum Bias + QCD events sample. The $0-15$ bin sample contain almost all Minimum Bias events, and also $15-20$ bin sample contain ∼2% of events which are Minimum Bias[67]. Each bin sample presents different hadrons p_T distribution; in particular, increasing $\hat{p}_T$ bin means increasing p_T of particles in hadronic jets. This is the reason of the choice of generation with different $\hat{p}_T$ bins[6].

More informations about the topology of this background sample will be given in the next chapters, where the analysis is explained with more details, and it will be possible to understand how Minimum Bias + QCD events represent a background in prompt J/Ψ's physics.

[6]Usually, as it will be shown in the next chapters, events with higher $\hat{p}_T$ gives a bigger contribution to the background in the various physics analysiss

Chapter 4

Reconstruction of $J/\Psi \rightarrow e^+e^-$ events with the CMS detector

As it was shown in Chapter 3, the reconstruction of $J/\Psi \rightarrow e^+e^-$ events is a big challenge in an experiment like CMS, due to the low-p_T of the electrons from the $J/\Psi \rightarrow e^+e^-$ decay - $\mathcal{O}$(GeV/c) - in the decay ($m_{J/\Psi} \simeq 3.097$ GeV/c^2) and, at the same time, to the huge amount of background produced in the p-p collisions at the LHC energies. As it was said in advance in Chapter 2, the two sub-detectors involved in electron reconstruction in CMS are the tracker (being $e^{\pm}$ charged particles) and the electromagnetic calorimeter (ECAL). The electron reconstruction procedure is the following:

- Reconstruction of the electromagnetic energy deposit in ECAL.
- Reconstruction of charged tracks in the tracker, starting from the single reconstructed *hits* in each layer.
- Find some criteria to match ECAL deposits with charged tracks, discriminating energy deposits and charged tracks left by other particles such as muons (μ), photons (γ) or charged hadrons (specially pions - π).

So it is now clear why the reconstruction of $J/\Psi \rightarrow e^+e^-$ events, and of low-p_T electrons more in general, is so "hard": the low-p_T makes difficult a good reconstruction of tracks in the big magnetic field of CMS and of ECAL energy deposits. Then, not only events topologically similar to the signal (i.e., with real e^+e^- pairs in the final state) can contribute to the background, but also events which can have a "good" match between ECAL deposits and charged tracks, specially charged hadrons produced in the p-p collisions which can be seen as *fake* electrons falling in the same phases space region of the e^+e^- pairs of the signal. This is the reason why Minimum Bias and di-jet QCD events are included as background in prompt $J/\Psi \rightarrow e^+e^-$ analysis.

Anyway, some algorithms have been developed by the CMS collaboration to effectively reconstruct electrons in the "busy" LHC environment. They are thought specially for the reconstruction of $Z^0 \rightarrow e^+e^-$ events (and more in general for events with electrons from decays of bosons heavier than J/Ψ's), where electrons have $p_T \sim \mathcal{O}(45\ \mathrm{GeV/c})$. So in this work it is introduced an alternative algorithm developed properly for $J/\Psi \rightarrow e^+e^-$ events (in particular, for prompt J/Ψ's).

In order to tune the "dedicated" algorithm Monte Carlo techniques are used, where (signal and background) events are first generated (Chapter 3) and then passed to the detailed simulation of the detector. The CMS collaboration has developed a software package for these purposes, which will be described in this chapter. Here are implemented also the algorithm of reconstruction of physical objects (from lower level, tracks, calorimeters deposits, etc., to higher level, $e^{\pm}$, $\mu^{\pm}$, γ, missing-E_T, etc.).

Then, a comparison between this algorithm and the official one used in the CMS collaboration is presented to show which is the one that gives a better signal/background ratio. Other issues relative to triggers will be described in Chapter 5, and more specific algorithms and selections used specifically for prompt J/Ψ events will be described in Chapter 6.

4.1 The Simulation and Reconstruction Software of CMS: `CMSSW`

The CMS simulation and reconstruction software, `CMSSW`[68], is a `C++`[69] framework that can be configured via `Python`[70] 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`[71] files (more informations about `ROOT` framework will be given in Chapter 6), and are thus directly browsable in `ROOT`. The EDM includes the dedicated services which are needed by simulation and reconstruction, and has also stored detector calibration and alignment condition. A schematic overview of the software framework is given in Fig. 4.1. The EDM is capable to hold different types of data from Monte Carlo simulations, raw detector readout, or reconstructed high level objects needed for the physics analysis. The CMSSW framework is modularized in the way that only necessary parts of the software are loaded for the individual job. During processing, data are passed from one module to the next via the `Event` class, and are accessed only through the `Event`. The output data is capable

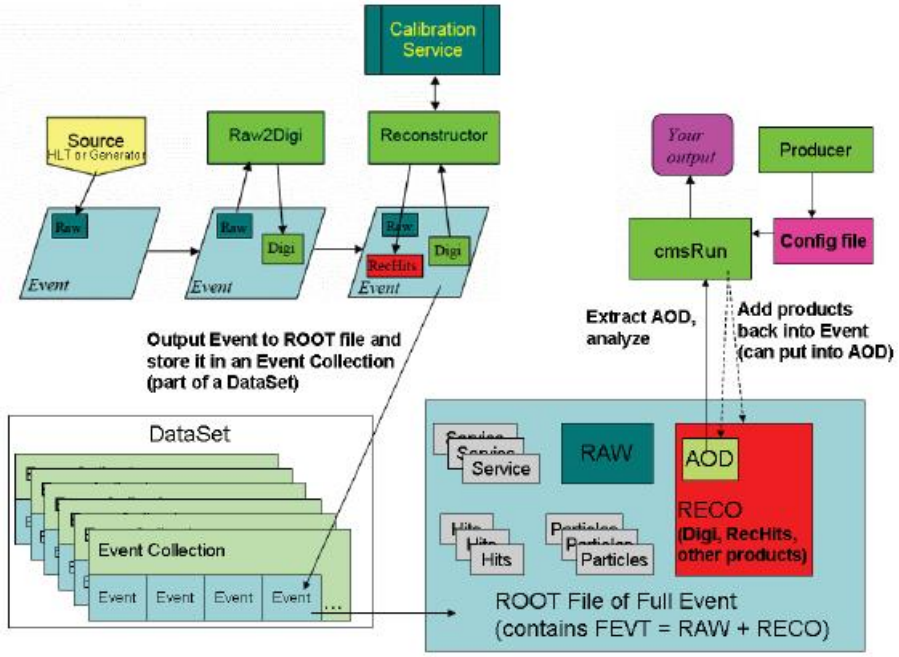

Figure 4.1: *A schematic overview of the* `CMSSW` *software framework.*

to hold different types of data from Monte Carlo simulations, raw detector readout, or reconstructed high level objects needed for the physics analysis.

CMSSW can be run feeding the desired `Python` configuration script into the executable `cmsRun`. The configuration file contains the modules (where are implemented the single 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 plug-in manager, finds out in which libraries the modules to be run are defined and loads them.

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

- `Source`: these modules are used either to load events from a `ROOT` file or to produce events running an event generator (such as `Pythia`, described in Chapter 3).
- `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 `C++ bool` 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, e.g., 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`, such as detector geometry and topology; these information is stored in the `EventSetup` object.

4.1.1 Events Generation, Detector Simulation and Reconstruction with `CMSSW`

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`[72] 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 (this is called the "`GEN`" step) the simulation of detector follows ("`SIM`" step). 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 p−p impaction point position distribution per bunch crossings. As it was explained in Chapter 2, a good modelling for the smearing is a three-dimensional gaussian with $\sigma_x, \sigma_y \sim 10-20$ μm and $\sigma_z \sim 15$ cm. The next step is the simulation of the interaction of particles with the detector. The description of these interactions is achieved using `GEANT4`[73]. Once energy deposits and multiple scattering effects in the CMS subdetectors are simulated (as well as other effects, such as *bremsstrahlung* for electrons passing through the detector material), the simulation of the signals produced by the subdetectors follows and it is passed to the electronics for the conversion into a numerical format. This step is knows as "digitization" of "raw" data (`RAW-DIGI`) (see Fig. 4.1). Then, reconstruction algorithms can be applied to obtain the so called *RecHits* from subdetectors (e.g. hits in tracker layers, position and deposited energy of a crystal interested by an electromagnetic shower in ECAL, etc.), and starting from them also higher level physics object can be obtained (tracks reconstruction, particle identification, etc.). This final step is called "`RECO`".

The chain described so far is often referred to as "full simulation" chain (*FullSim*). The most time consuming step of the full simulation is actually the simulation of detector effects and radiation interactions using `GEANT4`; the time needed to fully simulate one event can amount to several minutes. For this reason also a "fast simulation" (*FastSim*) of the detector effects was set up. In the fast simulation the `GEANT4` 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 parameterized functions that aim at reproducing the full simulation result.

Concerning the simulation of the hits in the tracker detector in the FastSim, 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 parameterized functions. The hit position and error are then assigned with a Gaussian smearing with respect to the analytically calculated impact point.

Calorimeter energy deposits in the FastSim are obtained in two steps. First the shower is simulated as if the calorimeter was 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].

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

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.

4.2 The Standard Electron Reconstruction in `CMSSW`

The version of the CMS software used in this work is the `CMSSW_2_1_X`. Here, the default electron reconstruction relies on the combination of both tracker and calorimetric informations. Three main steps can be identified in the electron reconstruction procedure: the energy "clustering" in ECAL, the matching between the pixel hits and the ECAL *clusters* to provide the track seeds and finally the inward-outward track reconstruction. The algorithms for these three steps are implemented in the `CMSSW` Electron Photon

[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.

package and they are currently employed both for the HLT and for the offline analysis.

4.2.1 The Calorimetric Energy Reconstruction

The clustering of the energy deposits in the electromagnetic calorimeter is the first step of the electron reconstruction procedure. A single electron generates a shower which develops in more than one crystal and the bremsstrahlung emission (due to the passage of electrons through the tracker layers) is responsible for a further spread of the energy in the bending plane, i.e. along the ϕ direction. It is therefore necessary to collect the energy of all the crystals which are involved.

The starting point is the search for crystals with energy above a certain threshold, the so called **seeds**. The *seeds* are ordered with decreasing energy and only the most energetic seed is kept among the adjacent ones. Starting from each seed, the energy deposits are then collected using two different reconstruction algorithms[74]: the `Hybrid algorithm`, used by default in the barrel, and the `Island algorithm`, employed in the endcaps. Two different algorithms are needed to consider the different geometry of ECAL crystals and map of the magnetic field in the barrel and endcap regions.

The *Island algorithm* moves in both directions in ϕ starting from the seed position and it collects all the crystals until it finds an energy rise compared to the previous crystal collected (this reasonably means that crystal belongs to another cluster) or a hole. Then, it moves one step in η from the seed position and makes another ϕ search. The same conditions used in the ϕ search also stop the search in η; the algorithm then comes back to the seed and it moves in the opposite η direction. A bremsstrahlung recovery procedure is then applied to create the so called **superclusters**, groups of clusters which collect the energy released both by the electron and by the emitted photons. The Island algorithm includes each crystal in one cluster only, to avoid double countings of energy. In this way, the energy deposits in the crystals which are below the threshold may remain unclustered, with a significative energy loss. On the other hand, small deposits of energy due to the noise or to pileup events are not clusterized either.

The *Hybrid algorithm* exploits the knowledge of the lateral shower shape in η and searches for bremsstrahlung energy in ϕ, requiring no *seeds*. First rows of 3 or 5 crystals along η direction are created and collected, then adjacent crystals in ϕ direction are clustered. At each stage, the clusters are required to have energy above a certain threshold. The Hybrid algorithm, a sort of super-clustering algorithm with different clustering steps, is designed to reconstruct relatively high energy electrons in the barrel; small deposits of energy are better described by the island clusters.

After the clustering, the electron energy is computed as the sum of the deposits in the crystals belonging to a supercluster. A correction factor depending on the number of

crystals in the supercluster is applied, to minimize the residual dependence of the energy scale both on the energy and on the pseudorapidity.

The position of the Super Cluster (*centroid*) is obtained by a weighted mean of the single crystals position:

$$\vec{x} = \frac{\sum_i W_i \vec{x}_i}{\sum_i W_i} \tag{4.1}$$

The energy density of an electromagnetic shower decreases approximatively such as an exponential moving transversely from the centre of the shower, so weights are defined as logarithms in the energy:

$$W_i = \ln\Big(\frac{E_i}{\sum_j E_j}\Big) \tag{4.2}$$

4.2.2 The Pixel Matching

The *superclusters* which are reconstructed in the ECAL are used as starting points for the electron track reconstruction. First, the supercluster position and energy are used to compute the electron transverse momentum. The expected hit position in the innermost pixel layer is estimated propagating the electron candidate from the supercluster position to the nominal vertex, taking into account its transverse momentum and the magnetic field. A search area is constructed around the expected point using the error in the ϕ measurement of the cluster and the spread of the vertex z. If a compatible hit is found in such region (within a loose cone with $\Delta\phi = 0.25$ and $\Delta\eta = 0.15$), it is used together with the supercluster to estimate the true vertex z position. The track is then propagated from this point to the second layer and if a second compatible hit is found the algorithm matching proceeds. This kind of search is performed for both charge hypotheses. Three possible hits are requested to start the electron reconstruction. The output of this search are the seeds which are used to start the track reconstruction.

4.2.3 The Track Reconstruction

Once the track seed has been identified, the full track reconstruction is performed going from the seed towards ECAL. In order to guarantee a good momentum resolution and high efficiency, the minimum number of hits required to keep the track is 5 and the χ^2 cut to check the compatibility of the hit which is found at each layer with the one which is expected is 30.

The algorithm used for the reconstruction is the **Kalman Filter**, which is briefly exposed below:

The Kalman Filter

The starting point is the theoretical model of a charged relativistic particle in a mag-

netic field. The trajectory is given by the equation of motion obtained by the Lorentz equation[75]:

$$m\gamma\frac{d^2\vec{x}}{dt^2} = kq\vec{v}\times\vec{B} \tag{4.3}$$

where k is a coefficient which takes into account the used units of measurement, $\vec{x}(t)$ is the vector position at the time t, q is the charge of the particle, m is the mass, γ is the relativistic coefficient, and $\vec{B}$ is the magnetic field.

In order to write the equation in terms of geometrical quantities eliminating the parameter t, the curvilinear arc length $s(t)$ was introduced: remembering that

$$\frac{d\vec{x}}{dt} = \frac{d\vec{x}}{ds}\cdot\frac{ds}{dt} = \frac{d\vec{x}}{ds}\beta c \Longrightarrow \frac{d^2\vec{x}}{dt^2} = \frac{d^2\vec{x}}{ds^2}(\beta c)^2 \tag{4.4}$$

The Eq. 4.3 can be written as follows

$$\frac{d^2\vec{x}}{ds^2} = \frac{kq}{|\vec{p}|}\frac{d\vec{x}}{ds}\times\vec{B} \tag{4.5}$$

where $\vec{p} = m\gamma\vec{v}$ is the relativistic momentum of the particle.

The Eq. 4.5 is a three-dimensional second order differential equation, and its solution is an helix with symmetry axis oriented to the direction of $\vec{B}$, so the minimum number of parameters to individuate a trajectory is **five**, called *the curvilinear parameters*[76].

So tracks are described by a five-dimensional state vector $\vec{y}$ containing the information about the momentum, the direction and the position at each point of the trajectory. The state vector can be written as a function of a coordinate $\vec{y} = \vec{y}(z)$ and its evolution as a function of z can be described by a set of differential equations. It is anyway sufficient to consider the state vector in a discrete set of points, like the intersections of the track with the detector. In this way the problem reduces to a discrete system of equations

$$\vec{y}(z_k) \equiv \vec{y}_k = \vec{f}_{k-1}(\vec{y}_{k-1}) + \vec{w}_{k-1} \tag{4.6}$$

where $\vec{f}_{k-1}(\vec{y}_{k-1})$ is the track propagator from the detector $k-1$ to the detector k and the random variable $\vec{w}_{k-1}$ is the process noise which incorporates a random disturbance of the track between z_{k-1} and z_k. In general, the state vector y is not directly observed and the **quantities $\vec{m}$ measured by the detector which are functions of the state vector** are observed (i.e., tracker `RecHit`'s are function of the state vector) with a distortion due to the measurement noise (multiple scattering, etc.) $\vec{\epsilon}_k$

$$\vec{m}_k = \vec{h}_k(\vec{y}_k) + \vec{\epsilon}_k \tag{4.7}$$

and should be inserted in Eq. 4.6 in place of $\vec{y}$.

The track fitting with the Kalman Filter[77] proceeds throughout three kinds of operations: the **filtering**, i.e. the estimate of the state vector with a local measurement, the

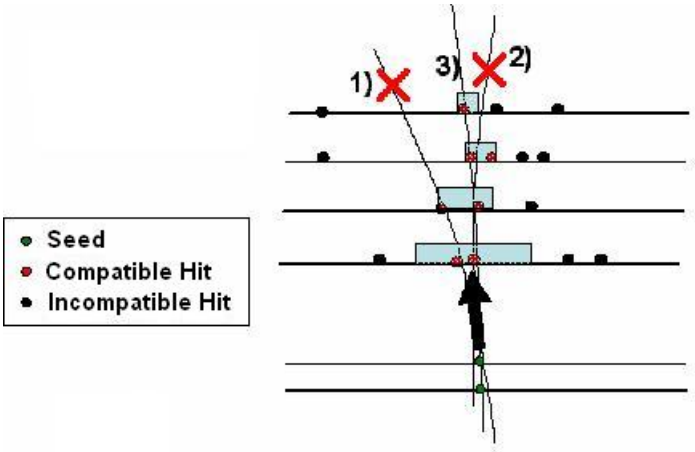

Figure 4.2: *Application of the Kalman Filter Algorithm to tracks reconstruction in the CMS tracker.*

prediction with the estimate of the state vector in the future, the **smoothing**, with the estimate of the state vector in the past, using all the measurements collected up to the present time. The track vector is extrapolated from the detector $k-1$ to detector k by means of the track model, then the extrapolated state vector is updated with the measurements on the detector k. The covariance matrix of the extrapolated state vector is computed by error propagation and the covariance matrix of the process noise between the detector $k-1$ and k is added to the propagated one. In order to improve the estimation of the state vector, the smoothing can be done running two filters both inward-outward and outward-inward (Fig. 4.2) and then combining both the predictions with the measured value.

In conclusion, the Kalman Filter is a least square estimator, but the big challenge is that the theoretical model is not a priori known and has to be obtained and updated every iteration k of the filter. Like all the least square estimators, it is the optimal filter when the system is linear and both $\vec{w}_k$ and $\vec{\epsilon}_k$ are Gaussian random variables, otherwise non linear filters can do a better job. The trajectory state at each layer is computed using both the predicted state and the measured hit. The compatibility is defined using the χ^2 of the residuals of the extrapolated track parameters on the new surface and the reconstructed hit on the same surface. If the χ^2 value exceeds a certain limit the hit is not taken. Not to lose efficiency at this stage, it was chosen to loose the tracker quality reconstruction requirements and to set the χ^2 cut to a very large value (100000). If many hits are found on a compatible layer, many candidate trajectories are grown in parallel. At the end only a maximum number of candidates per layer are kept.

The track reconstruction is started from pixels, where three aligned hits are used as starting *seed* for the Kalman Filter algorithm. The propagation function $\vec{h}_0(\vec{y}_0)$ is esti-

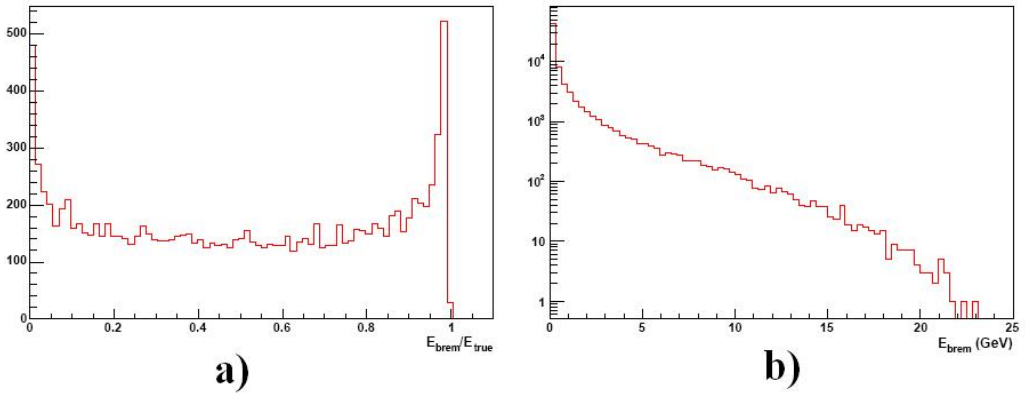

Figure 4.3: *Bremsstrahlung emission for 10 GeV/c p_T electrons in the ECAL barrel[56]. a) Fraction of the electron energy emitted by bremsstrahlung. b) Secondary particles energy. Only photons with energy higher than 10 MeV have been considered.*

mated fitting the pixels hits, then the next hit in the firsts tracker layers are searched (iteration $k = 1$) compatibly with the propagation function at iteration $k = 0$, and a new fit is made to update the propagation function to $\vec{h}_1(\vec{y}_1)$, and so on, as exposed before.

The Bremsstrahlung Problem

The main problem related to the electron reconstruction is the bremsstrahlung radiation in the tracker, which strongly affects both the momentum and the energy measurements. The material budget before the electromagnetic calorimeter, which is shown in Fig. 2.9 in Chapter 2 in term of radiation lengths, varies as a function of the pseudorapidity and it has its maximum of about 1.8 X_0 around $|\eta| = 1.5$. Such large amount of material, combined with the 3.8 T magnetic field, reflects in a significative bremsstrahlung emission. The importance of the effect is quantified in Fig. 4.3, which shows the fraction of energy emitted by bremsstrahlung in the tracker for 10 GeV/c p_T electrons in the ECAL barrel. About 50% of the electrons lose more than 50% of their energy and about 9% lose more than 95%. Fig. 4.3 also shows the energy of the secondary particles, which is often quite high. As can be expected, such large bremsstrahlung emission is directly related to the tracker material. In Fig. 4.4 the number of photons which are emitted is plotted as a function of the electron pseudorapidity for 30 GeV/c p_T electrons. The pattern clearly reproduces the tracker material budget.

The bremsstrahlung radiation affects both tracker and calorimetric measurements. Due to the kinematics of the bremsstrahlung emission, the photons move along the tangent to the electron trajectory. Their trajectory is a straight line while the electrons curve in the magnetic field and this results in a spread in the ϕ direction of the energy reaching

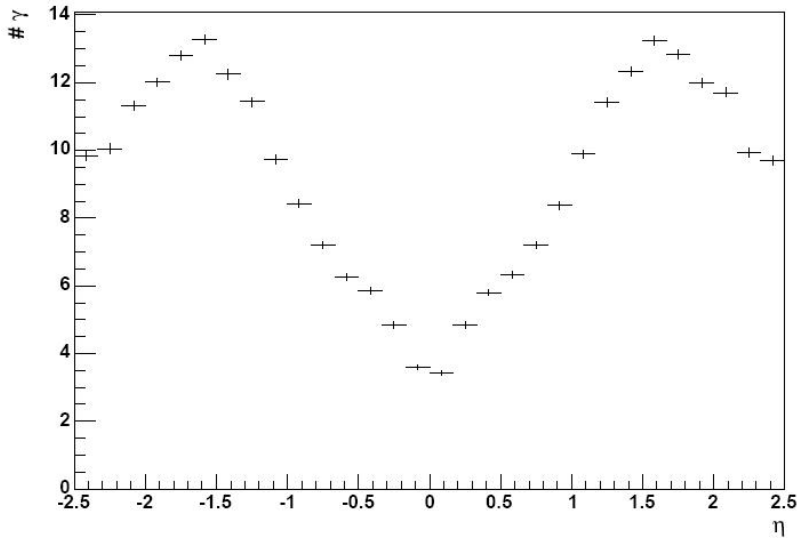

Figure 4.4: *Number of bremsstrahlung photons emitted as a function of the electron pseudorapidity for 30 GeV/c p_T electrons[56].*

ECAL. The impact points of the photon and of the electron on the ECAL surface are more distant if the photon is emitted at the beginning of the electron trajectory with respect to the case of late emission. The distance between the two impact points also decreases for high p_T. In case of late radiation the photon cluster partially merges with the electron one, so the loss in the energy measured with ECAL is lower than in case of early radiation.

For what concerns the tracker, when a photon is emitted the track gets more curved than predicted from the most probable value, hence biasing the estimation towards lower p_T values. This effect depends on the hardness of the photon which is radiated, so the tail is more evident for higher p_T tracks. Again, a late radiation has only a small effect on the reconstructed track parameters. Another important consequence of the bremsstrahlung emission on tracker measurements is the low number of hits in the tracks, which are often stopped by the hard χ^2 cut in the track updating when hard photons are emitted.

From the ECAL point of view, for instance, energetic clusters due to very hard photons are easy to be recovered by the superclustering algorithms, therefore a good energy estimation is possible despite such emissions.

4.2.4 Electron Track Reconstruction: the Gaussian Sum Filter Algorithm

It follows from what was discussed in the previous sections that the bremsstrahlung emission has to be carefully handled to have good performances in the electron reconstruc-

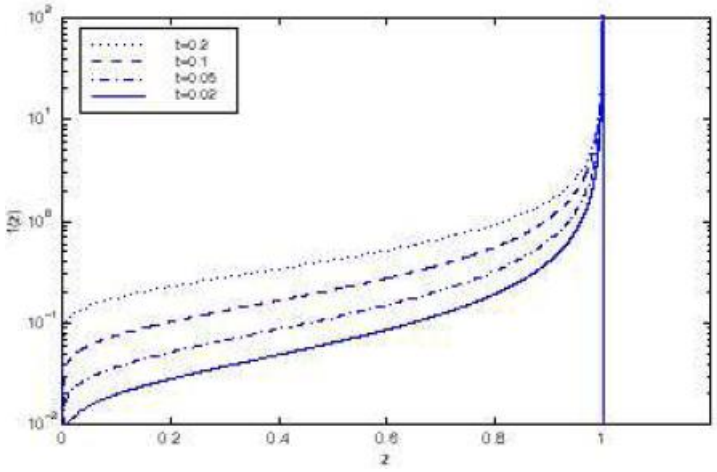

Figure 4.5: *Probability density function $f(z)$ for different values of the path length; z is the fraction of energy remaining after the material layer is traversed[79].*

tion both from the tracker and the calorimeter point of view. Bremsstrahlung recovery has to be efficiently done when reconstructing the energy in the ECAL; an important role is played also by the tracking algorithm which is used to reconstruct the tracks. In the following, will be discussed the one optimized for the electrons, which takes into account also bremsstrahlung: the **Gaussian Sum Filter**.

The Gaussian Sum Filter

The process noise $\vec{w}$ accounts both for the multiple scattering and the energy loss. In case of high energy electrons, the ionization loss can be neglected with respect to the energy loss due to bremsstrahlung[63]. The latter can be described using the Bethe-Heitler[78] model, in which the probability density function $f(z)$ of the electron energy loss is

$$f(z) = \frac{(-\ln z)^{c-1}}{\Gamma(c)} \tag{4.8}$$

where $c = t/\ln 2$, t is the thickness in units of radiation length of the material that the electron crosses, and z is the fraction of energy remaining after the material layer is traversed, and $\Gamma(c)$ is the Euler Γ-function[2]. The Bethe-Heitler probability density function, which does not depend on the particle energy but well reproduces the bremsstrahlung energy loss distribution, is shown in Fig. 4.5. While the multiple scattering can be well described with a single Gaussian, such approximation doesnt hold for the Bethe-Heitler distribution, so the use of the Kalman Filter is not properly correct in the case of electrons reconstruction. An adapted algorithm was developed for electrons tracks, called *Gaussian Sum Filter*[80], which is essentially a non linear generalization of the Kalman Filter

[2]In truth, in the Eq. 4.8 it is used the Γ notation introduced by Legendre: $\Gamma(c) = \int_0^{+\infty} c^{-1}\mathrm{e}^{-c}dt$.

and approximates the distribution of the process noise by a mixture of N Gaussians (Eq. 4.9). The resulting filter is a weighted (w_i) sum of Kalman Filters running in parallel, with weights which depend on the observations. Each Kalman Filter corresponds to one of the components of the mixture.

$$\sum_{i=0}^{N} w_i \frac{1}{\sqrt{2\pi}\sigma_i} \mathrm{e}^{-\frac{(z-\mu_i)}{2\sigma_i^2}} \equiv f(z) \tag{4.9}$$

In this way it is possible to fit the weights, means and standard deviations of each gaussian, and, at this level, Kalman Filter can be applied on single gaussian terms. Since the Gaussian Sum Filter corresponds to N Kalman Filter running in parallel, this brings to increase the combinatorial tries and so in a higher usage of CPU time than using the standard Kalman Filter. The Gaussian Sum Filter starts from pixels seeds found with the pixel matching of ECAL Super Clusters (SCs) as explained in section 4.2.2 to reduce the number of tries of GSF.

4.2.5 The `GSFElectron` Object: Electron Reconstruction in `CMSSW`

The electron reconstruction is made starting from the SCs in ECAL, then pixels seeds are found searching within a loose cone from SCs, so Gaussian Sum Filter algorithm is applied to reconstruct tracks. So it is obtained an object called `GSFElectron` which contains all informations about the SC and the reconstructed matched track. During the match, also other informations are obtained such as the $\Delta\phi$ and $\Delta\eta$ between the SC centroid and the track extrapolated at vertex, or the E_{SC}/p_{track} ratio.

In Fig. 4.6 it is shown the global electron reconstruction efficiency as a function of the pseudorapidity obtained on a MC sample of single electrons at different p_T. The drops at $|\eta| = 1.5$ and $|\eta| = 2.5$ are due to the crossing from ECAL barrel region to the ECAL endcap region and to the coverage of the tracker respectively.

4.2.6 The Electron Identification

The efficiency in Fig. 4.6 is referred to the `GSFElectron` reconstruction; additional selection criteria are applied to define a robust electron identification strategy (e-ID). At hadron colliders a huge QCD jets background is expected. Besides, simple electrons could be faked by processes such as $\pi^0\pi^+$ overlap and π^0 conversions or multiple electron candidates corresponding to a single supercluster could result when tracks from underlying events lie close to the electron track. Here a method to identify electrons is discussed. An electron is characterized by a narrow cluster in ECAL and a track pointing towards it. Different variables could be used to separate electrons from fakes.

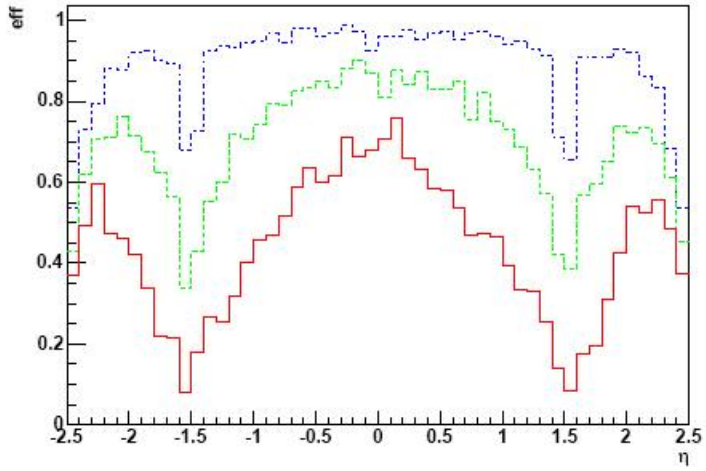

Figure 4.6: *`GSFElectron` reconstruction efficiency as a function of the track pseudorapidity for fixed p_T = 5 (red, solid line), 10 (green, dashed line), 30 (blue, dash-dotted line) GeV/c electrons. The results are obtained using GSF tracks. The E_T threshold for the superclusters has been lowered to 1 GeV.*

The different variables which could be used to discriminate between real and fake electrons mostly rely on the differences between the shower shapes in the calorimeter and on the combination of tracker and ECAL informations. The following variables were adopted in `CMSSW`[81]:

- **H/E**, the ratio between the energy which is deposited in the electromagnetic and the hadron calorimeter. Electrons tend to fully deposit their energy in the electromagnetic calorimeter, while hadrons leave a large fraction of their energy in the HCAL. The ratio is therefore close to zero for electrons and it is higher for hadrons.

- $\sigma_{\eta\eta}$ which exploits the fact that electromagnetic showers are usually narrower than hadronic ones, due to the limited range of nuclear interactions originating hadronic showers. It is defined as
$$\sigma^2_{\eta\eta} = \frac{\sum_i (\eta_i - \eta_{seed})^2 E_i}{\sum_i E_i} \tag{4.10}$$
with i running on the 5×5 crystals matrix around the seed, and it is a measure of the shower spread in η (low $\sigma_{\eta\eta}$ value ≡ low shower spread in η) not affected by bremsstrahlung (which spreads energy in ϕ).

- $\Delta\eta = |\eta_{track} - \eta_{SC}|$ and $\Delta\phi = |\phi_{track} - \phi_{SC}|$. Electron tracks directly point on the associated supercluster. These two variables compare the track direction with the supercluster position. In the case of $\Delta\phi$, the track is extrapolated from the vertex to the cluster position considering the magnetic field only. It is affected by bremsstrahlung emission and it is therefore energy dependent.

Off. e-ID sel.	LOOSE cuts		TIGHT cuts	
	barrel	endcap	barrel	endcap
H/E	0.115	0.083	0.015	0.075
$\sigma_{\eta\eta}$	0.0140	0.027	0.0092	0.028
$\lvert\Delta\eta\rvert_{\text{at vtx}}$	0.0090	0.01	0.0025	0.0065
$\lvert\Delta\phi\rvert_{\text{at vtx}}$ (rad)	0.09	0.04	0.025	0.020

Table 4.1: *Tabulated thresholds for the* `CMSSW` *default e-ID selections in CMS experiment[81].*

Studies on reconstruction and e-ID[81] shown that these selections variables are mostly uncorrelated for electrons.

Due to the dependence on the energy of some of the chosen variables, also e-ID depends on the energy scale of a particular signal. In the following, the default `CMSSW` electron reconstruction and e-ID will be applied for J/Ψ →e⁺e⁻ events and fully hadronic events (di-jet QCD events) to measure reco and e-ID efficiencies of e⁺e⁻ pairs for signal and background.

As signal sample is used the prompt J/Ψ in the High Radiation Scenario (where not-isolated electrons are expected) simulated in FullSim, and a sample of QCD events with $\hat{p}_T$ ≥30 GeV/c. This latter sample contains ∼1·10⁶ events and it was produced in FullSim by the CMS collaboration and available in the official DataBase Samples (DBS) as a reference sample for background studies in many physics study in CMS[86]. To obtain efficiencies it was made a comparison with Monte Carlo data. The considered `GsfElectron`'s are associated with the Monte Carlo truth following as matching criteria:

- `GsfElectron` has to include a Monte Carlo electron/positron from a J/Ψ within a $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.06$ cone.
- If more than one electron/positron in the event satisfy the previous criteria, the one with the lowest $\frac{p_T^{MC} - p_T^{GsfEl}}{p_T^{MC}}$ value is kept.

In Figs. 4.9, 4.8, 4.10, 4.11 are plotted the four e-ID selections distributions, obtained for an integrated luminosity of 1 pb^{-1}. Official e-ID foreseen two kind of selections with *loose* cuts or *tight* cuts according to how "hard" the selection/rejection have to be. The values of thresholds used are suggested in [81] and reported in Tab. 4.1.

The $\lvert\Delta\phi\rvert_{\text{at vtx}}$ and $\lvert\Delta\eta\rvert_{\text{at vtx}}$ selections are based on the GSF track ϕ and η estimated by the Gaussian Sum Filter algorithm. There are two ways to obtain those

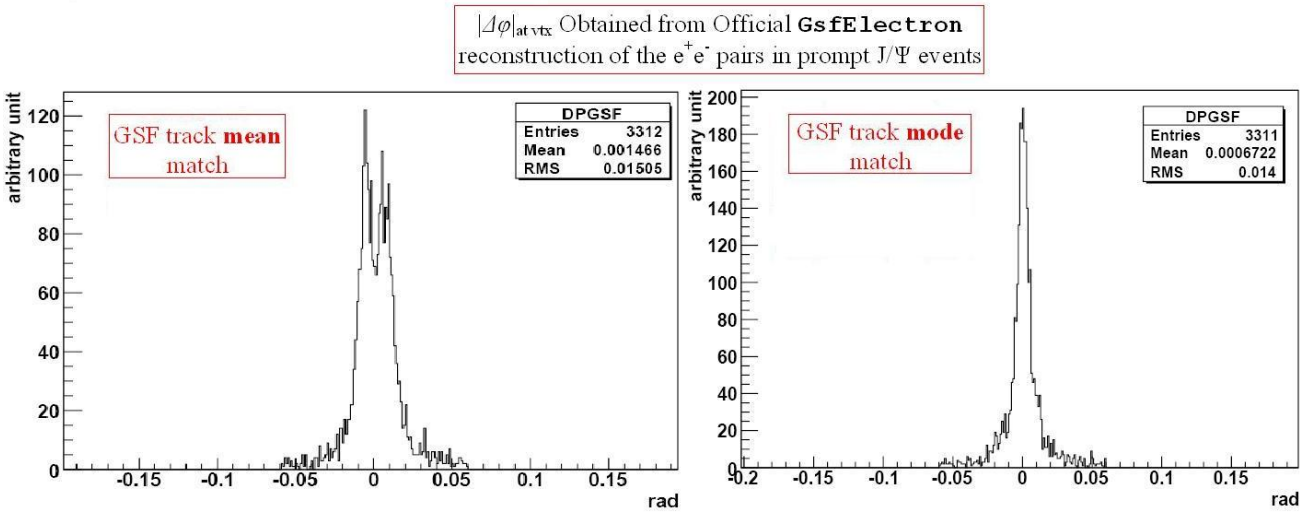

Figure 4.7: *Comparison between the usage of* ***mean*** *GSF track and* ***mode*** *GSF track in the* $|\Delta\phi|_{\text{at vtx}}$ *selection on* e^+e^- *pairs in prompt J/Ψ events.*

information: the first makes the weighted mean of track parameters of all the KF tracks used in the Gaussian Sum Filter (using as weights the ones in Eq. 4.9). The second consider only the KF track in the algorithm which has the higher extrapolated weight. In the first case we refer to **mean** GSF track and in the second to **mode** GSF track; in Fig. 4.7 there is the comparison of $|\Delta\phi|_{\text{at vtx}}$ distributions obtained from mean and mode algorithms. It is possible to see that the use of mode GSF tracks brings to more peaked distribution and so to more accurate estimation of the GSF tracks ϕ direction. Distributions shown in Figs. 4.10 and 4.11 are obtained with mode GSF tracks.

In the `CMSSW` default e-ID the H/E selection is the less powerful cut to optimize the signal-to-background ratio. On the other hand, the $\sigma_{\eta\eta}$ and the $|\Delta\eta|_{\text{at vtx}}$ selections seem to discriminate very well the signal from the background.

4.3 An Adapted e-ID and Reconstruction Algorithm for J/Ψ →e⁺e⁻ Events

In Chapter 3 it was shown that electrons in prompt J/Ψ's events have $p_T \sim \mathcal{O}(5$ GeV/c). From Fig. 4.6 we see that `GSFElectron` reconstruction efficiency drops for p_T lower than $p_T \simeq 5$ GeV/c. This is due to the fact that both reconstruction and e-ID are studied and optimized for $Z^0 \rightarrow e^+e^-$ events, where electrons have $p_T \sim \mathcal{O}(45$ GeV/c)

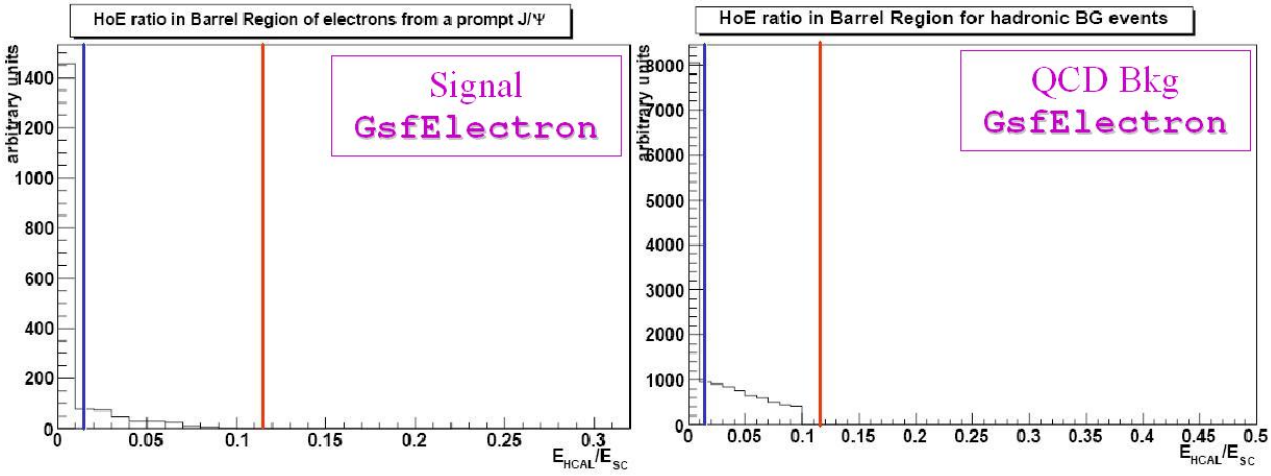

Figure 4.8: *Distribution of the ratio H/E for the e^+e^- pairs reconstructed with* `GsfElectron` *algorithm in prompt J/Ψ events. The blue lines are tight thresholds, the red ones are loose thresholds.*

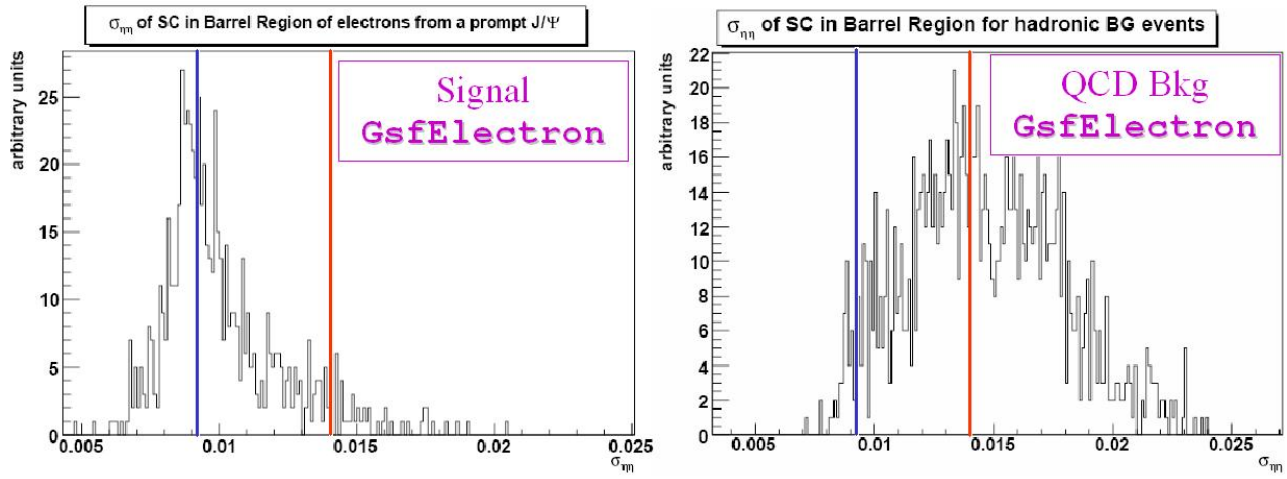

Figure 4.9: *Distribution of $\sigma_{\eta\eta}$ for the e^+e^- pairs reconstructed with* `GsfElectron` *algorithm in prompt J/Ψ events. The blue lines are tight thresholds, the red ones are loose thresholds.*

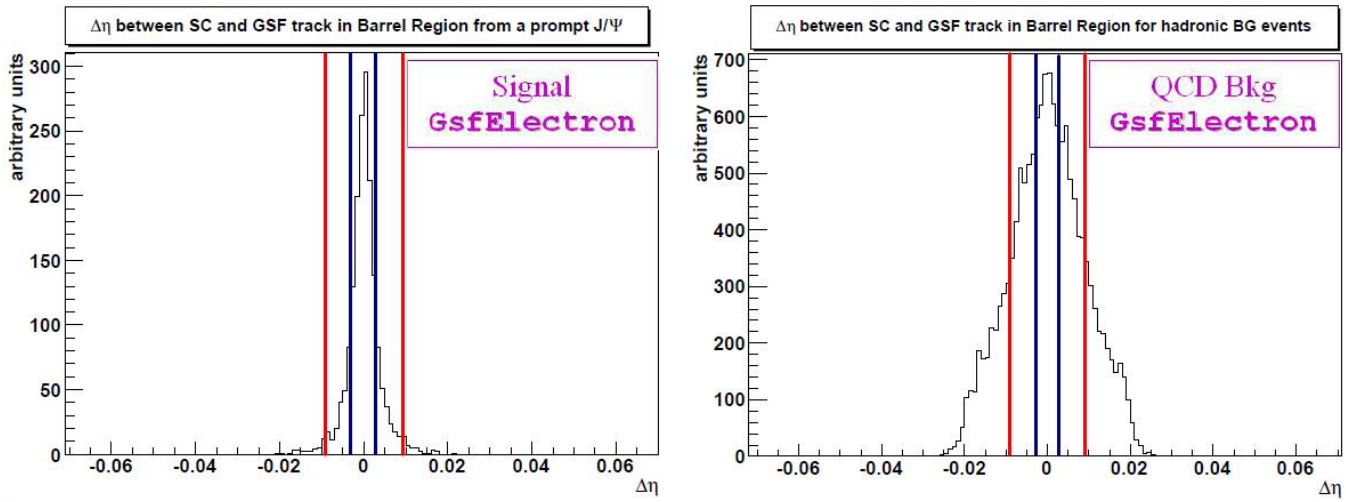

Figure 4.10: *Distribution of* $\Delta\eta_{\mathrm{at\ vtx}}$ *for the* e^+e^- *pairs reconstructed with* `GsfElectron` *algorithm in prompt J/Ψ events. Mode GSF tracks are used. The blue lines are tight thresholds, the red ones are loose thresholds.*

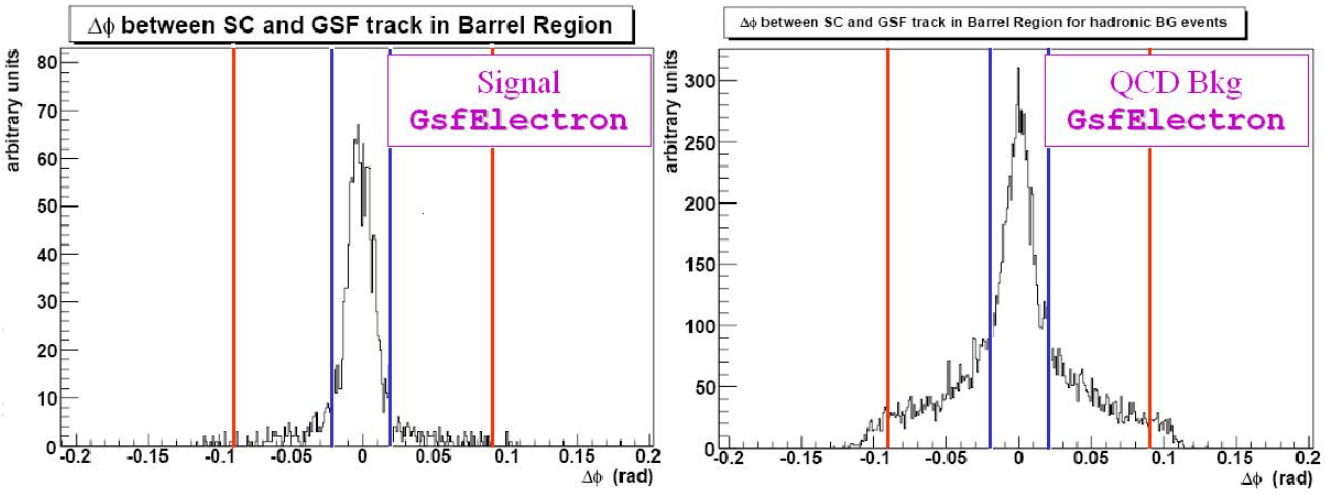

Figure 4.11: *Distribution of* $\Delta\phi_{\mathrm{at\ vtx}}$ *for the* e^+e^- *pairs reconstructed with* `GsfElectron` *algorithm in prompt J/Ψ events. Mode GSF tracks are used. The blue lines are tight thresholds, the red ones are loose thresholds.*

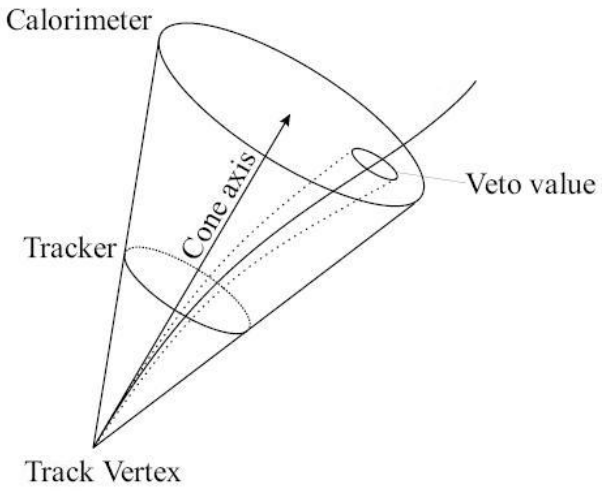

Figure 4.12: *Definition of Isolation Algorithm for a track. A "veto" cone is build closely around the track to exclude it from the algorithm computation.*

and are *isolated*. Experimentally, an electron is defined *isolated* if the sum of p_T's tracks included within a certain cone build around the electron's track (not including the electron p_T, see Fig. 4.12) is below a set threshold. To reconstruct J/Ψ →e⁺e⁻ events, specially prompt J/Ψ's, we have to consider that

- e⁺e⁻ pairs have $p_T \sim \mathcal{O}(5$ GeV/c).
- prompt J/Ψ's could be produced inside a big hadronic jet activity. So electrons could be surrounded by hadronic charged tracks resulting not-isolated.

Many analyses based on the `GSFElectron` reconstruction requires also some isolation criteria. These cannot be applied to the J/Ψ →e⁺e⁻ reconstruction in order to not-affect the efficiency and to avoid adopting selection criteria potentially highly model dependent.

Taking into account these requirements, both electrons reconstruction and e-ID are modified as follows:

- At least 2 ECAL SCs are required in the event, which match specific e-ID criteria.
- J/Ψ invariant mass is computed using **only** ECAL information: the *energy* of SC (E_{SC}) and its position (ϕ_{SC}, η_{SC}).
- The SCs positions are referred respect to the nominal CMS frame $(0, 0, 0)$, so η_{SC} has to be corrected for the spread in z of p-p collisions and ϕ_{SC} have to be corrected for the bending in the 3.8 T solenoidal magnetic field.

The choice of basing the analysis totally on ECAL measurements is due to the fact that J/Ψ analysis are refereed to the first 1-10 pb^{-1}, where a good comprension of the detector

is needed. In this way, the measurement is based only on one sub-detector (`GSFElectron` combine the measurements in the ECAL with the ones of the Tracker) and some study on the detector understanding can be done, such as crystals calibration. The J/Ψ →e⁺e⁻ channel can be useful as "monitoring" channel for the calibration, i.e. looking how the resolution of the invariant mass peak[3] changes as calibration conditions change. Some studies are reported in [82][83].

The computation of the invariant mass starts from the reconstructed four-momenta of the e⁺e⁻ pairs. Neglecting the electron mass (∼511 keV), four-momenta are univocally obtained by E_{SC}, η_{SC}, ϕ_{SC}. Being the position of ECAL SCs referred to the CMS frame, η_{SC} and ϕ_{SC} need to be corrected for the primary vertex displacement (i.e. the p-p collision point) and the bending in the magnetic field respectively.

The correction of η_{SC} is done knowing the coordinates of the primary vertex, which generally does not coincide with $(0, 0, 0)$. In particular, the spread in z can reach a width σ_z of 15 cm (the spread in x and y - 20 μm - can be considered negligible). The primary vertex is not a priori known, but is reconstructed. The primary vertex reconstruction is made using a Kalman Filter procedure. All tracks reconstructed in the event are sorted in increasing order of the distance of the closest approach of the track[4] and Kalman Filter is started with tracks with lower values of the distance. Iteratively, tracks are associated until they give a satisfactory χ^2 value, else tracks are rejected and used for secondary vertex reconstruction (i.e. vertex due to displaced decays, such as decay from B mesons). The algorithm for the vertex reconstruction is implemented in the `CMSSW` software, and further information about it can be find in [84]. Once primary vertex is reconstructed and z coordinate is known, the correction is made as shown in Fig. 4.13. It has to be pointed out that L is the distance between the z axis and the centroid of the shower in ECAL, which is slightly energy dependent and is longer than the bare distance between the z axis and the front face of ECAL crystals (∼1290 mm) since the electromagnetic shower requires few radiation length to develop inside the crystals. A check based on GEANT simulations was made, and the value of L was found to be 1.355 mm for J/Ψ →e⁺e⁻ electrons. With the correct value of L, the corrected value for SC pseudorapidity (ζ) is

$$\begin{aligned} \sinh(\zeta) &= \frac{L\sinh(\eta) - v_z}{L - \sigma} \quad \text{BARREL} \\ \sinh(\zeta) &= \frac{(D - v_z)\sinh(\eta)}{D - v_z - \sigma\sinh(\eta)} \quad \text{ENDCAP} \end{aligned} \tag{4.11}$$

[3] J/Ψ's natural width is ∼80 keV, so the reconstructed invariant mass peak is dominated in practice only by detector resolution.

[4] Technically, with "distance of the closest approach" of the track we mean the minimum distance between $(0, 0, 0)$ and the closest point of the extrapolated track. Further information will be given in Chapter 6.

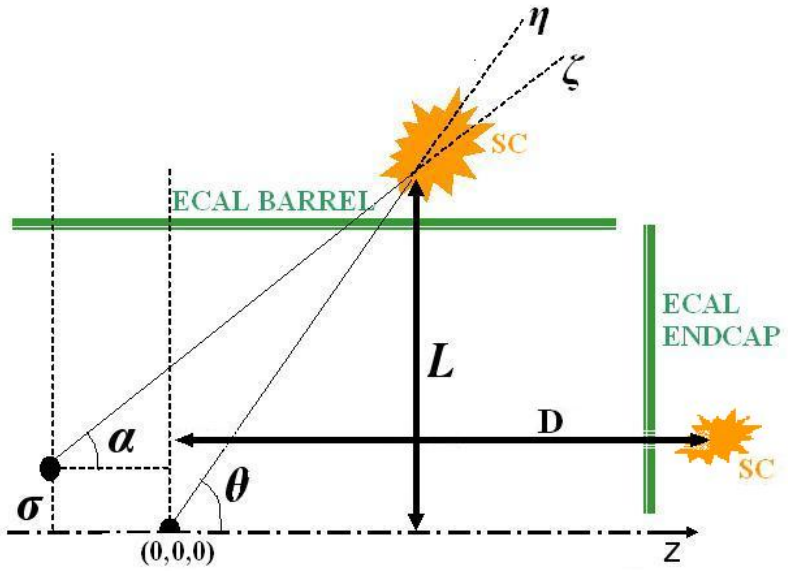

Figure 4.13: *Scheme in the yz plane of ECAL for η_{SC} correction.*

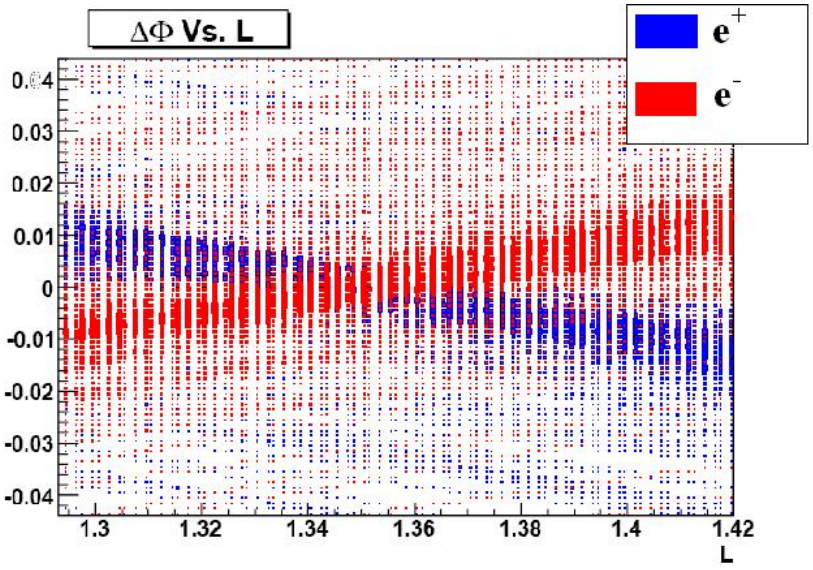

Figure 4.14: *Check on L value. The $\Delta\Phi$ distribution present a double peak for electrons and positrons which became a single peak in 0 when a correct value of L is used.*

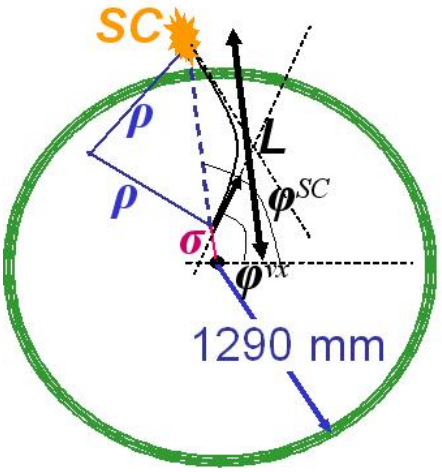

Figure 4.15: *Scheme in the xy plane of ECAL for ϕ_{SC} correction.*

where v_z is the z coordinate of the reconstructed vertex, and $\sigma = \sqrt{\sigma_x^2 + \sigma_y^2}$. For the D distance the same issues of L are valid, and the correct value has found in a similar way.

In Fig. 4.15 there is shown a scheme for the ϕ_{SC} correction. In principle, correction should take into account the magnetic field map and the presence of material budget (tracker) along the trajectory. Here, the corrected ϕ_{vx} value is obtained approximated by a simple helix backward propagation to the primary vertex. In this way, following the scheme in Fig. 4.15 and remembering that the track bending radius is given by $\rho[m] = 0.3B[T]p_T[GeV]$, the correction is given by

$$\phi_{vx} = \phi_{SC} + q \cdot \arcsin\Big(\frac{0.3B(L-\sigma)}{2p_T}\Big) \tag{4.12}$$

where q is the charge (± 1), B is 3.8 T, and p_T is the transverse momentum measured from the ECAL (the approximation $p_T \simeq E_T$ is valid for electrons). The value of L is estimated plotting the value of $\Delta\Phi$ difference between ϕ_{vx} and ϕ_{MC}, the electron ϕ obtained from MC truth, as a function of various values of L, differentiating the electrons from the positrons. A double peak is present in the distribution, one for positrons and one for electrons. For the right L value, the difference $\Delta\Phi$ present a single peak in 0 both for electrons and positrons. From Fig. 4.14 a value of ∼1360 mm is obtained.

After SCs are found, electron reconstruction is completed applying some e-ID criteria favourable for low-p_T and not-isolated electrons pairs. The e-ID is subdivided into two parts: first, *calorimetric selections* are applied, then a match with charged tracks is required to increase the rejection rate of hadronic fakes. So the e-ID is organized with the following selections applied in cascade:

- SC E_T cut.
- Cluster Shape of the SC (9/25).
- The "adapted" ratio H/E.
- The angular matching (in η and ϕ) with tracks.
- The particle bending compatibility between information obtained from tracker and information obtained from ECAL.

The E_T cut is based on the electron transverse energy measured from the SC. It is obtained as $E_T = E_{SC}/\cosh(\eta_{SC}^{corr})$, where η_{SC}^{corr} is the SC pseudorapidity corrected for the spread in z. This cut (**set to 5 GeV**), it is used for three reasons: 1) establish a p_T acceptance criteria since an implicit p_T cut is present due to trigger thresholds (see Chapter 5), 2) reduce the combinatorial background from low-p_T hadrons produced by QCD and MB events, and 3) Cutting on electrons energy, only boosted J/Ψ's (∼10 GeV) will be selected. The Lorentz Boost forward the electrons directions, and so a little separation angle between the e^+e^- is expected (this will be deeper treated in Chapter 6). Calling δ this angle, it can be obtained from the invariant mass (in Appendix A there is a brief explanation about invariant mass calculation) of pairs remembering that for relativistic pairs

$$M^2 = 2E_1E_2(1-\cos(\delta)) \quad \Longrightarrow \quad \cos(\delta) = 1 - \frac{M^2}{2E_1E_2} \tag{4.13}$$

So a cut on E_T should give a (loose) cut on δ; requiring a cut ≥5 GeV one obtain the constraint $\cos(\delta)\geq 0.8$, which rejects events where combinatorial brings to well-separated fakes ($\cos(\delta)<<1$).

The second selection relies on the exploiting of the cluster shape. Officially is used $\sigma^2_{\eta\eta}$, which has the advantage that it is not affected by bremsstrahlung effects which spread the SC along ϕ direction. Anyway, the cluster shape can be quantified also with the so called *9/25*; since the shower from a true electron is almost completely contained in a 3×3 array of crystals[85], the ratio 9/25 between the energy of the 3×3 and the 5×5 crystals arrays around the SC *seed* has a more pronounced peak around one in the electron case. Assuming that low-p_T electrons are not too much affected by bremsstrahlung, this variable was chosen because it is more "sensitive" to the fact that electromagnetic showers are narrower than hadronic ones: these last ones are generated via strong and nuclear interactions with short impact parameter, and the corresponding scattering brings to a shower development more transversally extended.

As other calorimetric selection, the ratio H/E has been exploited. Respect to the

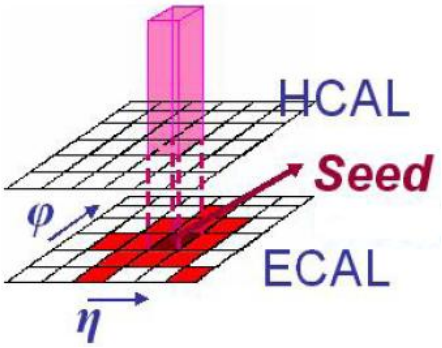

Figure 4.16: *Scheme of the implementation of the "adapted" H/E algorithm.*

CMSSW default definition, it was a bit modified. e^+e^- pairs could be not-isolated and so an "hard" selection based on H/E could reject many signal events. In the H/E ratio are used the energy reconstructed in ECAL and the energy reconstructed in HCAL in the region immediately above all the ECAL SC. If electron is not-isolated, than an high ratio should be expected. To reduce this effect, we have considered not all the HCAL region above the SC, but only scintillators immediately above the *seed* (i.e. the one-two most energetic crystals) of the SC, as it is shown schematically in Fig. 4.16.

After calorimetric selections, a tracker match of the SC is required. First it is required angular matching between the SC position (i.e. η and ϕ given by the centroid) and charged tracks. The match is made with variables reconstructed at vertex[5], because tracks are better reconstructed in the firsts tracker layers where the material budget is lower and so bremsstrahlung effects (and other) does not make worse the track parameters measurements. SC parameters matched with tracker informations are η_{SC}^{corr} and ϕ_{SC}^{corr} corrected as explained before. Selection is then based on $\Delta\eta = \eta_{track} - \eta_{SC}^{corr}$ and $\Delta\phi = \phi_{track} - \phi_{SC}^{corr}$.

The last selection for the e-ID is based on the compatibility between the information obtained from the ECAL and the information obtained from the tracker. In principle, an electron/positron bends in the magnetic field and leaves all its energy in ECAL crystals. So the p_T value obtained from the track and the E_T measured in ECAL (neglecting the electron mass) have to be the same. If an hadron does not leave all its energy in the ECAL, it should be $E_T \leq p_T$. Anyway, this is an ideal condition: bremsstrahlung could bring to inaccurate measurements of both p_T from tracker and E_T from the SC. As default, in the CMSSW code is implemented the compatibility between the information from ECAL and from the Tracker is made using the ratio E/p, where E is the energy of the SC and p is the momentum obtained from the matched track. This ratio is used to

[5]It is not properly the primary vertex, but the point of the closest approach to $(0, 0, 0)$.

e-ID sel.	barrel	endcap
E_T (GeV)	5	5
9/25	0.85	0.65
H/E	0.2	0.2
$\lvert\eta_{SC}^{corr} - \eta_{trk}\rvert$	KF: 0.009 GSF: 0.009	KF: 0.025 GSF: 0.025
$\lvert\phi_{SC}^{corr} - \eta_{trk}\rvert$ (rad)	KF: 0.02 GSF: 0.02	KF: 0.05 GSF: 0.05
$\frac{1}{E_T} - \frac{1}{p_T}$ (GeV^{-1})	KF: [-0.04,0.02] GSF: [-0.02,0.02]	KF: [-0.1,0.04] GSF: [-0.04,0.04]

Table 4.2: *Tabulated thresholds for the "adapted" e-ID selections.*

quantify the fraction of energy lost in the SC reconstruction due to bremsstrahlung. For our purposes, it was chosen to compare the bending radius ρ in the xy plane obtained from the two subdetectors. So, remembering that $p_T \sim \rho$, the compatibility between the two radius is sought requiring that the value of

$$\frac{1}{p_T} - \frac{1}{E_T} \tag{4.14}$$

is included within a confidence range (see Tab. 6.11 for the used thresholds). If more than one track which satisfy the angular matching condition is found, then the one with the highest bending compatibility is kept.

Both Kalman Filter tracks (KF) and Gaussian Sum Filter tracks (GSF) are used for the track-SC matching. The same selections are applied in both cases. It is important to stress that only calorimetric information are used for the physical analysis (Chapter 6), so no change in resolution is expected between the use of a KF track and a GSF track. The only expected changes can be in reconstruction efficiencies. Indeed, GSF track parameter allows a better resolution specially in the SC-track match, but the Kalman Filter algorithm results to be more effective than the Gaussian Sum Filter in track reconstruction for low-p_T particle; GSF is started when a pixel seed is found during the *Pixel Matching*, which can fail if p_T particle is low and escapes from the matching window.

The adapted reconstruction and e-ID algorithm is based starting from SCs in the ECAL and on calorimetric selections followed by tracker matching to improve the rejection of hadronic background. Both KF and GSF tracks are considered, and *mode* parameters are considered in the GSF case. Most of the considered selections are energy dependent, so they are applied in cascade starting from the calorimetric ones to the tracker matching criteria.

In Tab. 4.2 the values of thresholds used are reported. Assuming that at low E_T the error on the $\frac{1}{E_T} - \frac{1}{p_T}$ difference is dominated by ECAL resolution, a reference value for

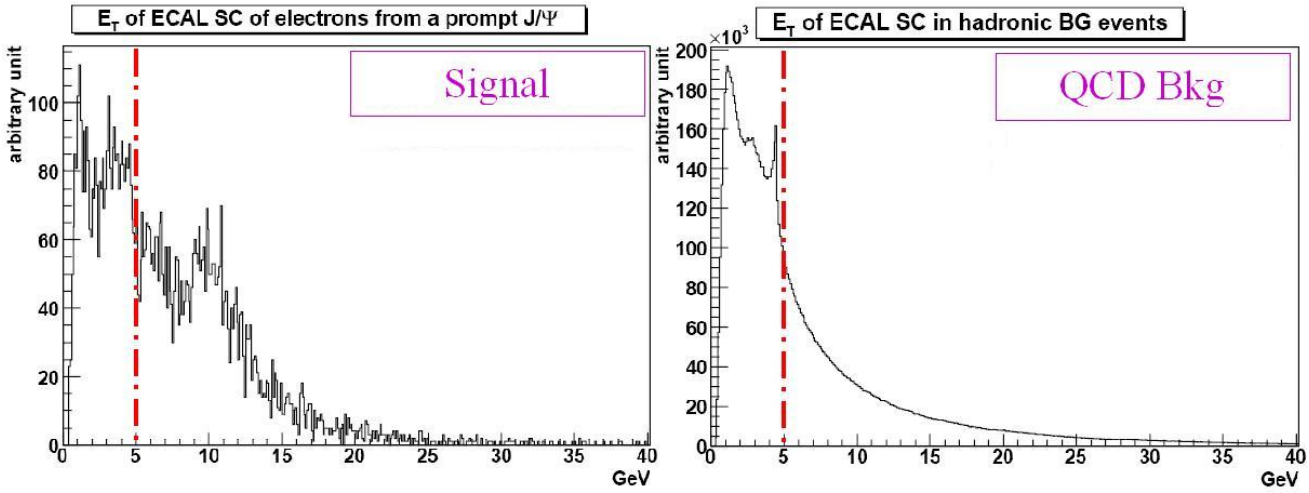

Figure 4.17: *Distribution of the E_T for the e^+e^- pairs reconstructed in ECAL with SCs in prompt J/Ψ events. The E_T is corrected for the spread in z of the Primary Vertex, i.e. $E_T = E_{SC}/\cosh(\eta_{SC}^{corr})$. The selection threshold is the red line.*

electrons with $E \sim \mathcal{O}(5$ GeV) can be estimated through the ECAL resolution formulæ (Eq. (2.7) in Chapter 2) to be:

$$\sigma(1/E) = \frac{\sigma(E)}{E^2} \simeq \frac{0.027\text{GeV}^{\frac{1}{2}}}{E^{\frac{3}{2}}} \oplus \frac{0.2\text{GeV}}{E^2} \oplus \frac{0.005}{E} \simeq 0.015 \qquad (4.15)$$

In Figs. 4.17, 4.18, 4.19, 4.20, 4.21 and 4.22 are plotted the e-ID selections distributions for the e^+e^- pairs in prompt J/Ψ events. The plots are refereed to an integrated luminosity of 1 pb^{-1}.

Distributions are obtained only for events which pass the same triggers at level 1 (L1, i.e. hardware based, as it will be better described in Chapter 5), in order to refer them "visible" events[6]. The level 1 trigger dedicated to electromagnetic particles ($e^{\pm}$, γ) selects at least two ECAL SCs with $E_T \geq 5$ GeV **or** at least only one SC with $E_T \geq 8$ GeV.

The E_T selection presents a double fall at ∼5 GeV and ∼8 GeV. This is due to the rejection made by the two level 1 triggers (exactly at 5 GeV and 8 GeV).

The 9/25 distribution is peaked in 1 for electromagnetic deposits. The presence of a pronounced peak at 1 also in QCD events probes that in hadronic events there is a pure electromagnetic part. More in-depth studies about this background composition will be made in Chapter 6.

After the track-matching, the distribution of $1/E_T - 1/p_T$ (bending compatibility between

[6] "Visible" term is used meaning events which pass the trigger selections foreseen in the CMS experiment (see Chapter 5).

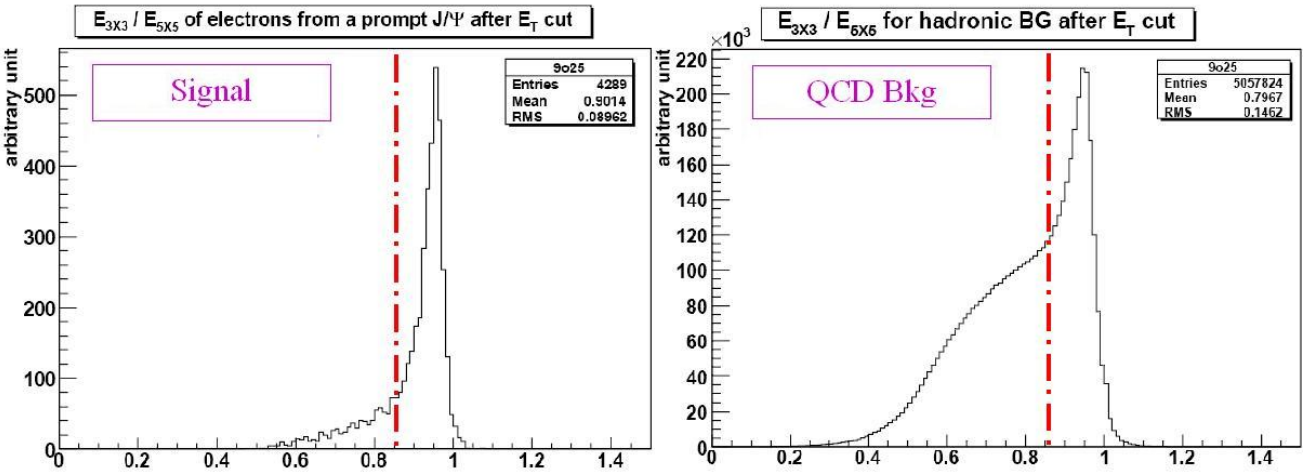

Figure 4.18: *Distribution of the ratio $E_{3\times3}/E_{5\times5}$ for the e^+e^- pairs reconstructed in ECAL with SCs in prompt J/Ψ events. The selection threshold is the red line.*

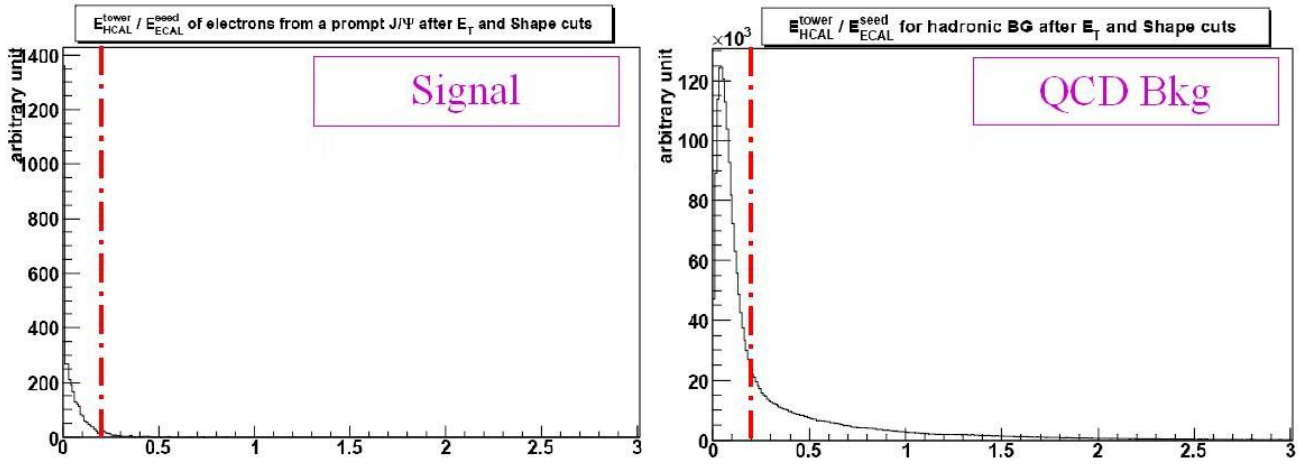

Figure 4.19: *Distribution of the "adapted" H/E ratio for the e^+e^- pairs reconstructed in ECAL with SCs in prompt J/Ψ events. The selection threshold is the red line.*

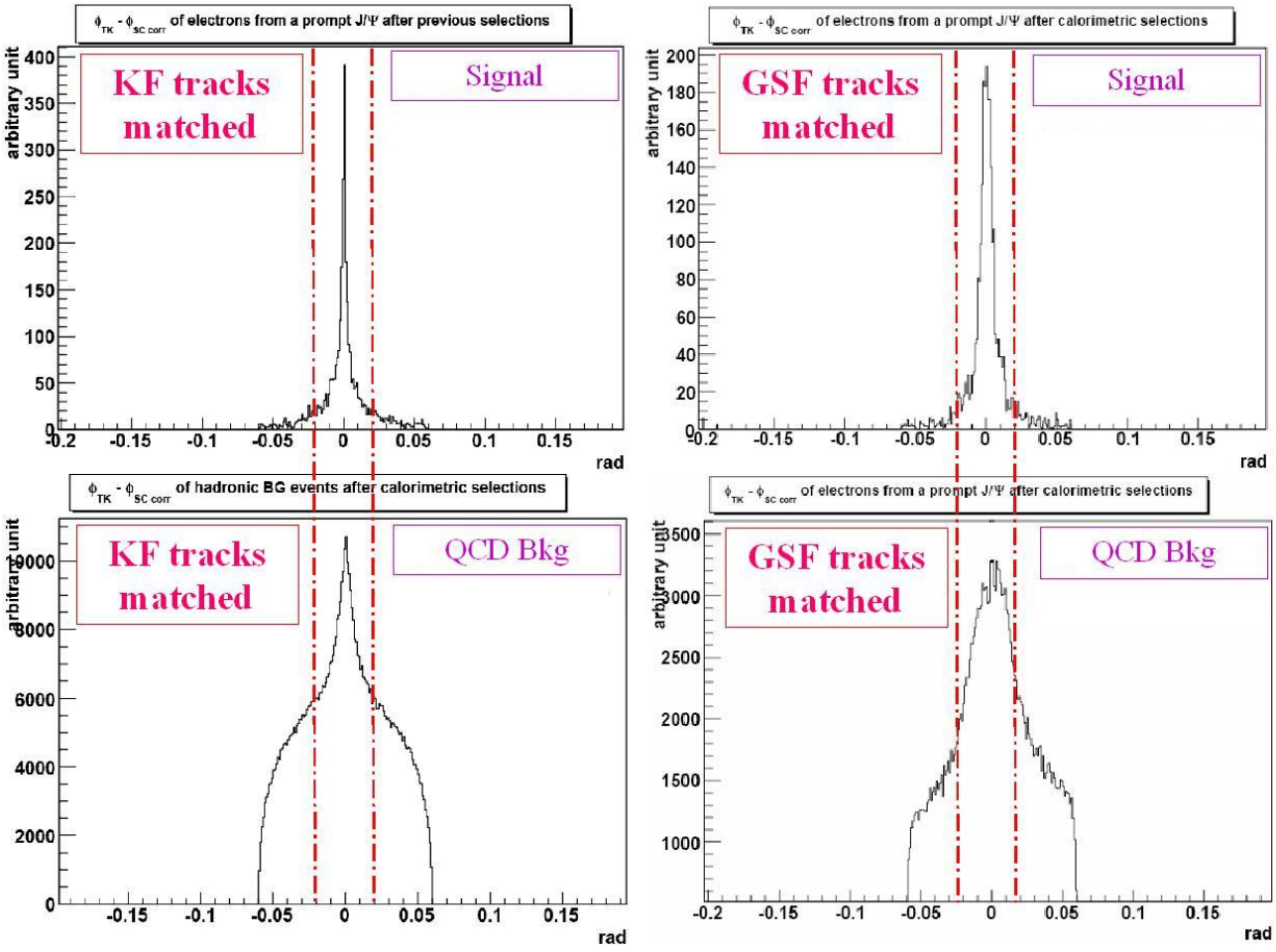

Figure 4.20: *Distribution of $\phi_{SC}^{corr} - \phi_{Trk}$ for the e^+e^- pairs reconstructed in ECAL with SCs in prompt J/Ψ events. Both KF and GSF tracks are considered. The selection threshold is the red line.*

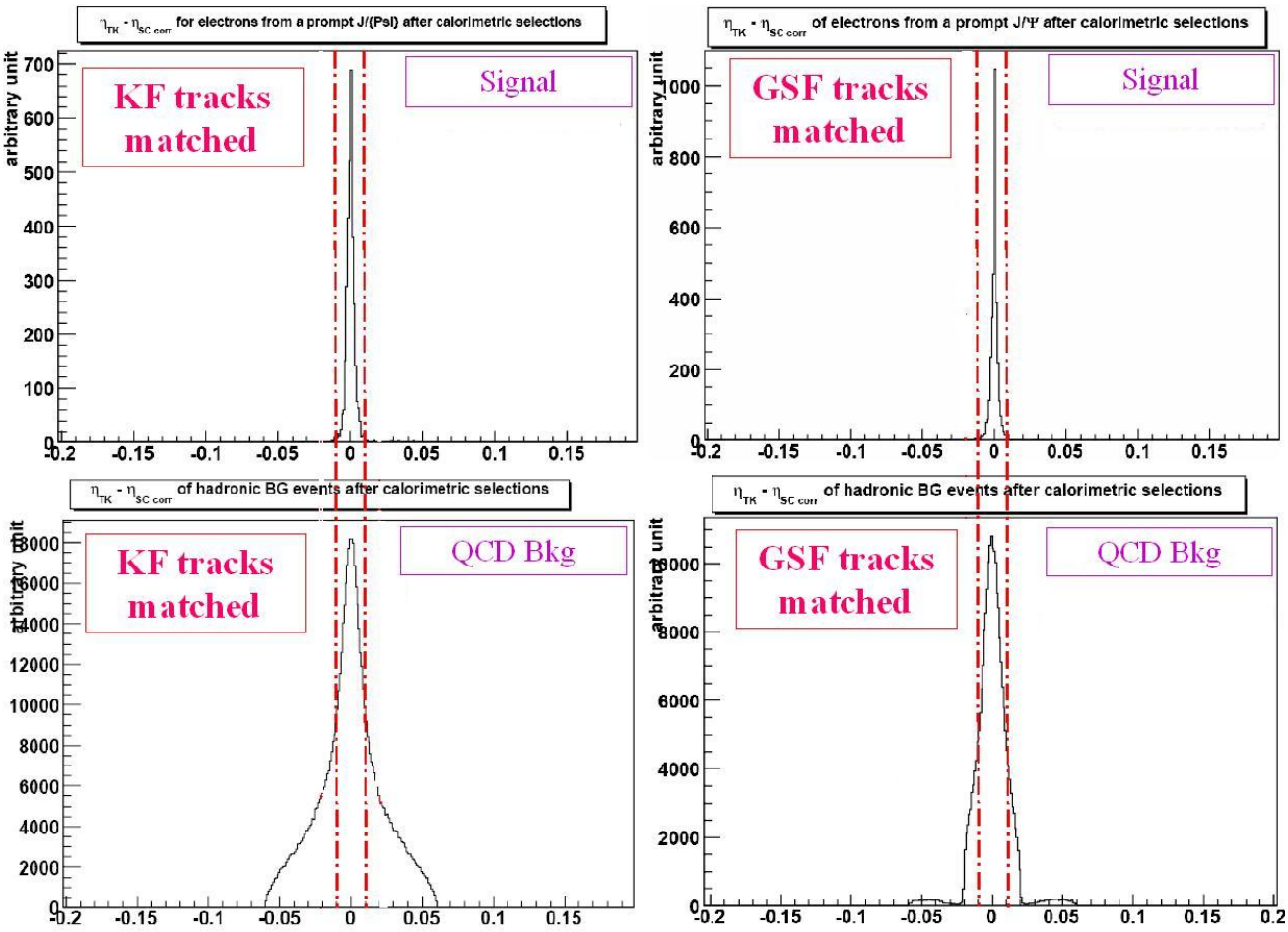

Figure 4.21: *Distribution of $\eta_{SC}^{corr} - \eta_{Trk}$ for the e^+e^- pairs reconstructed in ECAL with SCs in prompt J/Ψ events. Both KF and GSF tracks are considered. The selection threshold is the red line.*

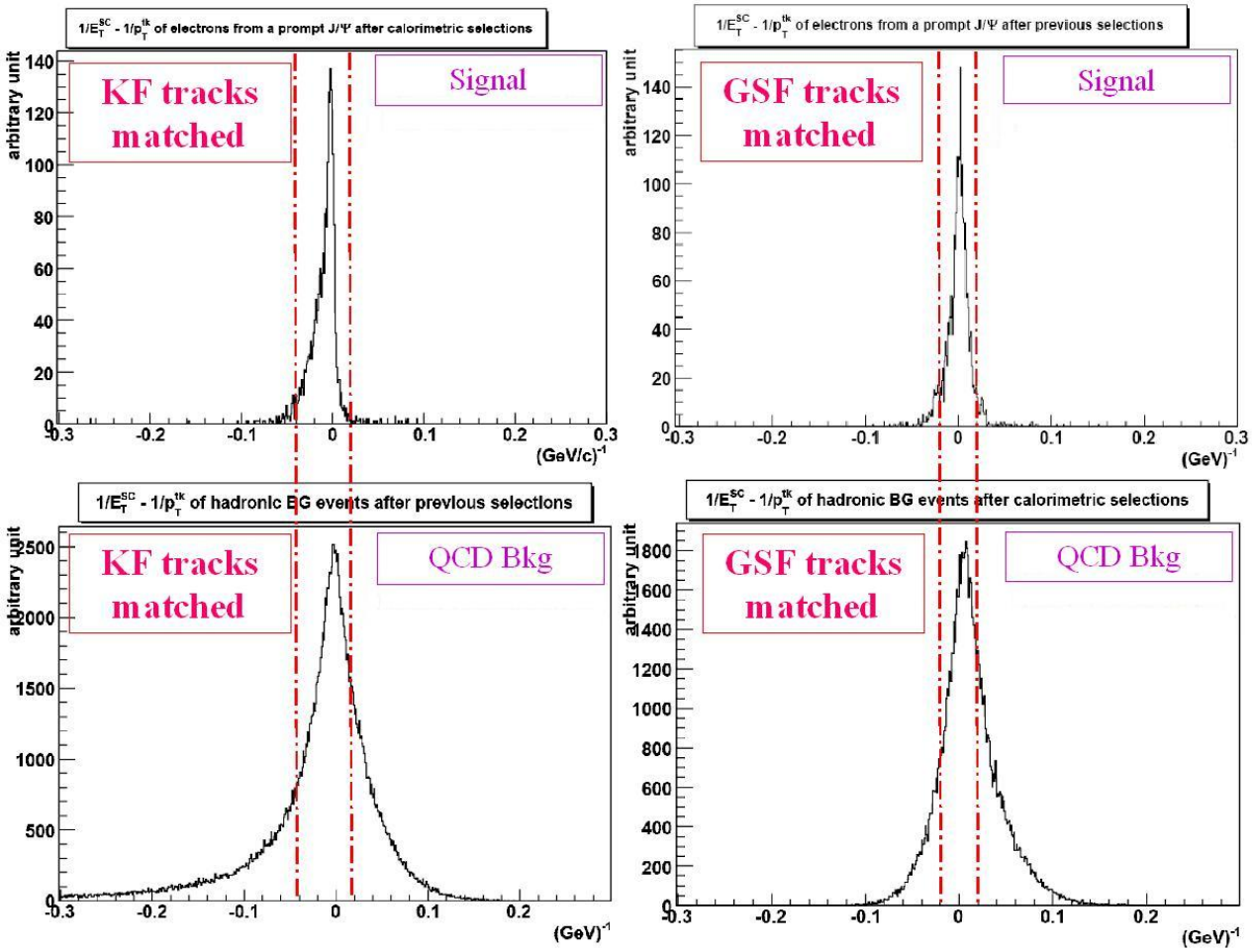

Figure 4.22: *Distribution of* $1/E_T - 1/p_T$ *for the* e^+e^- *pairs reconstructed in ECAL with SCs in prompt J/Ψ events.* E_T *is obtained from ECAL (*$E_T = E_{SC}/\cosh(\eta_{SC}^{corr})$*) and* p_T *from the matched tracks. Both KF and GSF tracks are considered. The selection threshold is the red line.*

informations obtained from ECAL and Tracker) presents asymmetric tails in the case of match with KF tracks. Asymmetries for this quantity can be due to two possible effects: energy radiated by bremsstrahlung along the track trajectory, and incompatibility between ECAL E_T and track p_T. The last effect is present specially in hadronic events, where more than one hadrons in a jet can leave energy deposit in the same SC, increasing the total energy reconstructed in the ECAL more than it can be expected from the single matched track. In Fig. 4.22 it is possible to see that if GSF tracks are used for the SC matching then more symmetric tails appear in the distributions. Gaussian Sum Filter algorithm takes into account bremsstrahlung effects better than the Kalman Filter one, correcting partially the asymmetry in the distributions.

4.4 Comparisons between the `GSFElectron`'s and the J/Ψ "adapted" algorithm

In this section will be shown a comparison of the selections used both in `CMSSW` default e-ID of `GSFElectron` and in the "adapted" e-ID. Efficiencies are then estimated by the use of Monte Carlo samples and simulation techniques implemented in `CMSSW`.

The comparison is made on events which pass the same triggers at level 1 explained in the previous section, in order to refer it on the same "visible" events giving more physical sense comparing efficiencies.

In the "adapted" reconstruction and e-ID the association with the Monte Carlo truth is made with the ECAL SCs, whit the corrected values for ϕ and η. The same criteria of the `GsfElectron`'s are used in the association.

In order to compare quantitatively the two reconstruction and e-ID algorithm, the number of events passing each selection are reported both for signal and background. For the signal, is reported the number of events where the whole e^+e^- pair is reconstructed and identified as electrons. This is possible controlling the Monte Carlo truth associated to the reconstructed electron, following the criteria exposed at the initial part of this section.

The e^+e^- pair candidates are considered as "reconstructed and identified" if there are reconstructed and identified at least two electrons with opposite charge (obtainable from tracks bending in the magnetic field) and with a computed invariant mass (correcting opportunely for the spread of the primary vertex and for the bending in the magnetic field in the case of the adapted reconstruction algorithm) within 3.1 GeV/c^2 ± 200 MeV/c^2.

In Tab. 4.3 there is the report of events (signal and background) passing official e-ID

Signal	LOOSE	%	TIGHT	%	**Background**	LOOSE	%	TIGHT	%
L1 Trig.	4398	100	4398	100	L1 Trig.	$2.94{\cdot}10^7$	100	$2.94{\cdot}10^7$	100
H/E	4398	100	2050	47	H/E	$2.94{\cdot}10^7$	100	$3.53{\cdot}10^5$	1.2
$\lvert\Delta\eta\rvert$	4303	98	4002	91	$\lvert\Delta\eta\rvert$	$4.41{\cdot}10^6$	15	4704	0.16
$\lvert\Delta\phi\rvert$	4134	94	3295	75	$\lvert\Delta\phi\rvert$	$2.50{\cdot}10^6$	8.5	$1.06{\cdot}10^6$	3.6
$\sigma_{\eta\eta}$	3997	91	1450	33	$\sigma_{\eta\eta}$	$7.94{\cdot}10^6$	27	64680	0.22
final `AND`	**3914**	**89**	**654**	**15**	final `AND`	$\mathbf{2.92{\cdot}10^6}$	**10**	**1232**	**0.042**

Table 4.3: *Number of events with reconstructed and identified e^+e^- pairs per pb^{-1} using official algorithms.*

Signal	# events	%	**Background**	# pairs	%
L1 Trig.	4398	100	L1 Trig.	$2.94{\cdot}10^7$	100
E_T	2404	55	E_T	$5.27{\cdot}10^5$	18
9/25	1540	35	9/25	$2.12{\cdot}10^4$	0.10
H/E	1491	34	H/E	$4.56{\cdot}10^3$	0.015

Tkr match	KF	%	GSF	%		KF	%	GSF	%
Ang. match	1020	23	888	19		819	0.0028	473	0.0016
Bend Comp.	**843**	**19**	**750**	**17**		**490**	**0.0016**	**310**	**0.0010**

Table 4.4: *Number of events with reconstructed and identified e^+e^- pairs per pb^{-1} using $J/\Psi \rightarrow e^+e^-$ algorithms.*

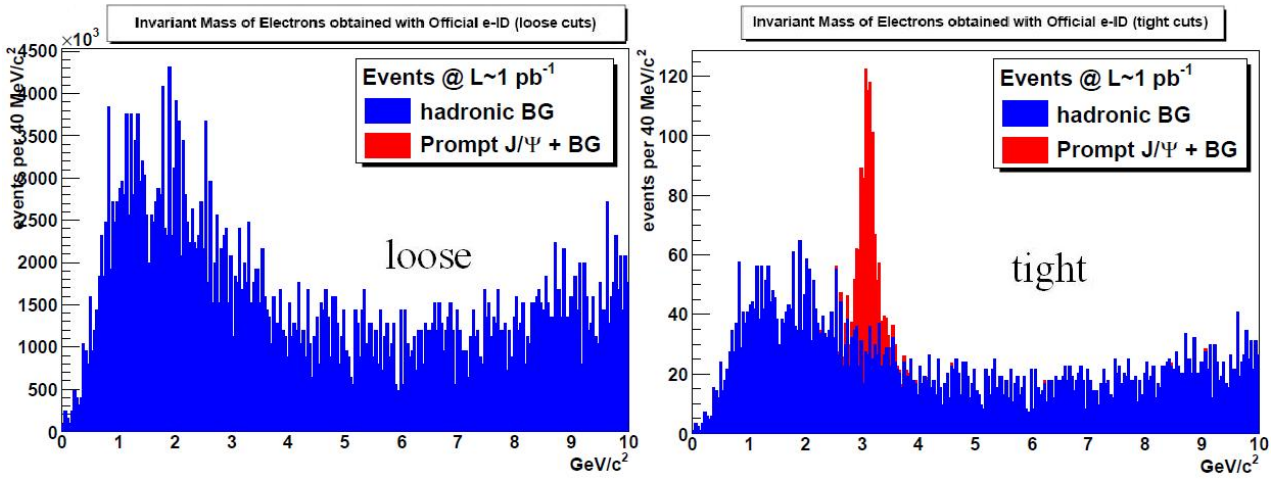

Figure 4.23: *Invariant mass distribution of* `GsfElectron` *pairs identified as* e^+e^- *pairs. Both loose and tight e-ID are shown. The chosen binning for the histogram is 40 MeV/*c^2 *per bin.*

selections on `GsfElectron`'s, and in Tab. 4.4 there is the same report for the adapted reconstruction. The numbers are refereed per pb^{-1}. In the case of official reconstruction every selection is applied "stand-alone", and only at the end of pairs identification it is required a logic `AND` of every selection[7]. In the adapted e-ID each selection is applied in cascade, and the angular matching with tracks is required with the logic `AND` of the $\phi_{SC}^{corr} - \phi_{Trk}$ and $\phi_{SC}^{corr} - \phi_{Trk}$ selections.

Finally, in Fig. 4.23 there are the invariant mass distributions of the `GsfElectron` pairs identified as e^+e^- pairs, showing the contribution of signal and background in the spectrum, both for loose and tight e-ID. In Fig. 4.24 there are the same distributions, using the SCs identified with the "J/Ψ →e^+e^- adapted" e-ID. The invariant mass is computed reconstructing the four-momentum of the reconstructed and identified objects with the respective E_T, η and ϕ (correcting angular kinematic variables in the case of SCs; the `GSFElectron`'s give directly the values at the track vertex) as described in Appendix A.

Conclusions

To quantify the how is good the reconstruction of the signal and the rejection of the background at the same time, we introduce now the *significance* (S), defined as (assuming

[7] `CMSSW` gives the possibility to put the response of each selection building a statistic *Likelyhood* function or using a *Neural Network*. Anyway, to make a comparison it is enough the requirement of a simple final `AND` (the so called *cut-based* e-ID)[87].

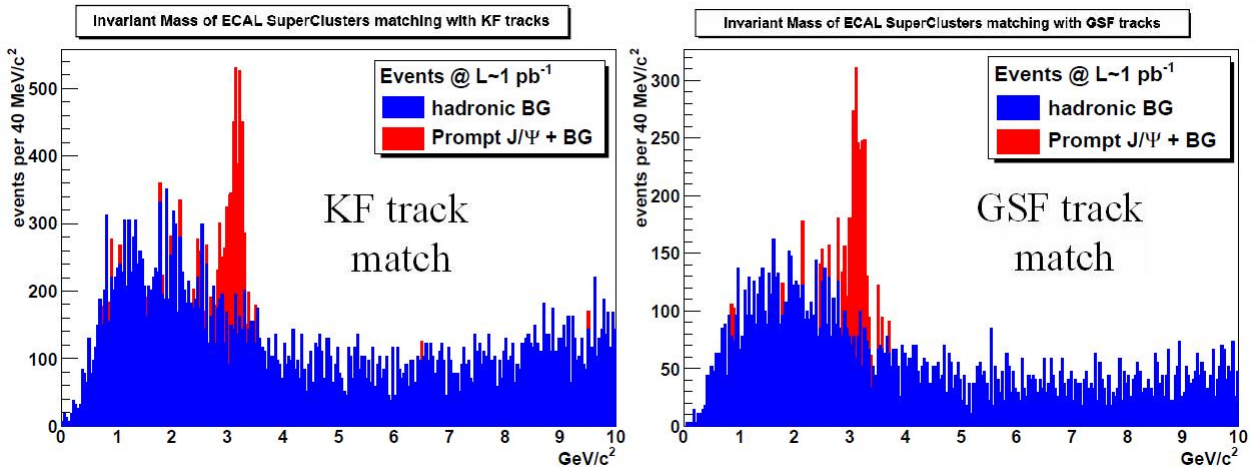

Figure 4.24: *Invariant mass distribution of SCs pairs identified as e^+e^- pairs after an adapted e-ID. Both the case of matching with KF tracks and GSF tracks are shown. The chosen binning for the histogram is 40 MeV/c^2 per bin.*

Poisson statistics in the histograms)

$$S = \frac{N_S}{\sqrt{N_B}} \tag{4.16}$$

where N_S is the number of reconstructed signal events and N_B the number of reconstructed background events (here we consider events within 3.1 GeV/c^2 ± 200 MeV/c^2). This parameter assesses the performance of the various reconstruction and identification algorithms. Looking Figs. 4.23, 4.24 and Tabs. 4.3, 4.4 the following significances after a referential integrated luminosity of 1 pb^{-1} can be obtained:

- $S_{OffLOOSE} = \frac{3914}{\sqrt{2.92 \cdot 10^6}} \simeq 2$
- $S_{OffTIGHT} = \frac{654}{\sqrt{1232}} \simeq 19$
- $S_{SC-KF} = \frac{843}{\sqrt{490}} \simeq 38$
- $S_{SC-GSF} = \frac{750}{\sqrt{310}} \simeq 42$

The first consideration is that loose selection in official e-ID is not good, because the J/Ψ peak is completely invisible on top of the huge background. On the other hand, the tight selection is the one with the higher significance. However, also the adapted e-ID algorithm presents high significances both in the match of SCs with KF tracks and GSF tracks. The tight e-ID selection seems to have a quite worse signal-to-background ratio than the e-ID based on the match SC-KF, and this last one is more effective, as it can be seen from the final numbers of reconstructed and identified signal events (843 against

654). As it will be shown in the following (Chapter 6) the big part of the background is composed by real e^+e^- pairs and can be further reduced with opportune selections. Since in the firsts pb^{-1}'s of collected data the measurement will be limited by data statistic, the adapted algorithms for the reconstruction and identification of J/Ψ →e^+e^- events based on match with KF tracks seems to be the better one. Then, a J/Ψ invariant mass peak obtained from the only ECAL measuraments is useful at the first runs of LHC, where detectors has still to be known.

The last observation is on the resolution of the reconstructed invariant mass peak. Being the natural width of J/Ψ ∼80 keV, the width of the peak in the histograms is affected only by experimental resolution. It is interesting in the case of reconstructed peak using the kinematic informations from ECAL SCs to know what is the variable which affects more than the others the resolution. The peak can be fitted with a gaussian to estimate the width. Anyway, the best method to fit the peak will be presented in Chapter 6 where analysis of data will be presented. For the moment, the resolution can be estimated simply by the FWHM[8] of the invariant mass peak. From Fig. 4.24 (after background subtraction) a FWHM∼140 MeV/c^2 is measured. The invariant mass M can be computed as

$$M = \sqrt{2E_1E_2(1-\cos\delta)} \tag{4.17}$$

where E_1, E_2 are the energy of the two SCs and δ is the angle between them, which depends from the angular variables (η and ϕ) of the two SCs. Neglecting the resolution due to δ measurement, and assuming that $E_1 \simeq E_2 \sim E$ =5 GeV, the total resolution on M can be obtained as

$$\sigma(M) = \frac{2E(1-\cos\delta)}{M}\sigma(E) \equiv \frac{2E^2(1-\cos\delta)}{M}\frac{\sigma(E)}{E} \tag{4.18}$$

Remembering the Eq. 2.7 in Chapter 2 valid for the CMS ECAL

$$\frac{\sigma(E)}{E} \simeq \frac{0.027\text{GeV}^{\frac{1}{2}}}{\sqrt{E}} \oplus \frac{0.2\text{GeV}}{E} \oplus 0.005 \tag{4.19}$$

and a value of δ (taking into account that boosted electrons are nearby close) that cosδ ∼0.8, from Eq. 4.18 can be obtained

$$\sigma(M) \sim 130\text{GeV/c}^2 \tag{4.20}$$

So it means that the higher effect on invariant mass resolution is given by the energy resolution achievable in the ECAL. This shows how the usage of J/Ψ →e^+e^- can be important for calibration purposes (even if the statistics on total reconstructable events

[8]≡ Full Width Half Maximum.

is not satisfying). On the other hand, this means that $J/\Psi \rightarrow e^+e^-$ depends strongly from the calibration conditions of the detector. For the study presented in this chapter the *START-UP* conditions were used (i.e. conditions at the really firsts runs of LHC, $\sim$1-2 pb^{-1}). Anyway, as it will be better described in chapter 6, these conditions are too pessimistic for an analysis refereed at 10 pb^{-1} for example. So higher reconstruction efficiencies can be obtained in the real start-up of LHC.

Chapter 5

The Trigger System at CMS: $J/\Psi \rightarrow e^+e^-$ Events On-line Selection

As discussed in Chapter 2, the LHC is designed to operate at a very high event rate and collision energy (40 MHz at high luminosities), and there is neither the space nor the acquisition bandwidth available to store all of the data produced.

The readout rate is particularly limited by data storage space. Each event in CMS produces approximately 1 MB of RAW data and the total data volume produced by the CMS detector is $\sim \mathcal{O}(100$ TB/s). This huge amount is too large to be handled by current storage and network capabilities and therefore has to be trimmed to more reasonable values through a sophisticated trigger cascade system. The main challenge is therefore the selection of interesting events among the huge amount of background (first of all Minimum Bias events) which has to be reduced to a manageable amount. The event rate which can be recorded for offline processing and analysis with the odiern technologies is in the order of **150 Hz**.

In this chapter, are described the techniques adopted in CMS to reduce the events rate to be stored, i.e. the trigger system, talking more in detail of electromagnetic trigger. Then, the trigger developed for $J/\Psi \rightarrow e^+e^-$ events is presented and validated, to obtain the rate of events filtered by triggers and which is the occupancy of the total band width.

5.1 The Data Acquisition System (DAQ)

In order to reach a band width of data in the order of 150 Hz, the CMS trigger and data acquisition system (TriDAQ)[88] comprises two stages of event selection. In Fig. 5.1 the architecture of the CMS DAQ and trigger system is shown.

The **Level-1** (L1) trigger[88] is primarily hardware-based processing system in order

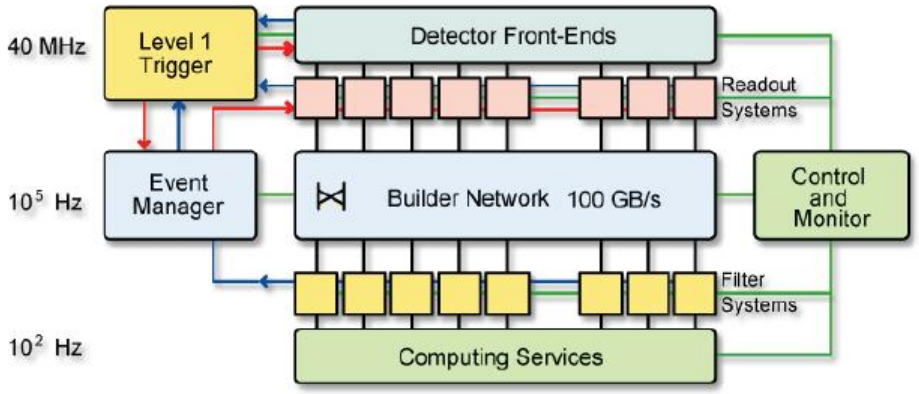

Figure 5.1: *A schematic overview of the TriDAQ.*

to handle the enormous data volume from the detector and provide a trigger decision in a very short time period of ∼3.2 μs. The goal of the L1 system is to reduce the data volume by an average factor of several hundreds (i.e. to 100 kHz for high luminosity, ∼30 kHz for low luminosities at LHC start up).

The **High Level** Trigger (HLT)[89] reduces the rate by a further factor of $\sim 10^3$ and is dominated by the use of computer farm using a multi-stage iterative approach to reconstruction with a limiting cut-off in allowed processing time. At this level, basic tracker information is used, and initial selection based on possible underlying physics events is made. The data produced at this level of filtering is then recorded to disk at approximately 100 Hz, satisfying the initial requirement of 150 Hz.

The purpose of the trigger is to discard the vast majority of events produced which is not interesting, i.e. low transferred momentum interactions (Minimum Bias events, so called because they are the nearly total fraction of filtered events if triggers thresholds were ideally lowered to 0) and retain data relating to events that cannot be immediately identified. Of course at the same time one must ensure that the trigger has the capability to distinguish signatures of possible new physics and store them for further study, such as Higgs or SuperSymmetry. Essentially, in CMS this reduces to a selection on the transverse energy, missing transverse energy and types of particle detected. In doing this one is placing as fewer constraints as possible on the physics available to the end-user (ignoring low energy physics for which the LHC was clearly not designed). This is often described as inclusive triggering and is particularly important for the first level trigger stage. Differently from other experiments, the trigger system in CMS doesn't synchronize the detectors to acquire data when a trigger signal occurs: in CMS (such as in other LHC experiments), detectors are synchronized to the clock signal given by LHC, i.e. detectors

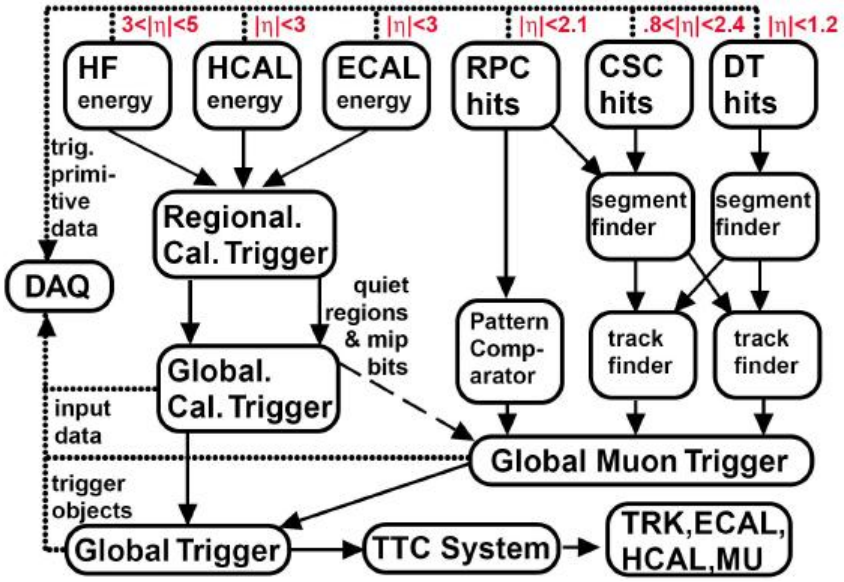

Figure 5.2: *A schematic overview of the L1 system.*

are ready to acquire every protons bunch crossing (every ∼25 ns at higher luminosity).

In addition to the requirement of efficient event selection, the trigger must also operate with minimal dead time[1], allowing one to maintain a high efficiency for the recording of useful events. This is achieved using buffers in the detector readout combined with a fast, efficient trigger processor.

5.2 The Trigger System

5.2.1 The Level-1 Trigger (L1)

The Level-1 trigger is responsible for the selection of events containing electrons, muons, photons, jets and missing transverse energy. Its output rate and speed are limited by the readout electronics and by the performances of the 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

[1] A period after a trigger during which data cannot be taken.

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. 5.2.

L1 Calorimeter Trigger

The input for L1 Calorimeter Trigger are the **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** in the ECAL considers domains of 5×5 crystals and sums the pulse shape data from a raw of 5 crystals to produce a strip energy; in the ECAL barrel, all the strip energies are also summed to compute the energy which is deposited in the trigger tower. Trigger data are then transmitted to the off-detector electronics through a serial link which operates at 800 Mb/s, which allows the data transmission every 25 ns. In case of a L1 accept, the data from the triggered events are also transmitted to the off-line electronics through a separate serial data link also working at 800 Mb/s. This allows reducing both the costs of the optical link system and the quantity of the off-detectors electronics by about a factor 8. Also the Regional Calorimeter Trigger works similarly. In this way, the Regional Calorimeter Trigger finds out electron, photon, tau 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 reconstructing roughly the bending radius in the magnetic field lines return in the yokes 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 (obtained as in RPC) 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 flags, 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.

The Trigger Table is a list of available L1's implemented in the Level-1 System electronics. For the various luminosity scenarios different rates of collisions and interactions are expected, so there are more Trigger Tables, one for each designed luminosity. In particular, for lower luminosities are expected fewer interactions rates, so triggers thresholds can be lowered. Analysis involving low-p_T particles, such as the reconstruction of $J/\Psi \rightarrow e^+e^-$ events, will be possible only in the low-luminosity phase of LHC ($\mathcal{L} = 8 \cdot 10^{29} \div 10^{31}$ $cm^{-2}s^{-1}$), when lowered trigger thresholds will allow to retain a sizeable part of the signal.

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, giving a safety margin of 3. The L1 trigger table is reported in Table 5.1. The trigger rates are computed based on the L1 emulator (`L1Emulator` of `CMSSW`) responce to a Monte Carlo Minimum Bias samples. Calling $\xi = \frac{\#\ L1\ passed}{\#\ tot\ sample}$ the fraction of events in the sample which pass a certain L1, the rate is computed as

$$R = \sigma_{MinBias} \cdot \mathcal{L} \cdot \xi \tag{5.1}$$

In [90] and [91] there is the L1 Table for the LHC start-up (respectively for $8 \cdot 10^{29}$ $cm^{-2}s^{-1}$ and $1 \cdot 10^{31}$ $cm^{-2}s^{-1}$). In particular, electromagnetic thresholds (e/γ) are lowered enough to have a good sample of filtered events whit low-mass resonances. In Tab. 5.2 are summarized the changes of electromagnetic L1 triggers at the start-up. The *prescaled* triggers are triggers with high rates which are accepted only one every n times (to keep the total band width within $\sim \mathcal{O}(10$ kHz) in the low luminosities at the start-up), where n is the prescale of the trigger.

L1 Triggers	E_T Thresholds (GeV 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)
$\mu\otimes$jet	- (15$\otimes$40)	- (0.8)	15.1 (32.5)
MinBias		0.9 (1.0)	16.0 (33.5)
Total			**16.0 (33.5)**

Table 5.1: *Trigger Table forseen for $\mathcal{L} = 8 \cdot 10^{32}\ cm^{-2}s^{-1}$ ($10^{34}\ cm^{-2}s^{-1}$).*

L1 Triggers	Thresholds (GeV)	Prascales at $8 \cdot 10^{29}$ $\mathbf{cm^2s^{-1}}$	Prescales at $1 \cdot 10^{31}$ $\mathbf{cm^2s^{-1}}$
Inclusive Isolated e/γ	5	1	1
	8	1	1
	10	1	1
	12	1	1
	15	1	1
	20	1	1
Di-e Di-γ	5	1	1
	10	1	1
	15	1	1
Total e.m. L1 Rate		**0.157 KHz**	**1.856 KHz**
Total L1 Rate		**5.813 KHz**	**8.643 KHz**

Table 5.2: *L1 Trigger Seeds Table foreseen for the LHC start-up. There are shown the new thresholds, prescales, the total electromagnetic L1 trigger rate and the total L1 trigger rate.*

5.2.2 The High Level Trigger (HLT)

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 a farm of $\mathcal{O}(10^3)$ 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 exibility to the system, the only limitation being the total bandwidth and the number of processors. The system can be easily 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. 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.

Tab. 5.3 summarizes the HLT requirements at low luminosity to match the cumulative rate of 100 Hz, according to what have to be reconstruct in the final state. The bigger part of CPU time is spent to events with final states which require complete tracking reconstruction (muons, b-tag[2]).

HLT algorithms are encoded in `CMSSW` and organized as an ensemble of algorithms organized into *modules*. Each module is then processed sequentially, in the so called **HLT paths**. There are about 100 HLT path implemented in the code. In each analysis the

[2]The b-tag is the algorithm to identify jets deriving from a quark b hadronization and fragmentation.

Final States	CPU times (ms)
1 e, 2 e	160
1 μ, 2 μ	710
1 τ, 2 τ	130
jet$\otimes E_T^{miss}$	50
e$\otimes$jet	165
Inclusive b jets	300
Total HLT Rate at $8 \cdot 10^{29}$ cm^{2}s^{-1}	**138 Hz**
Total HLT Rate at $1 \cdot 10^{31}$ cm^{2}s^{-1}	**145 Hz**

Table 5.3: *HLT Trigger Table foreseen for the LHC start-up. There are shown, according to the reconstructed final state, the CPU time[89], and the final HLT rate.*

most adapted HLT path have to be identified. HLT paths implemented in `CMSSW` are `EDFilter` (see Chapter 4), i.e. the final module of the path returns a boolian value which decides if the event have to be stored or not. Being bool values, a final logic `OR` of all HLT paths is made. If at least one is passed, than the event is recorded.

The first module of an HLT path is the *seed*: this module reads the return value of the L1 triggers (which are boolean values too) and if it is a positive response then the subsequent modules in the path are processed. Each HLT path has its proper seed, generally one of the L1 triggers or the logic `OR` of more L1 triggers. The seed takes also into account the prescaling if the L1 seeded is prescaled.

When HLT decide that an event can be stored, all `RAW` and `DIGI` data are recorded. Data are organized into `ROOT` files (see Chapter 4) where also triggers information (`HLT` data) are added; for this purpose, each return (boolean) value of each HLT path is recorded into the file as a bit (where if it is a "`0`" this means HLT path not passed, and "`1`" means HLT path passed). Each path is a bit, and all bits are sorted at the initial part of the file, and can be read in the off-line analysis. Each bit position corresponds to a path. Also L1 trigger informations are stored in the file, such as HLT bits: another sequence of sorted bits is recorded, where each bit position corresponds to a L1 trigger response.

In the following I describe which triggers are used for the J/Ψ →e⁺e⁻ analysis, how they work and I validate them with the Monte Carlo samples of prompt J/Ψ's and J/Ψ's from b hadron decay.

5.3 Triggers for J/Ψ →e⁺e⁻ events

The favourable L1's electromagnetic triggers foreseen for the LHC start-up will allow study of low-mass resonances in the e^+e^- decay channel. The Trigger Tables with lowered L1's will be available only for the start-up phase, i.e. the first 50-100 pb^{-1}. So, in this time range, it has to be collected as much data as possible to do a reasonable analysis.
So it is needed:

- A good L1 trigger seeding, to lose as less as possible signal events, taking under control the band width rate at the same time.
- An HLT with low signal rejection for each applied module of the path.

In the following, the trigger developed for J/Ψ →e^+e^- events is validated on Monte Carlo sample generated for this work: L1 triggers can be simulated by the *L1 emulator* implemented in `CMSSW`, with the Trigger Tables foreseen for $\mathcal{L} = 8 \cdot 10^{29}$, $1 \cdot 10^{31}$ $cm^{-2}s^{-1}$. So results can be fed in the L1 seed of the HLT path, and triggers efficiencies can be obtained comparing passed events with the Monte Carlo truth.

5.3.1 The L1 Trigger Seeding

The L1 seed has to collect as much e^+e^- pairs as possible. To do this, Di-e/Di-γ L1 triggers have to be considered, compatibly with the expected electrons p_T spectrum. In Fig. 5.3 it is shown the p_T of the hardest electron of the e^+e^- pair vs. the p_T of the lowest one obtained by a Monte Carlo simulation of prompt J/Ψ events. The generation includes the preselection cuts $p_T \geq 3.5$ GeV/c and $|\eta| \leq 2.5$ to consider only events effectively "visible" by the detector. The choice of the threshold in p_T in the preselection is made considering that the lowest L1 trigger threshold on a single electromagnetic candidate (e/γ) is of 5 GeV in E_T (3.5 GeV/c takes into account the effect of the resolution of the ECAL Trigger Towers, which could be also of the order of ∼1 GeV[92]). From Fig. 5.3 it is possible to notice that p_T spectrum is peaked at 3.5 GeV/c and decreases for higher values; then, there can be possible cases of asymmetric energies pairs, with an electron/positron ≥5 GeV/c and the positron/electron ≤5 GeV.
Taking into account the available L1 seeds and the spectrum of e^+e^- decaying from a J/Ψ, it was chosen a L1 seed composed by the logic `OR` of the Di-e/Di-γ L1 trigger with threshold of 5 GeV (called `L1_DoubleEG05` in `CMSSW`) and the Inclusive Single e/γ L1 trigger with threshold of 8 GeV (`L1_SingleEG08`). These two L1 triggers are the ones with the lowest thresholds and not-prescaled at the same time.
The `L1_SingleEG08` is passed if at least one ECAL Trigger Tower with a reconstructed E_T ≥8 GeV is found, while the `L1_DoubleEG05` is passed if two distinct Trigger Towers

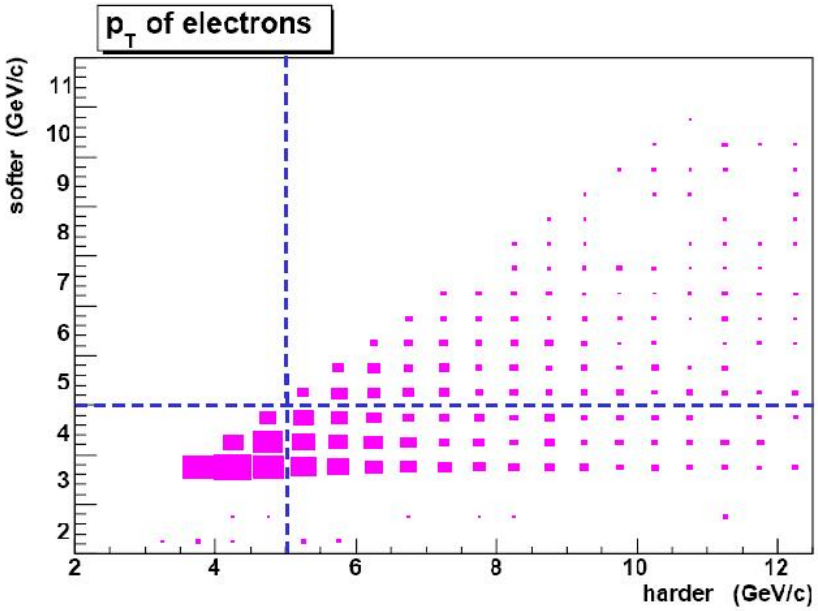

Figure 5.3: *p_T of the hardest electron vs. the p_T of the lowest expected in prompt J/Ψ events at LHC (10 TeV). The generation is preselected in p_T (≥3.5 GeV/c) and η (≤2.5) of e^+e^- pairs.*

with a reconstructed E_T ≥5 GeV are found. It has to be noticed that the $E_T \equiv E \cdot \sin(\theta)$ is computed with θ referred to the centre of CMS, and, being Trigger Towers domains of 5×5 ECAL crystals, the total E_T of the Trigger Tower is the sum of the single E_T of each crystal. A candidates pair needs to have e^+ and e^- enough apart to produce two distinct Trigger Towers of 5×5 crystals. This means that electron and positron have to be separated by a cone with $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \simeq 0.35$.

In Fig. 5.4 there is the distribution of the ΔR between the electron and positron in prompt J/Ψ events. Many events have close e^+e^- pairs, as it could be easily expected (higher values in electrons p_T correspond to enhanced J/Ψ's which give enough close pairs due to the Lorentz Boost), but not enough to be "not-visible" from the `L1_DoubleEG05` (ΔR >0.35 for most of pairs). Anyway, `L1_DoubleEG05` will consider the two highest E_T ECAL candidates. It is not unlikely that in some J/Ψ →e^+e^- events one of the two candidates creating the level 1 seed is not deriving from the J/Ψ decay but from some uncorrelated activity in ECAL (due to e/γ particles in the QCD event). These events are generally lost at the level of HLT (as described below).

In Tab. 5.4 are reported the L1 seed efficiencies and the number of passed events per pb^{-1} expected for prompt J/Ψ events and J/Ψ from B hadron decay events obtained by Monte Carlo Simulation. The J/Ψ from a B hadron decay sample has higher seed efficiencies (16% vs. 13%) than the prompt J/Ψ sample. In particular, the Di-e/γ trigger (`L1_DoubleEG05`) efficiency is higher more than a factor 2 (11% vs. 5%) between the two

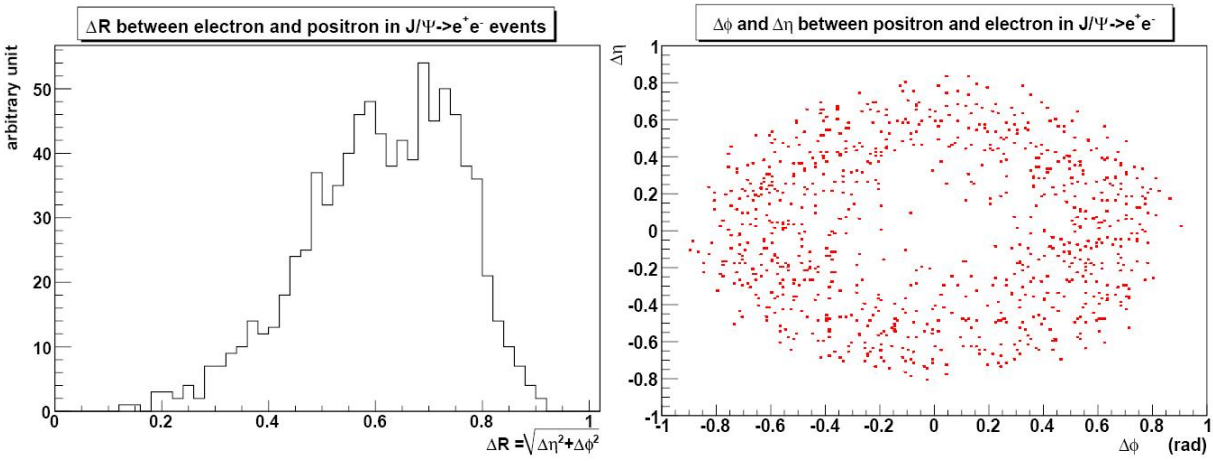

Figure 5.4: *ΔR (left) and $\Delta\eta$, $\Delta\phi$ distributions between electron and positron expected in prompt J/Ψ events at LHC (10 TeV). The generation is preselected in p_T ($\geq$3.5 GeV/c) and η ($\leq$2.5) of e^+e^- pairs.*

Sample	Prompt J/Ψ		J/Ψ from B hadrons	
L1's	evt/pb^{-1}	eff. (%)	evt/pb^{-1}	eff. (%)
L1_DoubleEG05	1670	5	2880	11
L1_SingleEG08	3670	11	3800	14
OR	4340	13	4190	16

Table 5.4: *L1 seed efficiencies and number of filtered events expected per pb^{-1} on J/Ψ's Monte Carlo samples.*

sample. This is due to the harder p_T spectrum of e^+e^- pairs of J/Ψ from B mesons decay (see Chapter 3).

The choice to require the logic `OR` is particularly good: increases the L1 seed efficiency of prompt J/Ψ events by a factor ∼2, from 5% to 13%.

5.3.2 The HLT Paths

Once events pass the L1 seed, they have to be processed by HLT path. The modules in the path have to accept J/Ψ →e^+e^- events as much as possible, keeping HLT rate within the foreseen band widths.

Dedicated HLT paths for J/Ψ →e^+e^- and ϒ →e^+e^- have been written and have been officially included in the `CMSSW` code few months ago[93][94]. Two different HLT paths have been developed for the LHC start-up, one for $\mathcal{L} = 8 \cdot 10^{29}$ cm^{-2}s^{-1} and one for $\mathcal{L} = 1 \cdot 10^{31}$ cm^{-2}s^{-1}, both for J/Ψ →e^+e^- and ϒ →e^+e^-.

For $\mathcal{L} = 8 \cdot 10^{29}$ cm^{-2}s^{-1} a relatively low rate of events is expected, so a level 2 HLT path has been chosen. In this path, only calorimetric informations are used.

This HLT locates the 5×5 crystals matrices in ECAL where Trigger Towers have been found by L1 seed. Starting from those regions, the *Hybrid Superclustering Algorithm* (in the barrel) and the *Island Superclustering Algorithm* (in the endcaps) are involved to reconstruct ECAL Super Clusters (SCs). In this way, there is no need to search in all η-ϕ ECAL geometry SCs, reducing the number of usages of superclustering algorithms in the event and so the timing of the HLT path. The SCs reconstructed in this phase of the HLT path are called **ECAL candidates**.

Once ECAL candidates are obtained, the HLT proceeds as follows:

- At least **two** ECAL candidates have to be found in the ECAL regions where L1 seed has given positive responses. Then, the two ECAL candidates have to satisfy the condition $E_T \equiv E_{SC} \cdot \sin\theta > 4$ **GeV**, where θ is referred to the CMS frame origin (0,0,0) (not to the real primary vertex). There is no case where more than two ECAL candidates can be found because the L1 seed is based on L1 triggers which sort the Trigger Towers in order of increasing E_T (see section 5.2.1) and only the first two candidates are considered.

  The 4 GeV thresholds it has been chosen because it has been estimated that the reconstruction efficiency of a SC with an energy deposit of at least 4 GeV within the 5×5 crystals domains located by L1 triggers is more than 90%[88]. In this way, it is possible to keep low trigger rates and high signal efficiencies at the same time.

- As next pass, it is applied **a "loose e-ID"** on ECAL candidates requiring only

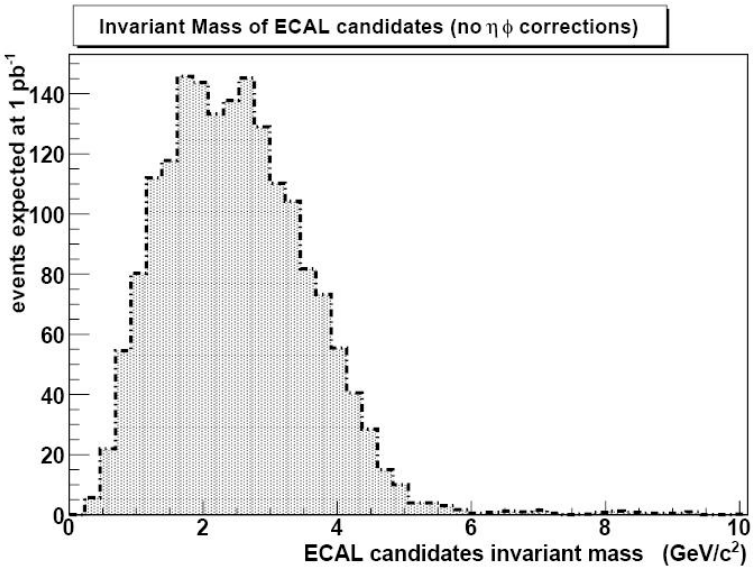

Figure 5.5: *Invariant Mass Distribution between two ECAL candidates in prompt J/Ψ →e^+e^- events.*

a selection on the cluster shape. It is used the $\sigma_{\eta\eta}$ introduced in Chapter 4 as selection parameter, with very loose thresholds (> 0.015 in the barrel and > 0.04 in the endcaps) to keep high passing efficiencies on signal events.

- Finally, it is required an **invariant mass filter** on the two ECAL candidates: for both of them, are used the E_{SC}, ϕ_{SC} and η_{SC} for the reconstruction of the invariant mass (see Appendix A), and it is required that the event must have a reconstructed invariant mass **between 2 and 4 GeV/c^2 in the J/Ψ →e$^+$e$^-$ case**. This loose range has been chosen because ϕ_{SC} and η_{SC} are refereed to the CMS frame origin and not to the real primary vertex, and ϕ is not corrected for the bending magnetic field, so the invariant mass distribution results to be more smeared than it can be expected by the only experimental resolution effects.

  In Fig. 5.5 there is an example of invariant mass distribution of ECAL candidates with no η and ϕ corrections in the case of J/Ψ →e$^+$e$^-$ events.

This path is implemented in the `CMSSW` code with the name `HLT_DoubleEM8e29_Jpsi`.

For $\mathcal{L} = 1 \cdot 10^{31}$ cm^{-2}s^{-1} it is expected that events rate increases by a factor ∼10, so the usage of the only calorimetric information could not be enough to keep the HLT rate acceptable. So the previous HLT path has been quite modified for this luminosity; in particular, it is a level 2.5 HLT, i.e. also pixels information are used. It has been chosen to not implement a level 3 HLT (i.e. with the full knowledge of tracks) to spare CPU

	HLT_DoubleEM8e29_Jpsi	HLT_DoubleEM1e31_Jpsi
ϵ(prompt J/Ψ)	38%	23%
ϵ(B→J/ΨX)	22%	10%
Rate 8e29	5.7 Hz	(0.2 Hz)
Rate 1e31	-	2.5 Hz

Table 5.5: *HLT paths efficiencies and trigger rates expected at the LHC start-up on J/Ψ's Monte Carlo samples. Efficiencies are expressed respect to events passing level 1 trigger requirements.*

time, assuming that the luminosity is enough low to guarantee low trigger rates (this issue has been then verified, see Tab. 5.5).

The ECAL candidates are supplied by the same L1 seed of the previous path, and this HLT path proceeds as follows:

- At least two ECAL candidates matching the L1 seed and with $E_T > 4$ GeV, as in the previous path.
- It is applied a "loose e-ID" as in the previous path, but requiring also a pixel matching in addition to the $\sigma_{\eta\eta}$ selection. Each ECAL candidate (SC) have to be matched with three aligned *hits* in Pixels detector layers within a window in ϕ in the direction of the SC. This procedure is very similar to the one implemented in the `GsfElectron` reconstruction; since this path is foreseen only for the start-up period, it is expected a lower number of hits occupancy in the pixels, so a larger $\Delta\phi$ window than in the usual reconstruction algorithm can be used. The pixels triplet hits is the same one which is used as *seed* for the GSF tracks reconstruction.
- Finally, it is required the same invariant mass filter of the previous HLT path.

This path is implemented in the `CMSSW` code with the name `HLT_DoubleEM1e31_Jpsi`.

In Tab. 5.5 there are the HLT efficiencies ϵ (defined as the ratio between the number of events passing the HLT and the number of events passing the L1 seed) and the expected HLT rates of the two paths. The highest HLT rates has been estimated in [93] considering true or fake electrons in Minimum Bias events, as it is usually made in trigger rates estimations. Three considerations can be made:

1) The trigger rates are acceptable for both luminosities. As a check, the path foreseen at $1 \cdot 10^{31}$ cm^{-2}s^{-1} has been applied also in the $8 \cdot 10^{29}$ cm^{-2}s^{-1} scenario: the resulting rate (0.2 Hz) shows that the requirement of "only" a level 2 HLT to increase the HLT efficiency is reasonable.

2) Starting from events which pass the L1 seed, there are more prompt J/Ψ events passing both HLT paths than in B→J/ΨX (by a factor ∼2), and this is due to the different action of the invariant mass filter on the two samples: due to the long mean life time of B hadrons (τ_B ∼1.6 ps) they cover a relatively long distance in the transverse plane (xy)[3] bringing to more displaced J/Ψ's than in prompt events. This mean that the invariant mass distribution results to be more smeared than in the prompt case.

3) Efficiencies seem to be low, for both the samples. Starting from the L1 seed, which already reject a big fraction of signal, only the order of 30% of events pass the HLT paths. If the preselection cut on electron p_T is increased from 3.5 GeV/c to 4 GeV/c in Monte Carlo generation it has been shown that HLT efficiencies (refereed to the events which pass the L1 seed) increase too[93]. This is due to the different number of events which pass the `L1_DoubleEG05` and `L1_SingleEG08` triggers in the two Monte Carlo preselections: if threshold is set to 3.5 GeV/c, there are more events with one $e^{\pm}$ at low p_T and the other above 8 GeV/c, passing only the `L1_SingleEG08`. The HLT paths for J/Ψ →e⁺e⁻ trigger include the requirement of two ECAL candidates with E_T above 4 GeV which rejects a big fraction of events with one $e^{\pm}$ p_T ∼3.5 GeV. As a check, it has been obtained the HLT paths efficiency refereed only to events which pass only the `L1_DoubleEG05`, i.e. computed as

$$\xi = \frac{n.\quad events\quad pass\quad HLT}{num.\quad events\quad pass\quad L1\_DoubleEG05} \tag{5.2}$$

In this way, it has been obtained $\xi = 85\%$ for the `HLT_DoubleEM8e29_Jpsi` and $\xi = 57\%$ for the `HLT_DoubleEM1e31_Jpsi`.

The only `HLT_DoubleEM8e29_Jpsi` and `HLT_DoubleEM1e31_Jpsi` paths seem to work very well only on events passing the `L1_DoubleEG05` and not on events passing only the `L1_SingleEG08`. This is a big restraint for a path which accept as L1 seed the logic `OR` of the two L1 triggers. To increase the number of signal events collected for the off-line analysis, it can be useful to associate another HLT path to the previous ones to recover events which pass only the `L1_SingleEG08`.

Among the HLT Trigger Tables foreseen for the LHC start-up (both at $8 \cdot 10^{29}$ cm^{-2}s^{-1} and $1 \cdot 10^{31}$ cm^{-2}s^{-1}), there is an Inclusive Single-e HLT path triggering events with a single electron with $E_T \sim \mathcal{O}(10$ GeV). This path is included in `CMSSW` with the name `HLT_Ele10_SW_L1R`, is a level 2.5 HLT and works as follows:

- The L1 seeding is made by the only `L1_SingleEG08`.
- In the 5×5 matrix crystals of the Trigger Tower located by L1 seed is applied the

[3] τ_B ∼1.6 ps corresponds to a covered distance of $\gamma c \tau_B \sim \frac{E_B}{m_B} \cdot 470$ μm, which can reach also to order of $1 \div 2$ mm.

	ϵ(prompt J/Ψ)	ϵ(B→J/ΨX)			
			Rate	**8e29**	1.7 Hz
HLT_Ele10_SW_L1R	49%	37%	**Rate**	**1e31**	20 Hz

Table 5.6: *Single Electron HLT efficiencies and trigger rates expected at the LHC start-up on J/Ψ's Monte Carlo samples.*

	ϵ(prompt J/Ψ)	ϵ(B→J/ΨX)
HLT_DoubleEM8e29_Jpsi	38%	22%
HLT_DoubleEM1e31_Jpsi	23%	10%
HLT_Ele10_SW_L1R	49%	37%
OR 8e29	64%	52%
OR 1e31	47%	38%
OR 8e29	∼2780/pb^{-1}	∼2180/pb^{-1}
OR 1e31	∼2040/pb^{-1}	∼1560/pb^{-1}

Table 5.7: *HLT efficiencies and number of collected events expected per pb^{-1} in J/Ψ's Monte Carlo samples.*

Hybrid SuperClustering Algorithm (barrel) or the *Island SuperClustering Algorithm* (endcap) to reconstruct the SC (ECAL candidate). Only events where the SC has $E_T \geq$10 GeV can proceed in the HLT path.

- Finally, the ECAL candidate (SC) have to be matched with three aligned *hits* in Pixels detector layers within a window in ϕ in the direction of the SC, to reject events where L1 seed is activated by photons.

In Tab. 5.6 there are reported the HLT trigger efficiencies (refereed to events which pass the L1seed) for J/Ψ →e⁺e⁻ events (both prompt and from B hadrons), and the HLT rates expected for the two start-up luminosities (obtained as usual on Minimum Bias events).

5.3.3 Final On-line J/Ψ →e⁺e⁻ Events Selection

In Tab. 5.7 a final report with HLT paths efficiencies is shown to compare how many events can be recovered requiring the logic OR between the J/Ψ →e⁺e⁻ dedicated HLT path and the single electron HLT path: the efficiencies increase of ∼40%, so the combination of a Single Electron HLT with the dedicated HLT seems to be a good idea to increase as much as possible the number of signal events selected for the off-line analysis.

Again in Tab. 6.21 there are also reported the number per pb^{-1} of signal events which pass the HLT for the luminosities foreseen for the LHC start-up. The numbers are ob-

tainable starting from the (Monte Carlo preselected) generated events expected at 1 pb^{-1} ($N^{MC}_{1pb^{-1}}$), computed in Chapter 3, i.e. ∼33400 prompt J/Ψ →e$^+$e$^-$ events and ∼26200 B→J/ΨX→e$^+$e$^-$X events. Defining ξ_{L1} the L1 seed efficiencies reported in Tab. 5.4 and ξ_{HLT} the HLT efficiencies defined in Tab. 6.21, the number of expected triggered events per pb^{-1} ($N^{trig}_{1pb^{-1}}$) is given by

$$N^{trig}_{1pb^{-1}} = N^{MC}_{1pb^{-1}} \cdot \xi_{HLT} \cdot \xi_{L1} \tag{5.3}$$

So ∼ 2000÷3000 prompt J/Ψ →e$^+$e$^-$ events per pb^{-1} and ∼ 1500÷2000 B→J/ΨX →e$^+$e$^-$X events per pb^{-1} can be collected for the off-line analysis after trigger selections.

Chapter 6

Measurement of prompt J/Ψ Differential Cross Section in the J/Ψ →e^+e^- Channel

In this final chapter is presented the analysis of the prompt J/Ψ differential cross section as a function of the J/Ψ reconstructed p_T in the decay channel J/Ψ →e^+e^-.

The measurement of this observable allows to discriminate which phenomenological model of the ones presented in Chapter 1 is the more adaptable to the data. Unlike the prompt J/Ψ polarization measurement, the cross section measurement needs a lower statistics and so it can be performed with the earliest data of LHC. Given the relatively high production cross section of heavy quarkonia at LHC ($\sigma_{c\bar{c}}$ ∼290 μb, see Chapter 3), this is one of the earliest analysis feasible at the LHC start-up. This work represents the first approach of a heavy quarkonia physics study in the e^+e^- decay channel, which can be considered complementary to the study made in the $\mu^+\mu^-$ decay channel[95].

The strong limitation of the study in the e^+e^- decay channel is given by the fact that electrons are reconstructed in the ECAL, where a huge amount of hadrons and soft E_T Minimum Bias particles arrive, and high thresholds triggers have to be implemented for the rejection of this background. This is opposed to the case of J/Ψ → $\mu^+\mu^-$ channel: muons are reconstructed in the muons chambers, which are placed outside the CMS solenoid; here, only high penetrating muons and neutrinos arrive, and lower trigger thresholds can be used for muons reconstruction (p_T ≥3 GeV).

Anyway, in the previous chapters we have shown that a good electron reconstruction can be made also at low E_T and that the low luminosities foreseen for the start-up allow a good trigger strategy for low-mass resonances decaying in e^+e^- pairs. Then, a dedicated e-ID based directly on ECAL super clusters (SC) has shown that signal reconstruction

efficiency can be increased by a factor ∼1.5 (see Chapter 4) keeping high the signal-to-background ratio. The J/Ψ →e$^+$e$^-$ channel is limited by the electron reconstruction and trigger algorithms, which place a threshold to $E_T \geq$5 GeV. Fortunately, theoretical models such as NRQCD and simple CSM start to differentiate for $p_T^{J/\Psi} \geq$9-10 GeV/c (see Chapter 1). Monte Carlo simulations made in Chapter 3 have shown that prompt J/Ψ's p_T spectrum in preselected events ($|\eta| < 2.5$ and $p_T > 5$ GeV/c) can be distinguishable for the different theoretical model of heavy quarkonia production.

In this chapter is presented the technique used in the measurement of prompt J/Ψ differential cross section and the study of the main backgrounds (already presented in Chapter 3 and 4) affecting this kind of measurement, and the off-line selections adopted to increase the rejection of the background. Then, it is exposed the statistical algorithm used to distinguish prompt J/Ψ's from the ones coming from the decay of B hadrons. Finally, some systematic errors in this measurement are estimated.

6.1 Differential Cross Section Measurement Technique

The e$^+$e$^-$ pairs are reconstructed in the region $|\eta| < 2.5$. For safety the differential cross section is measured and referred to the restricted region $|\eta| < 2.4$.
The production cross section σ of a given process is related with the number of produced events N refereed to a certain integrated luminosity by Eq. (6.1)

$$N = \sigma \cdot \int \mathcal{L} dt \tag{6.1}$$

so the cross section can be obtained as

$$\sigma = \frac{N}{\int \mathcal{L} dt} \tag{6.2}$$

The differential cross section as a function of p_T can be obtained deriving the (6.2)

$$\frac{d\sigma}{dp_T}(p_T) = \frac{1}{\int \mathcal{L} dt} \cdot \frac{dN}{dp_T}(p_T) \tag{6.3}$$

where $\frac{dN}{dp_T}(p_T)$ is the p_T derivative of $N(p_T)$, i.e. the number of events with a given p_T. The (6.3) function can be discretized such as in a histogram distribution of the discretized variable p_T in this way

$$\frac{d\sigma}{dp_T}(p_T) = \frac{N(p_T)}{\int \mathcal{L} dt \cdot \Delta p_T} \tag{6.4}$$

where Δp_T is the *binning width* of the histogram distribution, and $N(p_T)$ is the number of events which fill the p_T bin.

The measurement of the prompt J/Ψ differential cross section as a function of J/Ψ's p_T is obtained by filling an histogram with the reconstructed J/Ψ's p_T distribution. Basing this work on J/Ψ →e⁺e⁻ decay channel, the p_T of prompt J/Ψ's is obtained by the measurement of the e⁺e⁻ pairs: in particular, it is based on the conservation of the transverse component of the pairs four-momentum, i.e. on Eq. (6.5)

$$cp_{T_{J/\Psi}} = cp_{T_{e^+}} + cp_{T_{e^-}} \Longrightarrow p_{T_{J/\Psi}} = p_{T_{e^+}} + p_{T_{e^-}} \tag{6.5}$$

The obtained histogram has to be multiplied by a factor which takes into account the integrated luminosity that data refers, the binning width Δp_T, the number of prompt J/Ψ's reconstructed in each bin, and obviously the experimental efficiencies for which the measurement have to be corrected. Assuming that efficiencies, signal-to-background ratios depends from p_T, this factor will depend from J/Ψ p_T. So the total collected J/Ψ's are divided into various groups of bins in p_T, and the various multiplication factors are obtained for each p_T range of reference. The p_T ranges of each group are chosen to have at least a statistic of ∼800-1000 events per p_T range.

Obtained from the (6.4), the measurement is based on the following Eq. (6.6)

$$\frac{d\sigma}{dp_T}(J/\Psi) \cdot BR_{J/\Psi \to e^+e^-} = \frac{N^{reco}_{promptJ/\Psi}}{\mathcal{A}\epsilon_{trig}\epsilon_{reco} \cdot \lambda^{corr}_{trig}\lambda^{corr}_{reco} \cdot \int \mathcal{L}dt \cdot \Delta p_T} \tag{6.6}$$

where:

- $N^{reco}_{promptJ/\Psi}$ is the number of reconstructed prompt J/Ψ candidates in a given p_T bin. This is obtained by fitting the J/Ψ invariant mass spectrum with a gaussian signal hypothesis, as it will be better explained in the following. Anyway, background events have to be removed: reducible backgrounds can be reduced directly during the invariant mass fit, the irreducible background (given by non-prompt J/Ψ which however fill the invariant mass peak) is removed by a fit on the B hadrons decay length distribution.

- $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ is the the total efficiency for triggering and off-line reconstructing the J/Ψ events, as extracted from Monte Carlo. It takes into account also the geometrical acceptance ($\mathcal{A}$) of the detector ($|\eta_{e^\pm}| < 2.5$), and the efficiency due to the background subtraction.

- λ^{corr}_{trig} and λ^{corr}_{reco} are correction factors to the trigger and off-line efficiencies, respectively, as measured in data compared to the Monte Carlo simulation.

Those items have to be covered for the differential cross section measurement. In the next section, the backgrounds for this kind of analysis will be explained in detail.

6.2 The Backgrounds

There are two kinds of backgrounds:

- Events which present a similar final state composition in a similar phase space, and defined as *with the same topology.*
- Events which do not present a similar final state composition from a physical point of view, but detectors give the same response of the signal events from a technical point of view and so a similar "fake" final state composition.

In the measurement considered in this work, events where an e$^+$e$^-$ pair can be identified in a phase space region compatible with the one covered by the decay of a J/Ψ[1], i.e. at low p_T regions. Two processes represent the background of prompt J/Ψ →e$^+$e$^-$ events[2]:

- Non-prompt J/Ψ →e$^+$e$^-$.
- Minimum Bias and generic QCD events which can have included true or fake pairs with an invariant mass range near 3.1 GeV/c^2.

6.2.1 Non-prompt J/Ψ →e$^+$e$^-$ background

In those events J/Ψ's decay in e$^+$e$^-$ pairs and are generated not in the fundamental interaction during p-p collision but from the weak decay of B hadrons, in particular B mesons (B^0, B$^\pm$, B$_s$, B$_c$), as it was explained in Chapter 3. This background is not reduced by the various selections used for the e-ID. This represents an irreducible background. Anyway, the long mean life time of B mesons brings to displaced secondary vertices for the J/Ψ →e$^+$e$^-$ decays. A statistical method based on the measured distance between primary and secondary vertices allows to determinate the fraction of prompt and non-prompt events after all e-ID and off-line selections; this method will be better explained in the following sections.

6.2.2 Generic Hadronic background

In Chapter 4 it has been said in advance that the huge amount of background which has to be expected in this analysis (such as in every analysis which involves the ECAL detector) is given by hadronic events, such as Minimum Bias (σ ∼50 mb) and generic QCD events with production of various hadronic jets (σ ∼1 mb).

[1] i.e. e$^+$e$^-$ pair with an invariant mass close to the J/Ψ mass.

[2] The level of the Drell-Yan background in the same mass window is computed to be less than 1% with respect to the other background sources, and hence neglected from now on.

The Monte Carlo samples of this kind of background are simulated by `Pythia` generator and were yet presented in Chapter 3. Those events have not a topology similar to the signal: here, a big amount of hadronic jets (more or less collimated) are produced. Those jets contains stable[3] charged hadrons (most of them are expected to be $\pi^{\pm}$) or also true electrons which are generated for the most part by semileptonic decays of unstable particles. It can be that a charged hadron (or more than one) leaves an energy deposit in the ECAL crystals by nuclear interaction which simulates a SC left by an electron. If hadron is charged, then, there is also a charged track (reconstructible both with the Kalman Filter and the Gaussian Sum Filter algorithms) which match with the SC, so an e-ID algorithm can identify it as a full-blown reconstructed electron, but it is a *fake* in reality. Another example of *fake* is given by γ's (which originated a SC in the ECAL) emitted collinear to charged tracks.

An e^+e^- pair in hadronic events can be emulated by combinatorial coincidences of two reconstructed fakes, of one fake and one true electron, or also of two true electrons. The requirement of two electron candidates with opposite charges (observing the bending in the solenoidal filed of the matched tracks) can reduce the number of combinatorial pairs in hadronic background. Then, in Chapter 4 it has yet shown how to customize opportune selections for the e-ID to reject as much as possible the rate of fakes reconstruction. Also the trigger system (L1+HLT) already reduces the number of events, most of them from Minimum Bias. Anyway, the huge cross sections for those events makes not so rare that this background passes all on-line and off-line selections. Then, the low-p_T nature of these events, brings to have combinatorial pairs in a phases space region such that the most part of their invariant mass distribution fill the range between 1 and 4 GeV/c^2.

The simulation of this background is very difficult: a huge statistic is needed, because are expected ∼50 billions of Minimum Bias events per pb^{-1} and ∼1 billion of QCD events per pb^{-1}. In the CMS collaboration, this kind of events are generally produced in large numbers and placed in DBS area[86]. The production is made dividing the sample on $\hat{p}_T$ bins, in order to simulate more accurately events at higher $\hat{p}_T$ (since the cross section falls steeply at large $\hat{p}_T$)[4]. The division in $\hat{p}_T$ bins is the one reported in Tab. 3.4 in Chapter 3. In particular, the first four bins have to be considered (the ones with the higher cross sections): they contain almost all the Minimum Bias and QCD events (correctly weighted) expected for a given reference luminosity. The CMS production foresees ∼1÷10 million of events generated for each $\hat{p}_T$. For everyone the e-ID algorithm selections optimized for the J/Ψ →e⁺e⁻ reconstruction and identification are applied. The result is that no event

[3]It is defined as *stable* a particle with a mean life time higher than the time required to cross through all the detector.

[4]The cross section of a subprocess (at partonic level) should increase as the energy increase: anyway, the partons pdf's decrease as the energy increases, and so the production cross section decreases too.

of the firsts two bins sample passes the selections and only two events of the third sample passes the selection. However, the statistics is low (compared to the number of events expected): each sample referees to an integrated luminosity of just $10^{-4} \div 10^{-3}$ pb^{-1}. So the presence of significant background from Minimum Bias at low-$\hat{p}_T$ cannot be excluded on the basis of this simulation.

This represent the big problem for the analysis of low-mass resonances decaying in e$^+$e$^-$: the statistics of available Monte Carlo samples is not enough. On the other hand, it is not possible to simulate billions of events (each event size is ∼1 MB, and the CPU time needed for the Full Sim is ∼5 minutes per event, see Chapter 4). The Fast Sim of the detector response decreases the CPU time needed for the detector response simulation (∼2-3 seconds) but the problem of big amount of MB remains. I contribute to the study, started in the CMS collaboration, for an "electromagnetic enrichment" of hadronic background samples, i.e. the implementation of some filters at generator level which pass to the detector simulation only events which are probable to pass the e-ID selections. But this is a non-trivial challenge because it is not easy to predict the topology of events contributing after all electromagnetic selections. An example of implemented filter is to pass only events with collinear charged hadrons, grouping them within a cone with a certain open solid angle and with a total p_T above a certain thresholds. The study have to optimize those thresholds to have a good filter rejection, and, at the same time, no rejection of events which should have passed the reconstruction and identification selections.

The starting point of this study has been another study on the possible topologies of hadronic events. This part of the analysis was made by me, and in the following I will report some results about it. Anyway, the study of the background topology was important to the choice of selections of the e-ID algorithm and for the optimization of thresholds presented in Chapter 4, and for the supplementary off-line selections for the differential cross section measurement.

The Minimum Bias and QCD events Topology

To study the topology of hadronic events it was made a match between ECAL SCs and Monte Carlo truth to "tag" each SC with the particle which could have left energy deposit with good probability. The matching criteria are the following:

- SC has to include a Monte Carlo particle within a $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.06$ cone.
- If more than one particle in the event satisfy the previous criteria, the one with the lowest $\frac{1}{p_T^{MC}} - \frac{1}{E_T^{SC}}$ value is kept[5].

[5]It is used the same matching criteria of the J/Ψ →e$^+$e$^-$ dedicated e-ID for the Monte Carlo truth as it should be a track

MC particles	$\hat{p}_T\varepsilon[0,\infty)$	$\hat{p}_T\varepsilon[30,\infty)$
$\mathbf{p}/\bar{\mathrm{p}}$	8.7%	10.0%
$\mathbf{K}^{\pm}$	16.2%	16.5%
$\pi^{\pm}$	41.4%	41.3%
$\mathbf{e}^{\pm}$	8.8%	8.6%
γ	20.5%	19.5%
$\mathbf{n}/\bar{\mathrm{n}}$	4.4%	4.1%

Table 6.1: *Background composition (Monte Carlo truth matched with SCs) for low $\hat{p}_T$.*

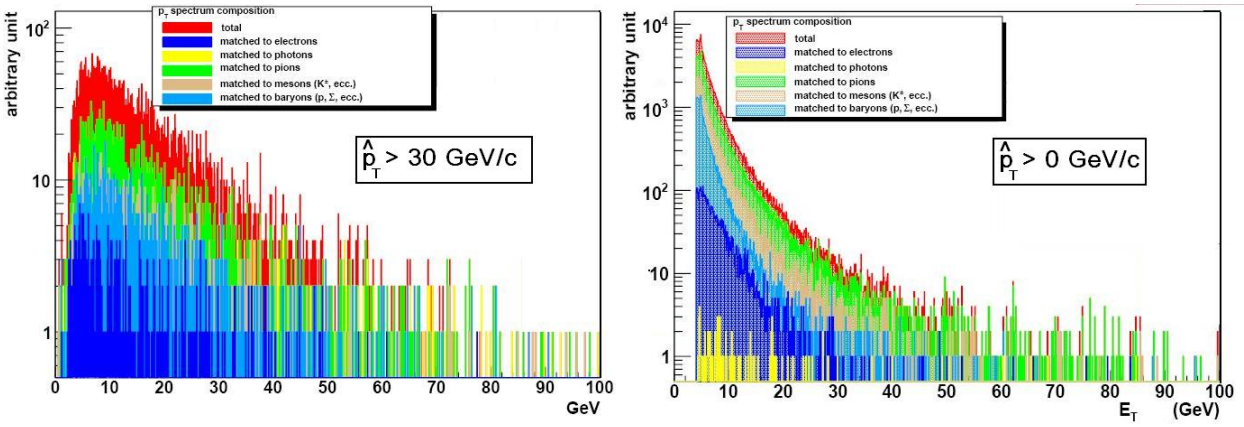

Figure 6.1: *E_T distribution of ECAL SCs in the two background samples (it is specified also the matched Monte Carlo truth).*

Two Monte Carlo sample were considered: one with $\hat{p}_T$ bin $30\div\infty$ considering official samples in DBS listed in Tab. 3.4 in Chapter 3, and the other simulated privately with the Fast Sim with $\hat{p}_T$ bin $0\div\infty$, i.e. containing basically Minimum Bias events with the (rightly weighted) rest of hard QCD events. The sample $\hat{p}_T$ bin $30\div\infty$ contains $\sim 6\cdot 10^6$ events; the sample $\hat{p}_T$ bin $0\div\infty$ was simulated with $\sim 6\cdot 10^7$ events.

In Tab. 6.1 are reported the composition of the SC matched to Monte Carlo truth for the samples: the composition is quite the same for the two samples. SCs seem originated by $\pi^{\pm}$'s energy deposit for the big part, but there is also a pure electromagnetic component ($\mathrm{e}^{\pm}$ and γ) ($\sim$30%). What changes between the two samples is the E_T deposited in the ECAL, as it is shown in Fig. 6.1.

It is interesting to study the origin of the pure electromagnetic component. In particular, the presence of so many photons could be suspicious. For this, it has been obtained also the "mother particles" from which photons are originated. In both sample $\sim$90% of π^0 and $\sim$10% of η are found. To better understand the problem, it has been also searched

the presence of particles collinear with photons, finding true soft p_T electrons and charged $\pi^{\pm}$ in the big part of cases. Charged hadrons in the same angular region of $\pi^0 \rightarrow \gamma\gamma$ and $\eta \rightarrow \gamma\gamma$ decays can account for a good amount of *fake* electrons. Also the following decays:

- $\pi^0 \rightarrow e^+e^-\gamma$: there is a non-negligible probability (Br∼1%[63]) of photons emitted collinear to electrons by boosted π^0.
- $\eta \rightarrow \pi^+\pi^-\gamma$ (Br∼1%[63]).

play a role, since the presence of charged particles collinear to photons makes possible the reconstruction of *fake* electrons (i.e. SCs matched with charged tracks).

The next step is to see how the composition changes applying the various e-ID selections (it is considered the match with KF tracks). For this purpose, it is considered the sample with $\hat{p}_T$ >30 GeV/c, because it gives a bigger statistics of events after e-ID selections. Here, the number of events surviving after selections expected per pb^{-1} are reported:

- composition: e$^{\pm}$-82.5%, $\pi^{\pm}$-12.7%, K$^{\pm}$-2.8%, p/$\bar{\text{p}}$-2.0%.
- number of candidate e$^+$e$^-$ pairs in the J/Ψ invariant mass region (3.1 GeV/c^2 ± 200 MeV/c^2): 373/pb^{-1}
- number of candidate pairs which are real e$^+$e$^-$ pairs: 248/pb^{-1}.
- (for candidates which are true electrons) composition of the mother particles: π^0-3.2%, $\tau^{\pm}$-4.8%, D-mesons-33.3%, B-mesons-58.7%.

After all electromagnetic selections the e$^+$e$^-$ pair candidates are for the most part effective e$^+$e$^-$ pairs. In Tab. 6.2 it is reported the composition of the mother particles for the all 248 e$^+$e$^-$ pairs/pb^{-1}. It is possible to see that the big fraction of real electrons comes from semileptonic decays of B-meson and D-meson, i.e. decays where the quark b→ $\bar{\text{c}}$e$^+\nu_{\text{e}}$ with the quark $\bar{\text{c}}$ →e$^-\bar{\nu}_{\text{e}}$X. In these decays, generally the e$^+$e$^-$ pair does not have a common vertex.

So the main features of the hadronic background topology of events which should pass the various e-ID selections are events enriched of semileptonic decays in electrons of B mesons and D mesons. Anyway, there is also an important part of background deriving from $\pi^{\pm}$'s and γ's which can be reduced by e-ID selections but not completely removed. Actually, there are some study about filters at Monte Carlo generator level which can filter from the detector simulation events where hadrons does not satisfy some conditions like collimation with charged particles above a certain threshold considering also the bending

hardest					
π^0	0	0	2	4	
$\tau^{\pm}$	1	1	4	7	
D	3	4	27	49	
B	5	7	48	85	
	π^0	$\tau^{\pm}$	**D**	**B**	**softest**

Table 6.2: *Composition of the mother of the e^+e^- particles after J/Ψ →e^+e^- dedicated e-ID selection. The table is organized as a matrix with the mother particles of the hardest electrons in the columns and with the mother particles of the softest electrons in the rows. The numbers are reconstructed J/Ψ candidates (each event may have more than one reconstructed candidate).*

in the solenoidal field, etc. The filters will be implemented for the official generation of large samples of background (∼20 millions for FullSim and ∼100 millions for FastSim) starting only from this winter. So for the analysis presented in this work will be used as reference background sample the one with $\hat{p}_T$ >30 GeV/c. Filters select events to pass which have charged particles which can be grouped within a cone and satisfying the following requirements:

- $\Delta R_{cone} = 0.2$
- total p_T of charged particle in the cone have to be ≥5 GeV/c.
- the ratio total p_T of charged *hadrons* over total p_T of $e^{\pm}$ or γ (in the cone) ≤ have to be ≤0.5.

With this implementation filter reach a rejection factor of passed events of $\sim \frac{1}{20}$.

6.3 Off-line Selections for Prompt J/Ψ →e⁺e⁻ Measurement

Over the e-ID algorithm, other off-line selections are used to select more properly signal events. Main sources of background derives, at this state, from true e^+e^- pairs deriving from heavy flavour (b, c) semileptonic decays and J/Ψ →e^+e^- decays generated from b hadrons decays. The non-prompt J/Ψ's is the more difficult one to reject, because it has a topology essentially identical to the prompt J/Ψ's, with the exception of the displaced vertex of the decay J/Ψ →e^+e^-.

The usage of btag algorithms[96] could be an important tool for the discrimination of

non-prompt J/Ψ events. Anyway, this analysis could be used also for the measurement of the differential cross section of J/Ψ's from B mesons decay, and the usage of btag algorithms to identify and reject non-prompt J/Ψ's could limitate the possible goals of the analysis. In section 6.1 it has explained that for the differential cross section measurement is important to determine the **number** of collected prompt J/Ψ's ($N^{reco}_{promptJ/\Psi}$) and not to discriminate *each* prompt event, so the usage of btag algorithm results to be also unnecessary. The $N^{reco}_{promptJ/\Psi}$ measurement will be made using a statistical fit which is able to estimate the fraction of prompt J/Ψ events in the total amount of collected J/Ψ's, as it will be better explained in the next sections.

So, at least at level of off-line selections, there is no need to discriminate prompt and non-prompt J/Ψ events, which can be considered together as the *signal*. Only hadronic events remain as real background. They are already well rejected by e-ID algorithm, as explained in Chapter 4. Anyway, studying the background topology, it has been shown that after the e-ID candidate pairs made by true e$^+$e$^-$ pairs are the big part of what remains. But those electrons are the result of B and D mesons decays. As each B decay, those electrons are displaced from the primary vertex. Then, it is not necessary true that e$^+$ and e$^-$ directions are nearby close as in J/Ψ →e$^+$e$^-$ events. Rather, in background events candidate pairs are the result of all the possible combination of ECAL SCs which pass the e-ID selections, so each candidate is often displaced from the other.

So two selections in addition to the others of the e-ID are used in this analysis to increase the rejection of background events, trying to reduce once more the candidate pairs composed by real e$^+$e$^-$ pairs:

- **A selection based on δ, angle between the two candidate electron/positron**, to reduce the combinatorial background.
- **A selection based on the distance of closest approach of the electron tracks from the primary vertex**, to reduce the electron/positron pairs from decays of B and C mesons, as described in the previous section.

No isolation criteria (see Chapter 4 for the definition) of the electron tracks was required because the selection should be model-independent (NRQCD foresees an enhanced hadronic activity around the prompt J/Ψ's direction).

The first selection reduces the background pairs at large angle. In Fig. 6.2 are shown the distribution of the δ angle between the candidate pairs for signals (prompt and non-prompt) and background (in the case of **real** e$^+$e$^-$ pairs). A selection of events with cosδ >0.85 reduce ∼28% of the background of real e$^+$e$^-$ pairs, keeping practically safe

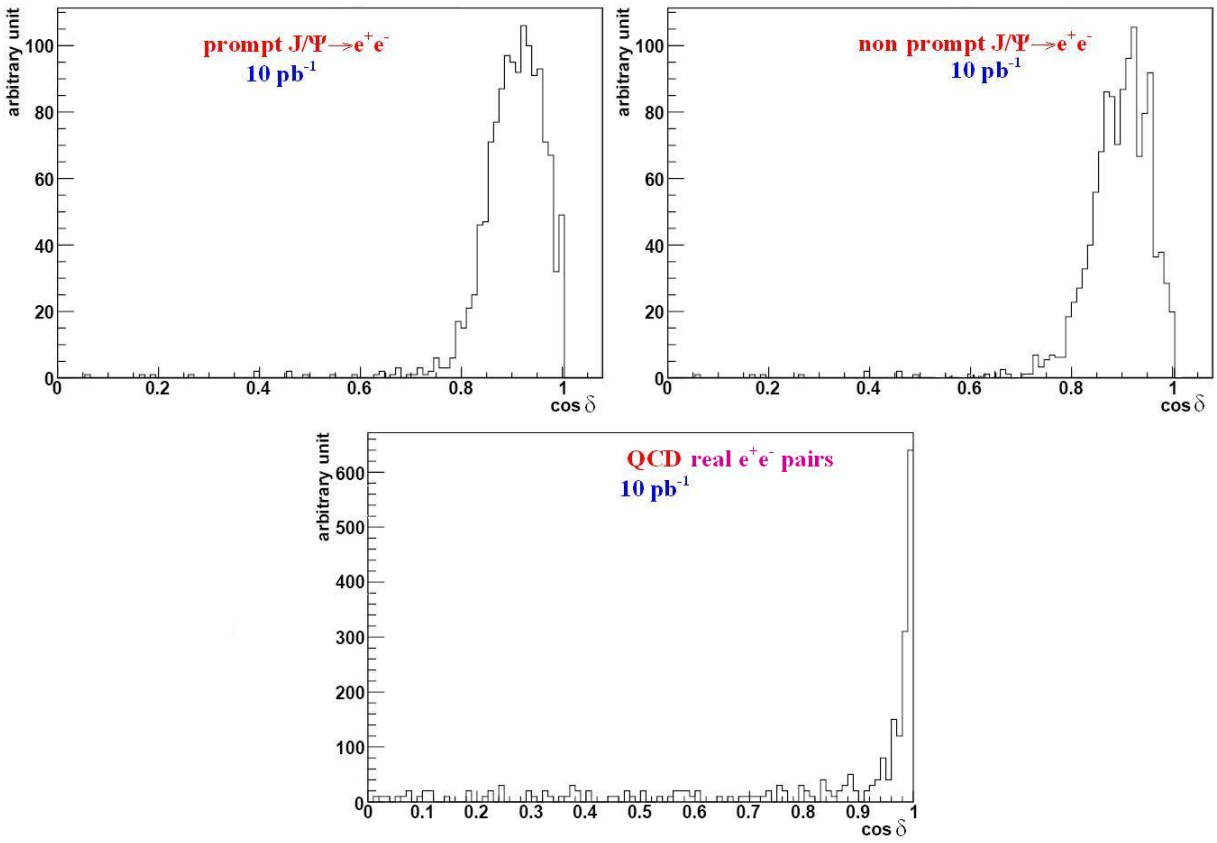

Figure 6.2: *cosδ distribution after e-ID selections for signals and QCD background (number of real e^+e^- pairs) expected after 10 pb^{-1}.*

hardest					
π^0	0	0	2	1	
$\tau^\pm$	0	1	1	1	
D	2	1	9	37	
B	2	0	36	20	
	π^0	$\tau^\pm$	**D**	**B**	**softest**

Table 6.3: *Composition of the mother of the e^+e^- particles after J/Ψ →e^+e^- dedicated e-ID and cosδ selection.*

all signals events. It is interesting to see the topology of background events rejected by this cut: in Tab. 6.3 there is an equivalent matrix such as the 6.2, after the cosδ selection applied in addition to the e-ID. It is possible to see a big reduction of pairs where both electrons derive from the decay of a B meson (or the decay of a D meson), while the case of b→ $\bar{c}e^+\nu_e$ with $\bar{c}$ →$e^-\bar{\nu}_e$X cascade decays (where an electron is from B and the other from D) is less affected by this cut. In the other case, a pair originated by two B mesons or two D mesons results from the combinatorial and so it has generally high value of δ.

The other off-line selection is based on the distance of the matched track from the primary vertex. The distance is computed by the so called **closest approach of the track** (distance of the closest approach, DCA). The closest approach is the nearest point of the extrapolated trajectory of the track to the primary vertex. In Fig. 6.3 there is a schematization of the DCA. Geometrically, it is the signed distance in the xy plane between the straight line tangent to the closest approach of the track to the CMS frame origin (from which the ϕ coordinate of the track is retrieved) and the primary vertex with coordinates ($PrimVx_x$,$PrimVx_y$)[97]. With this linear approximation, the DCA can be computed as the distance point-straight line in the plane, i.e.

$$PrimVx_x \cdot \cos\phi + PrimVx_y \cdot \sin\phi \tag{6.7}$$

The approximation in Eq. 6.7 is valid when the track has infinite or enough high momentum.

In Fig. 6.4 there are shown DCA distribution for prompt J/Ψ →e$^+$e$^-$ events and QCD events. In Tab. 6.4 there is the equivalent of 6.2 and 6.3 after the DCA selection is applied. A threshold of 100 μm has been chosen (it is expected that this cut reject inevitably a fraction of B→J/Ψ →e$^+$e$^-$ events).

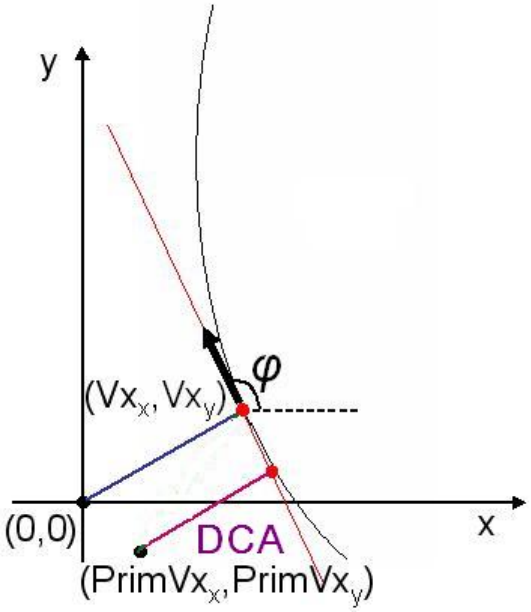

Figure 6.3: *Schematization of the distance of the closest approach (DCA) of a track.*

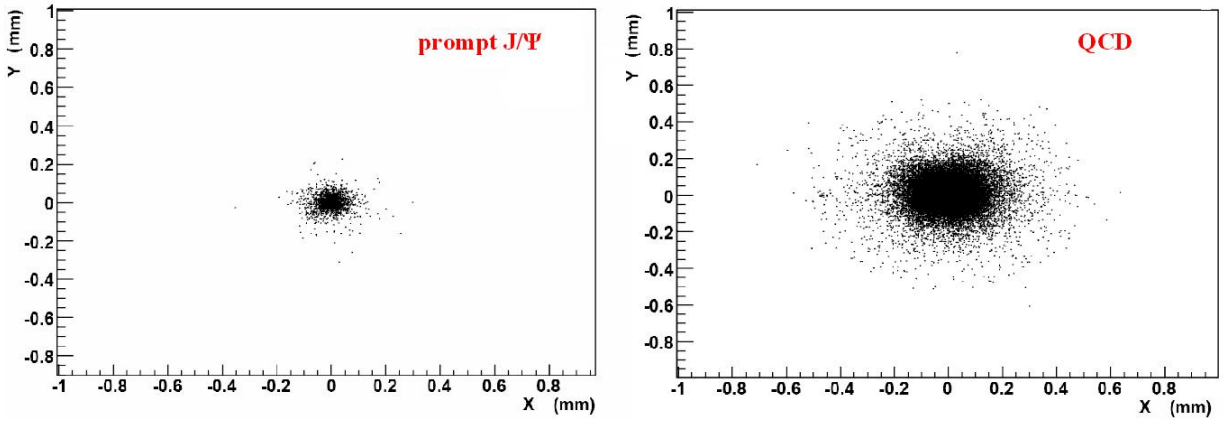

Figure 6.4: *Distribution of the DCA of tracks from the Primary Vertex for prompt signal (tracks of candidates associated to real e^+e^- at MC level) and QCD background (all candidates) after e-ID selections. Here is considered the match with KF tracks.*

hardest					
π^0	0	0	0	0	
$\tau^{\pm}$	1	1	0	1	
D	1	1	1	2	
B	1	1	3	5	
	π^0	$\tau^{\pm}$	**D**	**B**	**softest**

Table 6.4: *Composition of the mother of the e^+e^- particles after J/Ψ →e^+e^- dedicated e-ID and DCA selection.*

6.3.1 Off-line selections efficiencies of J/Ψ →e⁺e⁻ events.

For the measurement of J/Ψ's production differential cross section the candidate e^+e^- pairs have to be found: **candidates which pass all e-ID criteria and the two selections presented in the previous section with an invariant mass between (3.1 GeV/c² ± 200 MeV/c²) are considered as candidate e⁺e⁻ pairs generated by a J/Ψ decay**. It is chosen a range of 200 MeV/c² because it is nearby the resolution of the invariant mass peak of a J/Ψ →e^+e^- measured with the ECAL expected at the LHC start-up, as it was yet explained in Chapter 4.

In Tab. 6.5 are reported the various efficiencies of selections of events after each cut applied in cascade. Signals (prompt and non-prompt) efficiencies are computed by the number of events where the correct e^+e^- pair deriving from the J/Ψ decay passes the selections. For the hadronic background is reported the number of e^+e^- candidate combinations which survive after each selection. The triggers (both L1 and HLT) considered are the ones foreseen for the start-up luminosity of $\mathcal{L} = 8 \cdot 10^{29}$ $\mathbf{cm^{-2}s^{-1}}$.

In Fig. 6.5 it is shown the invariant mass distribution of the reconstructed pairs expected at 10 pb^{-1} applying e-ID, DCA and cosδ selections, considering both cases of KF tracks and GSF tracks matching in the e-ID. Only the background deriving from the $\hat{p}_T$ >30 GeV/c QCD sample is shown.

The implementation of the HLT, and the addition of the two off-line selections based on the DCA and on the cosδ rejects a big fraction of background. Minimum Bias events are not present in the Monte Carlo sample with enough statistic to be included in the analysis. However, a check was made with the private Monte Carlo sample simulated in FastSim with $\hat{p}_T \varepsilon [0, \infty)$ (Minimum Bias included). In Fig. 6.6 is shown the invariant mass distribution for the two Monte Carlo background samples requiring only the L1 trigger

prompt J/Ψ	# events	**non-prompt J/Ψ**	# events	**Background**	# pairs
L1 Trig.	40100		41860		$2.94{\cdot}10^8$
HLT	27650		21760		$4.94{\cdot}10^7$
E_T	22720		19280		$4.36{\cdot}10^6$
9/25	20640		17500		$1.14{\cdot}10^5$
H/E	18900		15310		31420

Tkr match	KF	GSF		KF	GSF		KF	GSF
Ang. match	11690	10180		11580	10040		7800	4530
Bend Comp.	10900	8950		10640	8740		4250	2635
DCA	9460	7370		4360	3570		1200	832
cosδ	8580	6620		3620	3020		720	450

Table 6.5: *Number of events with reconstructed and identified e^+e^- pairs per **10 pb^{-1}** in the invariant mass range (3.1 GeV/c^2 ± 200 MeV/c^2), after e-ID, DCA and cosδ selections. The considered triggers are for the low luminosity ($\mathcal{L} = 8 \cdot 10^{29}$ $cm^{-2}s^{-1}$ scenario).*

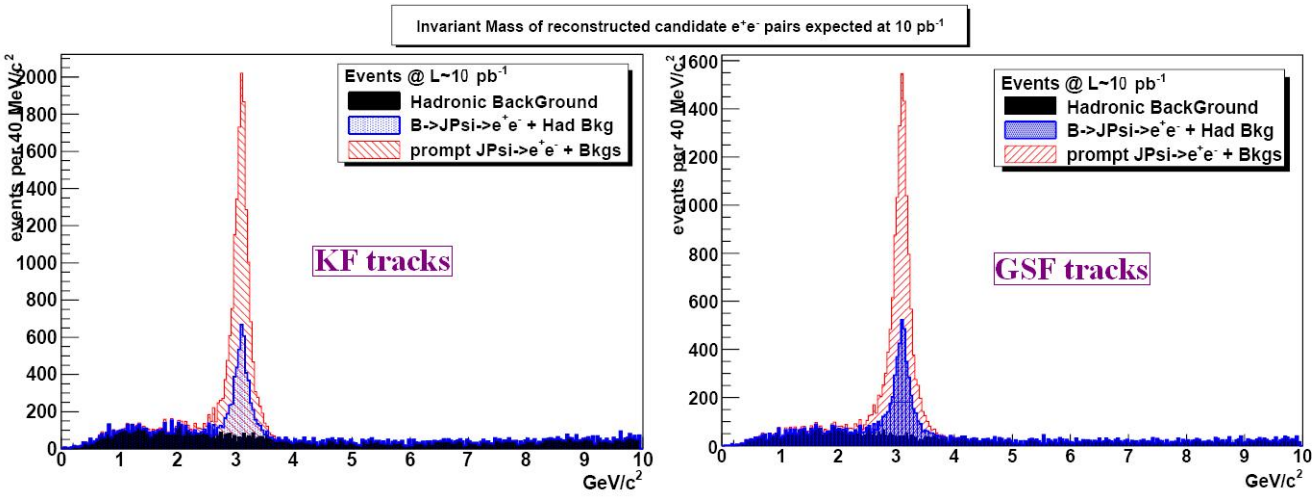

Figure 6.5: *Distribution of the invariant mass reconstructed using candidate e^+e^- pairs after on-line and off-line selections (e-ID, DCA, cosδ). There are shown distributions requiring a match with KF and GSF tracks in the e-ID.*

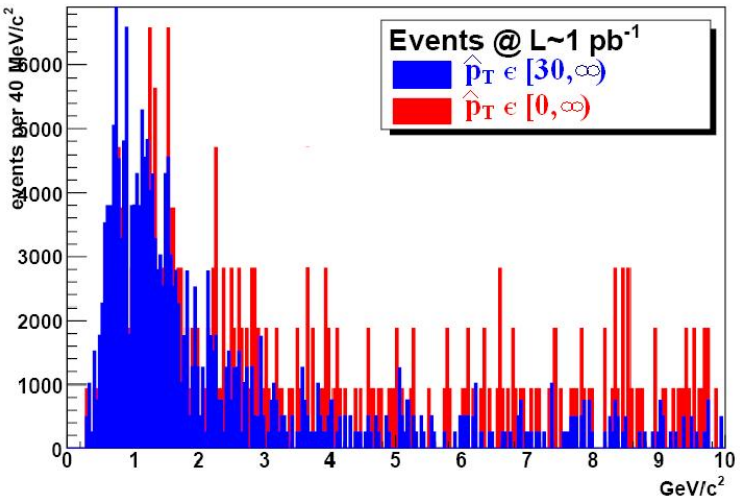

Figure 6.6: *Distribution of the invariant mass reconstructed using candidate e^+e^- pairs after the e-ID selections, for the official $\hat{p}_T\varepsilon[30,\infty)$ sample (blue) and for the private $\hat{p}_T\varepsilon[0,\infty)$ sample (red). Both samples are normalized to the integrated luminosity of 1 pb^{-1}.*

and a reduced set of e-ID selections (in order to keep sizeable statistics in the sample), as expected at 1 pb^{-1}: the two distribution seems to be similar, showing that Minimum Bias and more in general events with $\hat{p}_T$ <30 GeV/c^2 do not contribute sensitively to the background remnant after all selections, and that background is kept enough under control in the analysis.

6.4 Differential Cross Section measurement

In section 6.1 it has been shown that the differential cross section measurement in prompt J/Ψ's production is based on Eq. 6.6. In particular, the following measurements have to be made:

- $N^{reco}_{promptJ/\Psi}(p_T^{J/\Psi})$ number of reconstructed **prompt** J/Ψ's for a given J/Ψ p_T bin.
- $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ total acceptance, triggers and reconstruction efficiency.
- $\lambda^{corr}_{trig}\lambda^{corr}_{reco}$ correction factors to the efficiencies.

6.4.1 $N^{reco}_{J/\Psi}$ measurement

The number of prompt J/Ψ's is generally a function of the p_T and η of the J/Ψ ($N^{reco}_{promptJ/\Psi}$=$N^{reco}_{promptJ/\Psi}(p_T^{J/\Psi},\eta_{J/\Psi})$). It is obtained by the candidate pairs mass spectrum for each J/Ψ's p_T, similar to the mass distribution in Fig. 6.5 but requiring only events

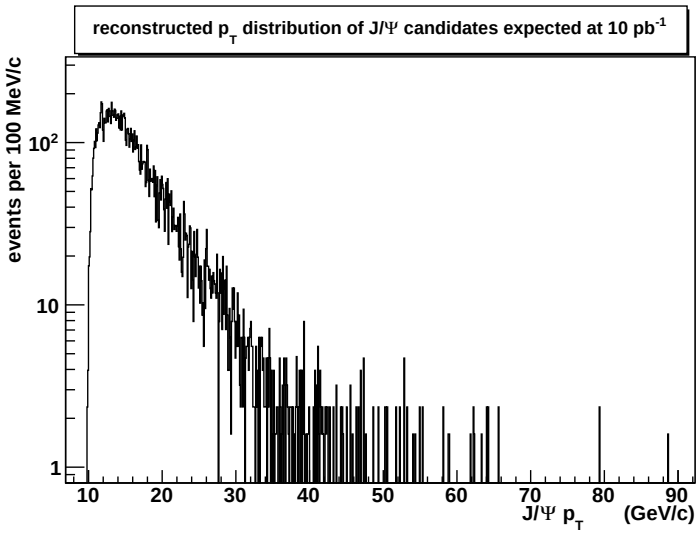

Figure 6.7: *Distribution of the reconstructed p_T of the J/Ψ candidates expected at 10 pb^{-1}. Here are considered e^+e^- candidate pairs requiring a match with KF tracks during the e-ID.*

with candidate pairs giving the p_T value, obtained using Eq. 6.5. The total number ($N^{reco}_{J/\Psi}$) of reconstructed J/Ψ's (prompt and non-prompt) in each $p_T^{J/\Psi}$ bin can be obtained by fitting the di-electron invariant mass distribution (m_{ee}) with a background hypothesis and a signal hypothesis. As anticipated in Chapter 4, signal distribution is affected totally by experimental resolution, so it is well described by a gaussian distribution:

$$\mathcal{G}(m_{ee}|A_{J/\Psi}, m_{J/\Psi}, \sigma_{J/\Psi}) = \frac{A_{J/\Psi}}{\sqrt{2\pi}\sigma_{J/\Psi}} \cdot \mathrm{e}^{-\frac{(m_{ee}-m_{J/\Psi})^2}{2\sigma_{J/\Psi}}} \tag{6.8}$$

The QCD background invariant mass distribution is indeed well described by a linear function, which approximates the background shape limited in the mass range $[m_{min},m_{max}]\equiv$[2.9 GeV/c^2,3.3 GeV/c^2]:

$$\mathcal{B}(m_{ee}|A_{bkg}, K_{bkg}) = A_{bkg} \cdot \left(\frac{1}{m_{max}-m_{min}} + K_{bkg} \cdot \left(m_{ee} - \frac{m_{min}+m_{max}}{2}\right)\right) \tag{6.9}$$

where K_{bkg} is the angular coefficient of the linear relation, and A_{bkg} is a normalization constant.

J/Ψ's p_T bins are the ones which then are included into Eq. 6.6. They are chosen to contain at least 800÷1000 reconstructed candidate e^+e^- pairs. In Fig. 6.7 it is shown the distribution of the reconstructed p_T of the J/Ψ candidates expected at ∼10 pb^{-1}. The distribution starts from 10 GeV/c, due to the 5 GeV thresholds in E_T for the single electron candidate applied in the e-ID. The binning here is 0.1 GeV/c. The chosen bins following the above criteria are reported in Tab. 6.6.

The fit was done on invariant mass histograms[6] obtained for each $p_T^{J/\Psi}$ bin using the

Bins (GeV/c)	Δp_T (GeV/c)	**# included J/Ψ's**
9.5÷11.5	2	1204
11.5÷12	0.5	982
12÷12.6	0.6	824
12.6÷13.2	0.6	906
13.2÷13.7	0.5	903
13.7÷14.2	0.5	845
14.2÷14.7	0.5	834
14.7÷15.3	0.6	884
15.3÷16.1	0.8	963
16.1÷17	0.9	850
17÷18	1	829
18÷19.5	1.5	896
19.5÷21.5	2	922
21.5÷25.5	4	921
25.5÷90	64.5	1064

Table 6.6: *The chosen $p_T^{J/\Psi}$ bins, the number of reconstructed J/Ψ's expected at 10 pb^{-1}, and the following sizes Δp_T which will have to be used in the Eq. 6.6.*

binned maximum likelihood method (see Appendix B): the method is implemented in the `TMinuit` package of the framework `ROOT`[71], which has been used for all the analysis computations presented in this chapter. The fitted function is the sum

$$\mathcal{B}(m_{ee}|A_{bkg}, K_{bkg}) + \mathcal{G}(m_{ee}|A_{J/\Psi}, m_{J/\Psi}, \sigma_{J/\Psi}) \tag{6.10}$$

and the chosen range for the fit is enlarged from 3.1 GeV/c^2 ± 200 MeV/c^2 to 2.5÷3.7 GeV/c^2 to better estimate the linear background shape. The number of reconstructed J/Ψ's ($N^{reco}_{J/\Psi}$) is then obtained with the fitted parameters as

$$N^{reco}_{J/\Psi} = \int_{2.5}^{3.7} \mathcal{G}(m_{ee}|A_{J/\Psi}, m_{J/\Psi}, \sigma_{J/\Psi}) dm_{ee} \tag{6.11}$$

In Tab. 6.7 there are reported the $N^{reco}_{J/\Psi}$ for each $p_T^{J/\Psi}$ bin obtained from the fits expected at 10 pb^{-1}. In Fig. 6.8 there are shown some examples of fits for some $p_T^{J/\Psi}$ bin and a total fit on the invariant mass distribution shown in Fig. 6.5.

Errors are obtained from statistical errors of fitted parameters, looking how integral values change as parameters are changed. The fitting algorithm gives information about the whole covariance matrix of fitted parameters, and to obtain the statistical error taking into account the various correlations the errors should have to be analytically propagated in the integral with the covariance matrix. To avoid heavy computations (for 5 parameters the covariance matrix has 15 independent elements), it has used the following method: a new fit is made fixing one parameter to the obtained value added to the estimated error and leaving free the others. The difference between the integral value obtained with this new fit and the integral of the initial fit gives the propagation of the error of the fixed parameter in the integral, taking into account the correlation of the other parameters. This procedure is then repeated for all the parameters. The square addition of those errors gives the statistical error of the integral due to the errors of the fitted parameters.

With the invariant mass distribution fit the number of reconstructed J/Ψ's per $p_T^{J/\Psi}$ bin is obtained. Anyway, this number includes both prompt and non-prompt J/Ψ's. Off-line and e-ID selections are not enough able to discriminate those events with a very similar topology.

To obtain the number of reconstructed *prompt* J/Ψ's ($N^{reco}_{promptJ/\Psi}$) per $p_T^{J/\Psi}$ bin another fit based on the decay length of B mesons (the only difference in the topology of prompt and non-prompt events) is used. The same procedure used by the CDF collaboration at Tevatron[98] is followed. It gives, for each $p_T^{J/\Psi}$ bin, the fraction f_B of reconstructed

[6]There is a conspicuous number of events per bin around the J/Ψ mass peak to use the *binned maximum likelihood* fitting method.

Bins (GeV/c)	$N^{reco}_{J/\Psi}$	$N^{reco}_{J/\Psi}$ **from MC signal**
9.5÷11.5	1490±62	1453
11.5÷12	862±34	876
12÷12.6	1002±41	978
12.6÷13.2	1033± 47	1028
13.2÷13.7	841±34	853
13.7÷14.2	796±33	801
14.2÷14.7	704±28	698
14.7÷15.3	840±30	845
15.3÷16.1	867±35	864
16.1÷17	844±36	838
17÷18	798±30	788
18÷19.5	892±36	887
19.5÷21.5	875±36	869
21.5÷25.5	929±37	937
25.5÷90	1003±38	998
Total fit	13775±276	13713

Table 6.7: *$N^{reco}_{J/\Psi}$ per $p_T^{J/\Psi}$ bin obtained by a fit on invariant mass distribution. Errors are obtained by statistical errors on fitted parameters. The number from the total fit (13775) is different from what is expected from Tab. 6.5 (8580+3620=12200), which is referred only to events with an invariant mass in the range 3.1 GeV/c^2 ± 200 MeV/c^2.*

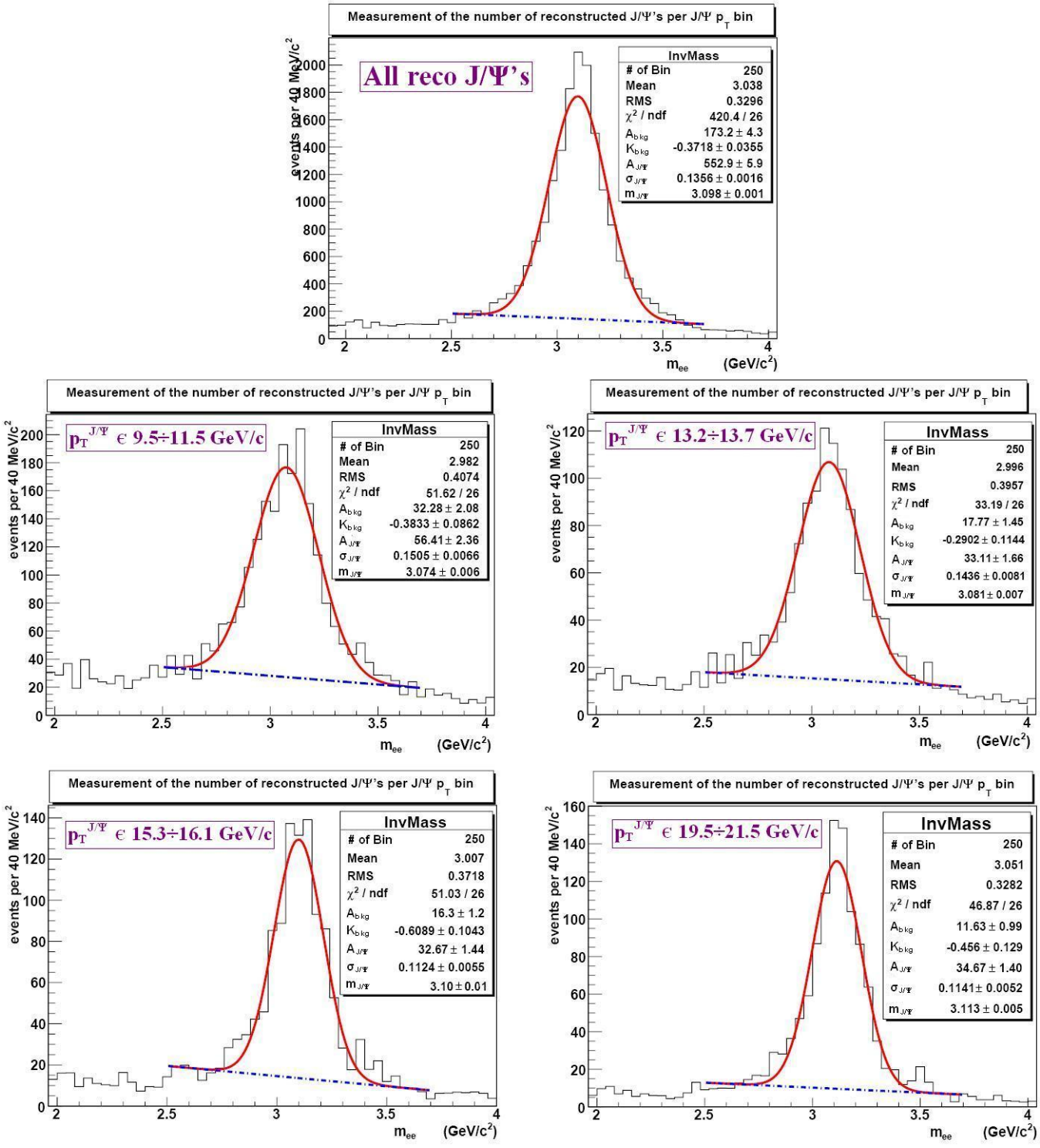

Figure 6.8: *Fits of invariant mass distribution of e^+e^- candidate pairs for various $p_T^{J/\Psi}$ bins (from the top, the whole reconstructed sample, 9.5÷11.5 GeV/c, 13.2÷13.7 GeV/c, 15.3÷16.1 GeV/c, 19.5÷21.5 GeV/c). The red line is the total fitted function (Eq. 6.10); the blue line is the background contribution (Eq. 6.9).*

non-prompt J/Ψ's in the whole collection of reconstructed J/Ψ's obtained with the fit on the invariant mass distribution. In this way, the number $N^{reco}_{promptJ/\Psi}$ can be obtained as $N^{reco}_{promptJ/\Psi} = (1 - f_B) \cdot N^{reco}_{J/\Psi}$.

6.4.2 $N^{reco}_{promptJ/\Psi}$ measurement (the pseudo-proper decay length)

Prompt charmonia states decay immediately after being produced[7]. In contrast, B-hadrons have long lifetimes in the order of ps, because the decay of a B-hadron to a J/Ψ is mediated by the weak interaction b→c+W. This implies that J/Ψ events from the decays of B-hadrons are likely to be displaced from the primary vertex. Special relativity gives the relation between the time decay of the B-hadron in the CMS frame (t) and the time decay (*proper time*) in the B-hadron rest frame (t'):

$$t' = \frac{t}{\gamma} \tag{6.12}$$

where γ is the relativistic coefficient $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ and $\beta = v/c$, with v relative velocity of the two frames. For high γ values, the mean life time of the B-hadron measured in the CMS frame results to be dilatated than the rest frame and the measured decay length (l) results to be dilatated too. The relation between the measured length l and the proper decay time t' can be obtained by a Lorentz boost as

$$t' = \frac{t}{\gamma} = \frac{l}{\gamma\beta c} = l\frac{m}{p} \tag{6.13}$$

where the relativistic momentum relation $p = m\gamma\beta c$ is used.

In the rest frame the decay time is distributed following an exponential decay law:

$$N(t') = N_0 \cdot \mathrm{e}^{-\frac{t'}{\tau}} \tag{6.14}$$

where τ is the mean life time. From the distribution of the proper decay length the mean life time τ and the fraction f_B of J/Ψ's originating from B decays can be measured. This kind of fit should be good to separate the prompt and non-prompt components. In order to do this, an excellent performance of secondary vertex reconstruction is required (σ ∼100 μm). From an experimental point of view, this performance is achievable with the CMS silicon pixels especially in the transverse plane. It is possible to improve the measurement of t' by the projection of l in the transverse plane, where the Eq. 6.13 becomes

$$t' = l\frac{m^B}{p^B} \equiv l\sin\theta\frac{m^B}{p^B\sin\theta} = l_{xy}\frac{m^B}{p^B_T} \Longrightarrow l'_{xy} = l^B_{xy}\frac{m^B c}{p^B_T} \tag{6.15}$$

[7] $c\bar{c}$ states decay in lepton/anti-lepton pairs via electromagnetic interaction (τ ∼10^{-20} s).

where θ is the polar angle referred to z axis. In the Eq. 6.15 $l'_{xy} = ct'$ is defined as **proper decay length**, m^B is the B-hadron mass and p_T^B is its transverse momentum. The distance l_{xy} is the measured distance between the reconstructed vertex of the e^+e^- pair from J/Ψ decays and the primary vertex of the event *along the B-hadron transverse momentum direction*, i.e. it is $l_{xy}^B = \vec{L} \cdot \vec{p}_T^B / |p_T^B|$, where $\vec{L}$ is the three-vector distance between primary and secondary vertices.

Pseudo-proper Decay Length Distribution

To discriminate prompt and non-prompt components, it is introduced the **pseudo-proper decay length** l''_{xy} defined as

$$l''_{xy} = l_{xy}^{J/\Psi} \frac{m^{J/\Psi} c}{p_T^{J/\Psi}} \tag{6.16}$$

where $m^{J/\Psi}$ is the J/Ψ mass, $p_T^{J/\Psi}$ is the J/Ψ transverse momentum and $l_{xy}^{J/\Psi}$ is the measured distance between the reconstructed vertex of the e^+e^- pair from J/Ψ decays and the primary vertex of the event *along the J/Ψ transverse momentum direction*, i.e. $l_{xy}^{J/\Psi} = \vec{L} \cdot \vec{p}_T^{J/\Psi} / |p_T^{J/\Psi}|$. In the J/Ψ case vertex and momentum are the only available information we have, both in prompt J/Ψ candidate events (where no B decay is present) and in B→J/ΨX events (where most of the times it is not possible to fully reconstruct the B meson decay). It is preferred to project the "apparent" B decay length along the J/Ψ momentum direction, since this is a more stable variable, specially in prompt J/Ψ events, where the displacement of the J/Ψ vertex (and its direction) is dominated by the vertex determination resolution. In the Eq. 6.16 l''_{xy} does not coincide with the transverse-proper decay length of the B-hadron in the rest frame (l'_{xy}). In Fig. 6.9 there is shown a scheme of the proper decay length and pseudo-proper decay length in non-prompt J/Ψ events topology.

It is interesting to see how the measurement of the mean life time of B-hadrons change if proper decay length or pseudo-proper decay length are used. In Fig. 6.10 there are shown the distributions of proper decay length and pseudo-proper decay length obtained from Monte Carlo informations at generator level in the non-prompt J/Ψ events sample. Decay length has expected to have an exponential decay law which is derived from Eq. 6.14:

$$N(t') = N_0 \cdot e^{-\frac{t'}{\tau}} \Longrightarrow N(l') = N_0 \cdot e^{-\frac{l'}{c\tau}} \tag{6.17}$$

This function is fitted to the distributions in Fig. 6.10 to obtain the mean life time τ of B-hadrons. While using the proper decay length the correct B life time (1.52 ps) is obtained as expected, by using the pseudo-proper decay length the τ is under-estimated

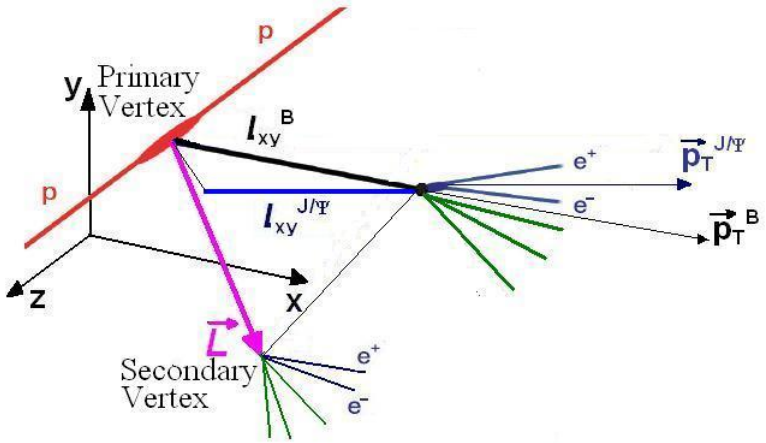

Figure 6.9: *A scheme of the proper decay length (l'_{xy}) and pseudo-proper decay length (l''_{xy}) in non-prompt J/Ψ events topology.*

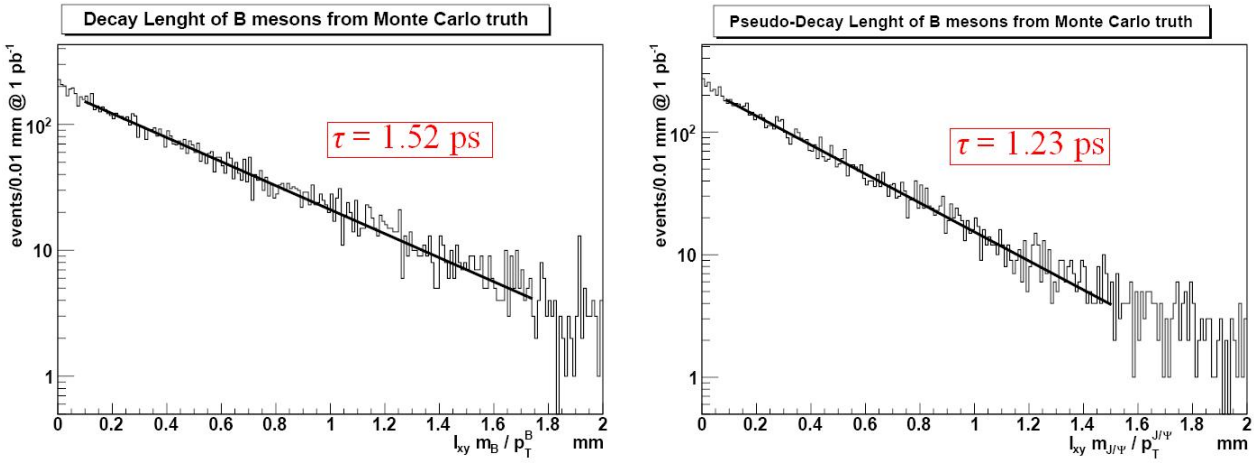

Figure 6.10: *Fit on proper decay length distribution (left) and pseudo-proper decay length distribution (right) obtained from MC truth in the non-prompt J/Ψ events Monte Carlo sample (1 pb^{-1}).*

by a factor 0.8, which have an expected value of ∼1.5 ps.

The pseudo-proper decay length is reconstructed to obtain the fraction of non-prompt J/Ψ's in the whole amount of reconstructed J/Ψ in the invariant mass peak. First of all the J/Ψ decay vertex is reconstructed from tracks selected by the e-ID algorithm and associated to the electron/positron. Both KF tracks and GSF tracks are used for this purpose, and differences in the reconstruction performances will be show in the following. In the `CMSSW` code there is implemented a package (`KinematicVertexFitter`[99]) which fits the vertex of two tracks extrapolating the helix trajectories. The goal of a *kinematic fit* is to improve the resolution of experimental measurements, and test hypothesis by introducing constraints derived from physics laws (such as invariant mass of the two candidate electron/positron tracks) into the minimization problem. It is based on a Least Means Squared minimization with Lagrange multipliers. GSF tracks and KF tracks helix extrapolation is made by the `TransientTrackBuilder`. The *kinematic fit* retrieves the reconstructed vertex position, its corresponding covariance matrix, the χ^2 and the number of degrees of freedom assigned to the vertex during the last refit. The state of the fitted vertex is described by 7 parameters: the (x, y, z) position in the global frame, the estimated (p_x, p_y, p_z) of the decayed particle at the vertex and the mass of the decayed particle, calculated as a result of previous fits or hypothesis assigned to the final state fitted tracks. Then, the method `VertexDistance3D`[100] of the `KinematicVertexFitter` class computes the distance of this secondary vertex from the primary vertex (the three-vector $\vec{L}$) with the associated error. The $l_{xy}^{J/\Psi}$ is computed using the $p_T^{J/\Psi}$ obtained by the `KinematicVertexFitter` during the fit (i.e. $\sqrt{p_x^2 + p_y^2}$).

The $p_T^{J/\Psi}$ which has to be used in the Eq. 6.16 is instead obtained adding the reconstructed E_T of single electron/positron candidates, such as in Eq. 6.5.

Finally, in Eq. 6.16 is used the invariant mass m_{ee} of the e^+e^- candidate pairs instead of the well-known value of J/Ψ mass (∼3.097 GeV/c²). This is done to remove possible experimental systematics effects from the ratio.

The different shapes of l''_{xy} distribution of prompt and non-prompt J/Ψ's allows a good discrimination between the two samples. At Monte Carlo truth level the pseudo-proper decay length for the prompt component is equal to zero ($\tau \leq 10^{-20}$ s); for reconstructed J/Ψ's, the distribution is described by a resolution function, which takes into account the intrinsic resolution due to the finite performance of the tracking reconstruction. The resolution function ($\mathfrak{R}(l''_{xy})$) depends strongly by the misalignment conditions. In Ref.[101], where the most recent measurements of differential cross section of charmonia production are presented, the resolution function is the convolution[8] of the sum of two gaussian with

[8]Given $f(x)$ and $g(x)$ two real functions, the **convolution** of the two functions is defined as $\mathcal{C}(x) \equiv$

a double-sided exponential function in order to describe the little tails of the prompt l''_{xy}. Anyway those measures refer to data token after many years of work of the Tevatron and CDF, and detectors are well calibrated and aligned. In CMS, the work presented in this thesis is refereed to the first 1÷10 pb^{-1}, where detectors need to be well-calibrated and aligned. Principally tracker alignment affects the resolution shapes, which are as wider as alignment is worse. Anyway, CMS detector have already made a long work of alignment with cosmic data[102], and the misalignment conditions are strongly better than the conditions which were expected for the LHC start-up. For this reason, in the Monte Carlo simulation chain, *the start-up conditions are used for the ECAL calibration* (which needs necessarily a certain amount of data during the real data taking at the start-up) and *the ideal conditions*[9] *are used for the tracker alignment.*

For prompt J/Ψ component of the pseudo-proper decay length, in this study I used simply a gaussian as resolution function, i.e.

$$\mathfrak{R}(l''_{xy}|\mu_{res},\sigma_{res}) \equiv \mathcal{G}(l''_{xy}|\mu_{res},\sigma_{res}) \tag{6.18}$$

where the function $\mathcal{G}$ is defined as in 6.8.

The pseudo-proper decay length for non-prompt J/Ψ's has a more complicated shape. The l''_{xy} distribution of B-hadrons is expected to have an exponential behaviour. However, it was noted that the pseudo-proper decay length defined in Eq. 6.16 is different from the proper decay length defined in Eq. 6.15, so the distribution of l''_{xy} for the non-prompt component is not perfectly exponential, since the transverse momentum $p_T^{J/\Psi}$ does not coincide with the transverse momentum p_T^B. To estimate the shape at Monte Carlo truth level, we notice that the relation between the pseudo-proper and the proper decay length (in the assumption that $l_{xy}^{J/\Psi} \sim l_{xy}^B$) is given by

$$l''_{xy} \simeq \frac{m^{J/\Psi}/p_T^{J/\Psi}}{m^B/p_T^B} \cdot l'_{xy} \tag{6.19}$$

i.e. the pseudo-proper decay length depends from two factors: the proper decay length l'_{xy}, which has an exponential distribution ($\mathcal{E}xp(-l'_{xy}/c\tau)$), and the ratio $\frac{m^{J/\Psi}/p_T^{J/\Psi}}{m^B/p_T^B}$ distributed as it is shown in Fig. 6.11. It is possible to see that the ratio has approximatively a gaussian like distribution $\mathcal{G}(l|\mu_{MC},\sigma_{MC})$[10]. So the l''_{xy} in non-prompt J/Ψ events is at Monte Carlo level distributed, being the result of two contribution, as the convolution between the two distributions from a statistical point of view[103], i.e.

$$\mathfrak{D}(l''_{xy}|\mu_{MC},\sigma_{MC},\tau) = \mathcal{G}(l''_{xy}|\mu_{MC},\sigma_{MC}) \otimes \mathcal{E}xp(-l''_{xy}/c\tau) \tag{6.20}$$

$f \otimes g = \int_{-\infty}^{+\infty} f(x)g(x-x')dx'$.

[9]The *ideal* means condition after a good alignment, generally after 100pb^{-1}÷1 fb^{-1}. The conditions used in this simulation are refereed to 100 pb^{-1}.

[10]The fit of the distribution peak with a gaussian gives a normalized $\chi^2 \simeq 1$

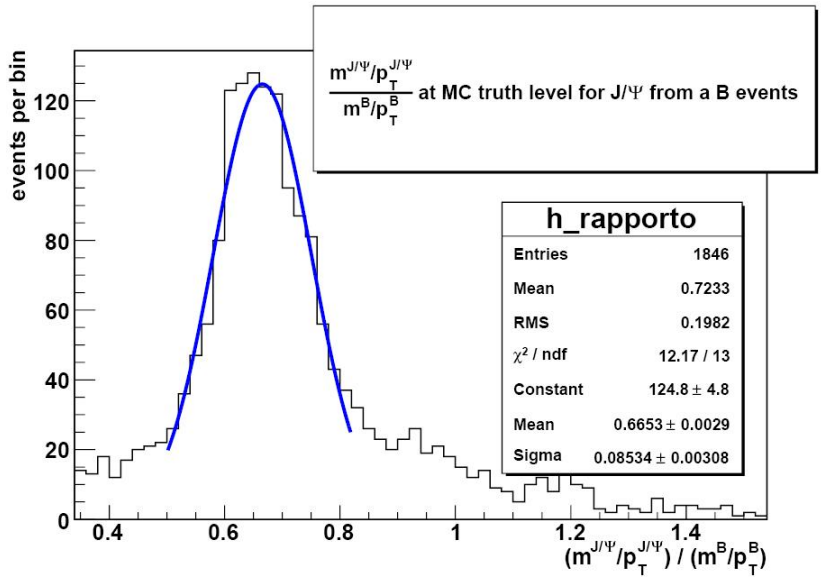

Figure 6.11: *The ratio* $\frac{m^{J/\Psi}/p_T^{J/\Psi}}{m^B/p_T^B}$ *distribution obtained from MC truth in the non-prompt J/Ψ events Monte Carlo sample where the two* e^+e^- *have* $p_T \geq 5$ *GeV/c (*$\sim$*1* pb^{-1}*).*

To describe the l''_{xy} distribution for non-prompt J/Ψ events at reconstruction level, Eq. 6.20 must be convoluted with the same resolution function as that for prompt J/Ψ's (Eq. 6.18):

$$\mathfrak{R}(l''_{xy}|\mu_{res},\sigma_{res})\otimes\mathfrak{D}(l''_{xy}|\mu_{MC},\sigma_{MC},\tau) \equiv \mathcal{G}(l''_{xy}|\mu_{res},\sigma_{res})\otimes\big(\mathcal{G}(l''_{xy}|\mu_{MC},\sigma_{MC})\otimes\mathcal{E}xp(-l''_{xy}/c\tau)\big) \tag{6.21}$$

Remembering the associative property of the convolution[11] and that the convolution of two gaussians is a gaussian again[12][104], Eq. 6.21 can be written as

$$\mathcal{G}(l''_{xy}|\mu_{conv},\sigma_{conv})\otimes\mathcal{E}xp(-l''_{xy}/c\tau) \tag{6.22}$$

For this reason, the fit is not able to distinguish the single parameters μ_{MC}, σ_{MC}, μ_{res} and σ_{res}, but only the "new" parameters μ_{conv}, σ_{conv} introduced in Eq. 6.22. However, this does not represent a problem for our purposes.

The Likelihood Function in the Fit to Obtain the Non-prompt Fraction

The fit on the invariant mass of the candidate e^+e^- pairs results the inclusive number of reconstructed J/Ψ's ($N^{reco}_{J/\Psi}$). In order to measure the differential cross section of prompt J/Ψ's production basing on Eq. 6.6, only the prompt component ($N^{reco}_{promptJ/\Psi}$) must be

[11] $f(x)\otimes(g\otimes h)(x) \equiv (f\otimes g)(x)\otimes h(x)$, being $f(x), g(x)$ and $h(x)$ three real functions.
[12] $\mathcal{G}(x|\mu_1,\sigma_1)\otimes\mathcal{G}(x|\mu_2,\sigma_2) = \mathcal{G}(x|\mu_{conv},\sigma_{conv})$, with $\mu_{conv} = \mu_1+\mu_2$ and $\sigma_{conv} = \sqrt{\sigma_1^2+\sigma_2^2}$.

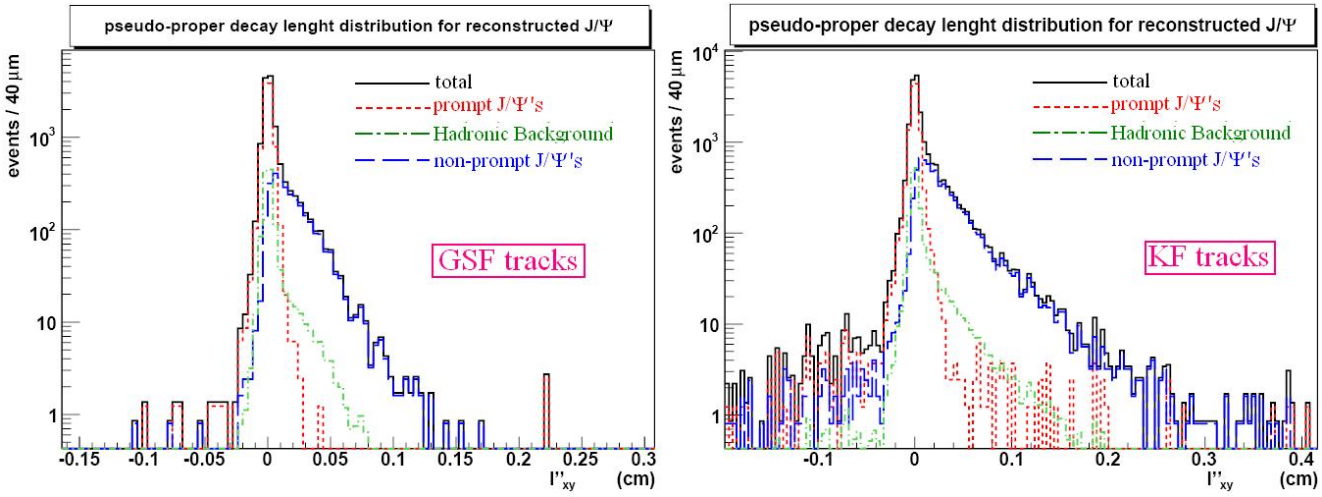

Figure 6.12: *l''_{xy} distribution of reconstructed J/Ψ after 10 pb^{-1}, both for candidates identified using KF tracks and GSF tracks. The various contributions (prompt events, non-prompts events, QCD background events) are shown.*

counted. To obtain the fraction of non-prompt component (f_B), a fit on the distribution of the reconstructed pseudo-proper decay length (as explained above) is performed:

$$l''_{xy} = l^{J/\Psi}_{xy} \cdot \frac{m_{ee}}{p_T^{J/\Psi}} \tag{6.23}$$

In Fig. 6.12 are shown the distributions of the reconstructed l''_{xy} expected after 10 pb^{-1}, for candidate electrons/positrons identified with e-ID based both on KF tracks and GSF tracks (which are used to measure $l^{J/\Psi}_{xy}$). It should be noticed that negative values of the l''_{xy} come from resolution effects, where the vectors of the J/Ψ momentum and the one joining the primary and secondary vertices in the transverse plane (see Fig. 6.9) are in opposite directions.

The first observation is that the pseudo-proper decay length is under-estimated in the case of GSF tracks: the exponential part of the distribution falls down quickly (the tail is limited to ∼1 mm, and a drop of the distribution is already visible at ∼0.5 mm) while in the case of KF tracks the tail exceeds the value of 2 mm. This is due to the fact that the reconstruction of GSF tracks is affected by a loss of efficiency for high values of the distance $\vec{L}$ between the primary and secondary vertex. KF tracks are less affected by this effect, and the reason stands in the different *seeding* of track reconstruction. In the usual Kalman Filter, track *seed* is build from all combinatorial of hits triplets in the pixels detector. In the Gaussian Sum Filter, it is required a match of at least three aligned hits within a window in ϕ defined from the position of a ECAL SC. Displaced track vertices could have *seed* hits in the pixels detector out from the ϕ window, so the pattern recognition for the track trajectory reconstruction with the Gaussian Sum Filter

algorithm cannot be started.

This brings to an important conclusion about this physics study, in a decay channel with electrons in the final states: the usage of candidates obtained by the e-ID based on GSF tracks, such as the usage of standard `GsfElectrons` (which are based on GSF tracks) too, are not properly indicated to have a good discrimination of prompt and non-prompt events based on a fit on the pseudo-proper decay length distribution.

Considering only the case of candidates reconstructed with the e-ID based on KF tracks, the fit is performed as follows: As for the invariant mass fit, reconstructed J/Ψ →e⁺e⁻ events are divided in the binning in $p_T^{J/\Psi}$ as exposed in Tab. 6.6. The fit is then applied for reconstructed events included in each bin; unlike in the invariant mass fit, the pseudo-proper decay distribution does not have many events for each bin in l''_{xy} ($\geq$50$\div$100), specially in the exponential tail for high values of l''_{xy}, so it has been chosen to use as fit algorithm the **unbinned maximum likelihood fit** (see Appendix B), which allows good performance also in the case where statistic for each single bin of a distribution is not enough high. The invariant mass m_{ee} and the pseudo-proper decay length l''_{xy} of each event is simultaneously fitted by a log-likelihood function $\mathfrak{L}$ which is defined as:

$$\ln\mathfrak{L} = \sum_{i=1}^{N} \ln\mathfrak{F}(l''_{xy}, m_{ee}|parameters) \tag{6.24}$$

where the sum is made on the N events reconstructed in each $p_T^{J/\Psi}$ bin. Each likelihood function $\mathfrak{F}$ is the product of two independent distributions (l''_{xy} and m_{ee}), so can be written as

$$\mathfrak{F}(l''_{xy}, m_{ee}) = f_{sig} \cdot \mathfrak{F}_{sig}(l''_{xy}) \cdot \mathfrak{M}_{sig}(m_{ee}) + (1 - f_{sig}) \cdot \mathfrak{F}_{bkg}(l''_{xy}) \cdot \mathfrak{M}_{bkg}(m_{ee}) \tag{6.25}$$

where f_{sig} is the fraction of signal J/Ψ events (i.e. both prompt and non-prompt), $\mathfrak{F}_{sig}(l''_{xy})$ and $\mathfrak{F}_{bkg}(l''_{xy})$ are the functional forms describing the J/Ψ pseudo-proper decay length distribution for the signal and background respectively, and $\mathfrak{M}_{sig}(m_{ee})$ and $\mathfrak{M}_{bkg}(m_{ee})$ are the functional forms describing the invariant mass distributions for the signal and background respectively. These last ones are the same functional forms used in the invariant mass fit to obtain the $N^{reco}_{J/\Psi}$, i.e. $\mathfrak{M}_{sig}(m_{ee})$ coincides with Eq. 6.8 (gaussian peak) and $\mathfrak{M}_{bkg}(m_{ee})$ coincides with Eq. 6.9 (background linear approximation in the considered invariant mass range).

The signal pseudo-proper decay length distribution is expressed as:

$$\mathfrak{F}_{sig}(l''_{xy}|f_B, \mu_{res}, \sigma_{res}, \mu_{conv}, \sigma_{conv}, \tau) = f_B \cdot \mathfrak{F}_B(l''_{xy}) + (1 - f_B)\mathfrak{F}_{prompt}(l''_{xy}) \tag{6.26}$$

where f_B is the fraction of J/Ψ's from B-hadron decays, and $\mathfrak{F}_B(l''_{xy})$ and $\mathfrak{F}_{prompt}(l''_{xy})$ are the pseudo-proper decay length distributions for non-prompt and prompt J/Ψ's respectively, i.e. $\mathfrak{F}_{prompt}(l''_{xy})$ coincides with Eq. 6.18 (resolution of the primary vertex), and

$\mathfrak{F}_B(l''_{xy})$ coincides with Eq. 6.22 (resolution of the primary vertex convoluted with the exponential B decay).

With regard to the background going in Eq. 6.25, this has a bit complicated function. Part of the background is from the long lived B-hadrons that produce electrons with opposite sign such as B→ ce$^-\bar{\nu}X$ and c→ e$^+\nu X$, which brings to the exponential tail visible in Fig. 6.12. The other part are instead hadrons identified as electrons fake which are produced in hadronic interaction in the primary vertex, and has the form due to the finite resolution of the vertices detector. So the expression for $\mathfrak{F}_{bkg}(l''_{xy})$ is the sum of those two contributions:

$$\begin{aligned}\mathfrak{F}_{bkg}(l''_{xy}|f',\mu_{res},\sigma_{res},\mu'_{conv},\sigma'_{conv},\tau') = (1-f')\cdot\mathfrak{R}(l''_{xy}|\mu_{res},\sigma_{res})+ \\ + f'\cdot\mathcal{G}(l''_{xy}|\mu'_{conv},\sigma'_{conv})\otimes\mathcal{E}xp(-l''_{xy}/c\tau')\end{aligned} \tag{6.27}$$

where the resolution function $\mathfrak{R}(l''_{xy}|\mu_{res},\sigma_{res})$ is the same used for the prompt signal (Eq. 6.18), and the function part deriving from semi-leptonic decays of B hadrons has the form of Eq. 6.22 used for the non-prompt signal, but where parameters $\mu'_{conv},\sigma'_{conv},\tau'$ are not necessarily the same of Eq. 6.22.

The prompt, non-prompt J/Ψ and background Monte Carlo samples are added together according to the expected cross sections and statistics. The unbinned maximum likelihood was performed in all the 15 $p_T^{J/\Psi}$ bins according to Tab. 6.6, and for each bin the fraction of non-prompt signal events which are reconstructed in the J/Ψ invariant mass peak (f_B) are retrieved. In this way, the number of prompt events which has to be included in Eq. 6.6 is given by

$$N^{reco}_{promptJ/\Psi} = (1-f_B)\cdot N^{reco}_{J/\Psi} \tag{6.28}$$

In Fig. 6.13 are shown the fits of the pseudo-proper decay length distributions for the various $p_T^{J/\Psi}$ bins. The fitted function is the one in Eq. 6.25, $\mathfrak{F}(l''_{xy},m_{ee})$, which is a bi-dimensional function. So the fit has to be implemented in parallel both for the pseudo-proper decay distribution (l''_{xy}) and the invariant mass of the candidate pairs (m_{ee}). Anyway, the unbinned maximum likelihood method explained in Appendix B can be generalized also in multidimensional cases[103]. There is a package in the `ROOT` framework, `RooFit`[105], which allows multidimensional fits and also convolutions fits; convolutions are calculated both analytically (when possible) and with numerical algorithms. In this work, the `RooFit` package has been used for the pseudo-proper decay length fit; the analytical expression of the convolution between the gaussian and the exponential functions has been used too. **The unbinned maximum likelihood works for distribution normalized to the unitary**. The `RooFit` package normalizes to the unitary the distributions using numerical approximation[105], retrieves the fitted parameters, and then

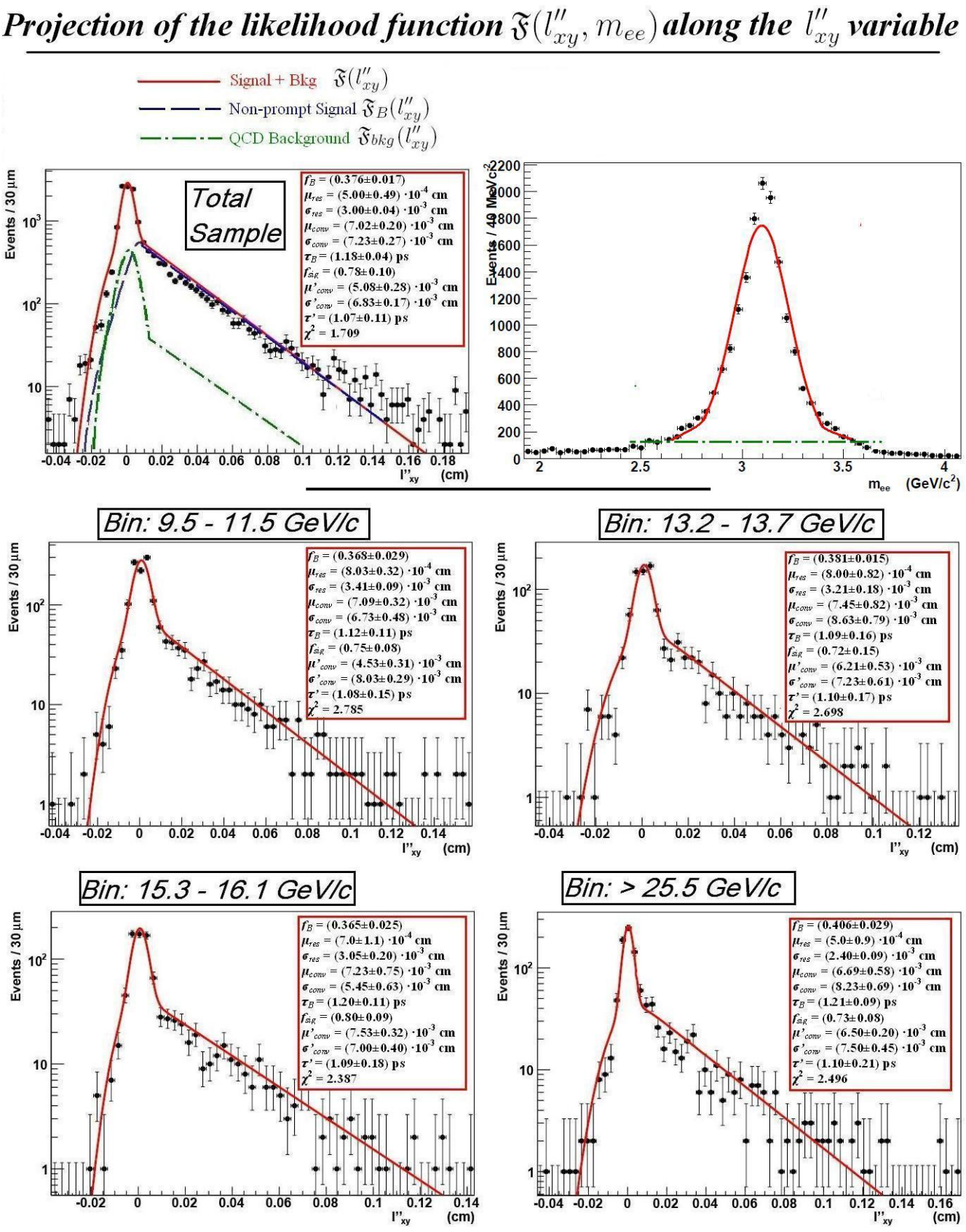

Figure 6.13: *Fits of pseudo-decay length distribution of reconstructed J/Ψ's for various $p_T^{J/\Psi}$ bins (from the top, the whole reconstructed sample, 9.5÷11.5 GeV/c, 13.2÷13.7 GeV/c, 15.3÷16.1 GeV/c, >25.5 GeV/c). The red line is the total fitted function (Eq. 6.25); the blue line is the non-prompt component contribution ($\mathfrak{F}_B(l''_{xy}) \cdot \mathfrak{M}_{sig}(m_{ee})$); the green line is the background contribution ($\mathfrak{F}_{bkg}(l''_{xy}) \cdot \mathfrak{M}_{bkg}(m_{ee})$).*

parameters	fitted values
f_B	(0.376±0.017)
μ_{res}	(5.00±0.49)·10^{-4} cm
σ_{res}	(3.00±0.04)·10^{-3} cm
μ_{conv}	(7.02±0.20)·10^{-3} cm
σ_{conv}	(7.23±0.27)·10^{-3} cm
τ_B	(1.18±0.04) ps
f_{sig}	(0.78±0.10)
μ'_{conv}	(5.08±0.28)·10^{-3} cm
σ'_{conv}	(6.83±0.17)·10^{-3} cm
τ'	(1.07±0.11) ps
μ_{mass}	(3.12±0.12) GeV/c^2
σ_{mass}	(0.14±0.06) GeV/c^2
K_{bkg}	(-0.07±0.03)
χ^2/ndf	1.709

Table 6.8: *Parameters of function in Eq. 6.25 fitted on all selected e^+e^- candidate pairs.*

superimposes the fitted functions to the histograms. Since distribution are normalized to the unitary, parameters concerning the normalization of the fitted functions are not retrived.

In Tab. 6.8 all the parameters fitted with the unbinned maximum likelihood on all collected e$^+$e$^-$ pairs is reported. In particular, the number of total collected J/Ψ candidates has been obtained with the fitted f_{sig} parameter: the total number of collected candidates (N_{ee}) with an invariant mass between 0 and 10 GeV/c^2 is multiplied by the parameter f_{sig}.

$$N_{ee} = 21543, \quad f_{sig} = 0.78 \pm 0.10 \Longrightarrow N^{reco}_{J/\Psi} = N_{ee} \cdot f_{sig} = 16804 \pm 2154 \qquad (6.29)$$

which has to be compared with the value estimated with the binned maximum likelihood fit reported in Tab. 6.7, 13775±276. In Tab. 6.7 the expected number of reconstructed J/Ψ's at Monte Carlo truth level was also reported (13713). The binned maximum likelihood seems to give a better estimate of $N^{reco}_{J/\Psi}$. Probably, this is due to the low number of background events which are fitted, so the unbinned maximum likelihood fit (which normalizes the distributions to the unitary) retrieves the parameters which describe the background with high uncertainties.

In Fig. 6.14 there is shown the fraction of non-prompt reconstructed J/Ψ's obtained for each $p_T^{J/\Psi}$. In Tab. 6.9 there are reported the same numbers, with the number of

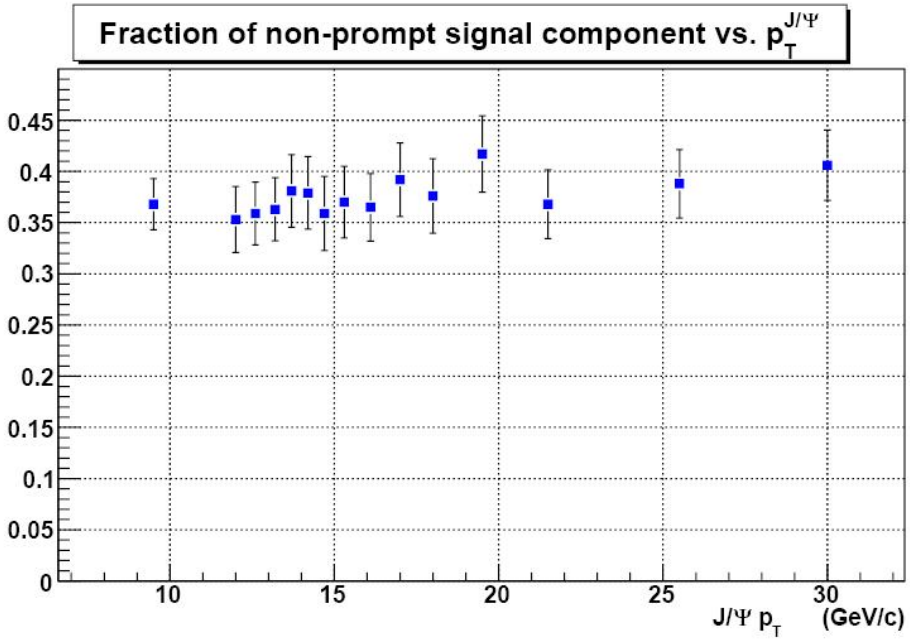

Figure 6.14: *Fraction of non-prompt reconstructed J/Ψ's (f_B) as a function of $p_T^{J/\Psi}$.*

prompt reconstructed J/Ψ's $N^{reco}_{promptJ/\Psi}$ for each bin, obtained with Eq. 6.28.

6.4.3 $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ determination

In the cross section measurement is very important to correct the number of reconstructed J/Ψ for the various efficiencies to obtain the number of *all J/Ψ produced* in the p-p collisions. This is done in Eq. 6.6 dividing for the quantity $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$, the total J/Ψ selection efficiency, which depends on the geometrical acceptance of the detector ($\mathcal{A}$), the trigger efficiency (ϵ_{trig}) and the off-line selection (ϵ_{reco}).

The total J/Ψ selection efficiency is a function of the *true* (i.e. not the reconstructed values, but retrieved from Monte Carlo truth) p_T and η of the reconstructed J/Ψ's; **it is determined from Monte Carlo** as follows:

$$\mathcal{A}\epsilon_{trig}\epsilon_{reco}(p_T^{J/\Psi}, \eta_{J/\Psi}) = \frac{N^{rec}_{J/\Psi}(p_T^{J/\Psi}, \eta_{J/\Psi})}{N^{gen}_{J/\Psi}(p_T^{J/\Psi}, \eta_{J/\Psi})} \tag{6.30}$$

where $p_T^{J/\Psi}$ and $\eta_{J/\Psi}$ are the true value of transverse momentum and pseudorapidity of the generated J/Ψ, $N^{rec}_{J/\Psi}$ is the number of reconstructed J/Ψ's in the phase space region $p_T^{J/\Psi}$ and $\eta_{J/\Psi}$, and $N^{gen}_{J/\Psi}$ is the total true number of generated J/Ψ's, i.e. *before applying the generator di-electron filter*. It has to be noted that the definition of Eq. 6.30 is different from all the previous definitions of efficiencies used in this works, which was with respect

Bins (GeV/c)	$N^{reco}_{J/\Psi}$	f_B	$N^{reco}_{promptJ/\Psi}$
9.5÷11.5	1490±62	0.368±0.029	942±58
11.5÷12	862±34	0.353±0.032	558±35
12÷12.6	1002±41	0.359±0.029	644±39
12.6÷13.2	1033±47	0.363±0.025	658±40
13.2÷13.7	841±34	0.381±0.015	521±25
13.7÷14.2	796±33	0.379±0.035	494±35
14.2÷14.7	704±28	0.359±0.033	451±29
14.7÷15.3	840±30	0.370±0.030	529±32
15.3÷16.1	867±35	0.365±0.025	551±31
16.1÷17	844±36	0.392±0.027	513±32
17÷18	798±30	0.376±0.023	498±26
18÷19.5	892±36	0.417±0.035	520±38
19.5÷21.5	875±36	0.368±0.029	553±34
21.5÷25.5	929±37	0.388±0.022	568±31
25.5÷90	1003±38	0.406±0.029	596±37
Total fit	13775±276	0.376±0.017	8597±291

Table 6.9: *$N^{reco}_{J/\Psi}$, f_B and $N^{reco}_{promptJ/\Psi}$ for each $p_T^{J/\Psi}$ bin obtained by a fit on invariant mass distribution. Errors are obtained by statistical errors on fitted parameters.*

to the filtered events. Indeed, for the cross section measurement (such as for all physical variables) we are interested in the true number of J/Ψ's in the η region $|\eta|$ <2.5, not only in the ones passing the generator filter. Then, the definition in Eq. 6.30 is referred only on **prompt J/Ψ** events.

The $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ determination, being totally based on Monte Carlo, is potentially subjected to relevant systematics effects.

One systematic is the particular Monte Carlo model used for the sample generation: efficiencies change considering different models for reference.

More in particular, the polarization of the J/Ψ affects the decaying e$^+$e$^-$ pairs spectrums, and therefore influences the kinematic acceptance in Eq. 6.30. The considered Monte Carlo samples of prompt J/Ψ in this work are generated with no-polarization, following the measurements made with Tevatron data. Anyway, NRQCD theory foresees same level of polarization for J/Ψ's at high p_T values, which can be seen at LHC. So the polarization choice has to be included at least as systematic effect in the total selection efficiency measurement. Once the polarization parameter α (see Eq. 1.22 in Chapter 1) will be measured with real LHC data, the total reconstruction efficiency will be computed by a weighted sum of the total efficiencies from transversely, longitudinally and non polarized J/Ψ production

$$(\mathcal{A}\epsilon_{trig}\epsilon_{reco})_\alpha = (1-\eta_L-\eta_T)(\mathcal{A}\epsilon_{trig}\epsilon_{reco})_0 + \eta_L(\mathcal{A}\epsilon_{trig}\epsilon_{reco})_L + \eta_T(\mathcal{A}\epsilon_{trig}\epsilon_{reco})_T \quad (6.31)$$

where η_L and η_T represent respectively the fraction of J/Ψ's polarized longitudinally and transversally.

Another systematic effect is due to the accuracy in the $p_T^{J/\Psi}$ reconstruction. One event belonging (according to the Monte Carlo truth) to one p_T bin might get reconstructed in a different p_T bin.

Anyway, this last systematic effect and the previous ones will be treated more specifically in the next sections, where all systematic effects on the cross section measurements will be presented.

In Tab. 6.10 there are reported the total J/Ψ selection efficiencies for each $p_T^{J/\Psi}$; in Fig. 6.15 it is reported as a function of J/Ψ's p_T and η.

6.4.4 Measurement of the correction factors λ_{trig}^{corr} and λ_{reco}^{corr}

The correction factors λ_{trig}^{corr} and λ_{reco}^{corr} are needed to correct the dependence of the total selection efficiencies retrieved in the previous section from the Monte Carlo simulations. Any deviation to the total efficiency that could arise in real data taking can be determined

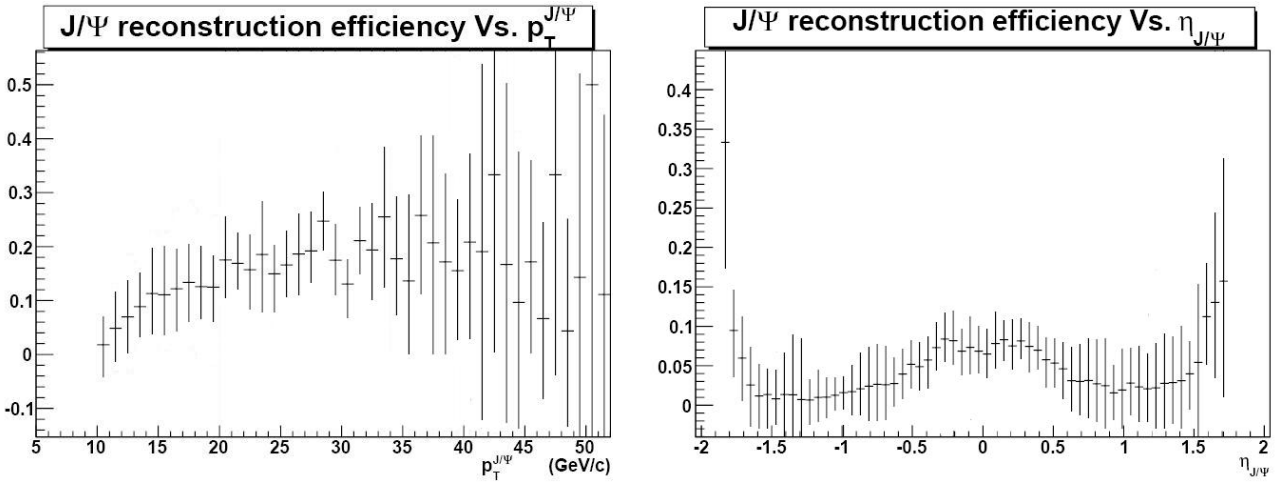

Figure 6.15: *$\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ as a function of J/Ψ's p_T and η.*

Bins (GeV/c)	$\mathcal{A}\epsilon_{trig}\epsilon_{reco}$
9.5÷11.5	0.02607±0.00092
11.5÷12	0.0496±0.0031
12÷12.6	0.0610±0.0035
12.6÷13.2	0.0722±0.0041
13.2÷13.7	0.0770±0.0051
13.7÷14.2	0.0846±0.0058
14.2÷14.7	0.1042±0.0070
14.7÷15.3	0.0882±0.0061
15.3÷16.1	0.1053±0.0068
16.1÷17	0.1091±0.0073
17÷18	0.1105±0.0077
18÷19.5	0.1191±0.0076
19.5÷21.5	0.1336±0.0091
21.5÷25.5	0.1403±0.0091
25.5÷90	0.167±0.011

Table 6.10: *The total J/Ψ selection efficiency $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ per $p_T^{J/\Psi}$ bin. Statistical errors are reported.*

on data itself, using the **tag and probe** methods[106]. As previously mentioned, there will be both corrections to trigger efficiencies as well as to the offline efficiencies.

Known low-mass resonances such as J/Ψ and ϒ are used: the e^+e^- pairs are considered, and it is then assumed that the offline reconstruction efficiency can be expressed as the product of the individual electrons efficiencies:

$$\epsilon^{J/\Psi}_{offline}(p_T^{J/\Psi}, \eta_{J/\Psi}) = \epsilon_{e^+}(p_T^{e^+}, \eta_{e^+}) \cdot \epsilon_{e^-}(p_T^{e^-}, \eta_{e^-}) \tag{6.32}$$

both in the data and in the Monte Carlo simulated samples. A *tag* is a well identified electron/positron. The existing of a tag electron/positron usually indicates the existence of the other leg from the same resonance. A *probe* is an ECAL SC or Tracker track that is likely from the other positron/electron coming from the resonance. The method begins to count when both *tag* and *probe* exists. In this way, efficiencies on single electron/positron can be estimated directly from data. The correction factors to efficiencies are the ratio between real data and Monte Carlo simulation efficiencies:

$$\lambda^{corr}_{reco}(p_T^e, \eta_e) = \frac{\epsilon_e^{data}(p_T^e, \eta_e)}{\epsilon_e^{MC}(p_T^e, \eta_e)} \tag{6.33}$$

Ideally, $\lambda^{corr}_{reco}(p_T^e, \eta_e)$ should be one, if the Monte Carlo simulation used is perfect.

In a similar way, the trigger efficiency correction can be determined. Always the *tag and probe* method is used. The two HLT path, single electron and J/Ψ dedicated path have to be exploited on J/Ψ events. In particular, in events where the single electron path has given positive response the J/Ψ dedicated path is exploited, and vice versa. Events where a J/Ψ is off-line reconstructed are *tagged*; the *probe* are given by events where the hlt path is passed. In this way, trigger efficiencies can be estimated directly from data. The J/Ψ trigger efficiency, as computed with this method, is therefore the measure of the probability of a reconstructed J/Ψ event to fire the J/Ψ trigger:

$$\epsilon_{trig}(p_T^{J/\Psi}, \eta_{J/\Psi}) = \frac{N_{trig}(p_T^{J/\Psi}, \eta_{J/\Psi})}{N_{offlinereco}(p_T^{J/\Psi}, \eta_{J/\Psi})} \tag{6.34}$$

Then, the correction factor is defined as

$$\lambda^{corr}_{trig}(p_T^{J/\Psi}, \eta_{J/\Psi}) = \frac{\epsilon^{data}_{trig}(p_T^{J/\Psi}, \eta_{J/\Psi})}{\epsilon^{MC}_{trig}(p_T^{J/\Psi}, \eta_{J/\Psi})} \tag{6.35}$$

The correction factors can be obtained with real data. In this work, the differential cross section measurement is made with Monte Carlo sample data, so there is no sense to compute correction factors here. Here, the error associated to the usage of Monte Carlo is

estimated with the systematics presented in the previous section. When the measurement will be made with real data, the knowledge of the correction factors will allow a less dependence from the Monte Carlo, and so from the systematics above mentioned.

6.5 Systematics Uncertainties

In this section an estimate of possible systematic uncertainties is given. It must be noticed that some of the systematics will come from the experimental setup (such as the one from the luminosity or alignment), others from the particular selections chosen for the J/Ψ selection, and others from the particular models implemented in the Monte Carlo. Others may arise when real data taking will start, for example, the corrections to the trigger and offline reconstruction efficiencies.

The following systematic uncertainties have been considered:

Luminosity Determination

The Van-der-Meer scan of the LHC beams[107] will be able to measure the luminosity with an uncertainty of the order of 5% or less already at the LHC start-up[108] (even if uncertainty can be of the order of 10% for higher luminosity). As a comparison, in the J/Ψ measurement of CDF II[109], that uncertainty was about 5.9% for an integrated luminosity of 39.7 pb^{-1}.

Parameterization of the Invariant Mass Fitting function

As shown in Section 6.4.1, a Gaussian function is found to describe quite accurately the di-electron invariant mass in each $p_T^{J/\Psi}$ bin. The number of events will be determined from a fit to the di-electron invariant mass distribution, and any dependence of the fit affects this measurement. In order to investigate any dependence on the fit, a new fit range has been chosen to see how the linear approximation of the background and the integral result of the fitted gaussian change for each $p_T^{J/\Psi}$ bin. In Tab. 6.11 there are reported the new numbers compared with the old ones. The systematic error is estimated as the differences.

Parameterization of the Resolution and B-lifetime model

For the description of the resolution model a gaussian has been used. The model has been changed to a double-gaussian to estimate the systematic due to the pseudo-proper decay fitted model. As systematic effect, also the B-lifetime model has been changed:

Bins (GeV/c)	**old** $N^{reco}_{J/\Psi}$	**new** $N^{reco}_{J/\Psi}$	ϵ_{sys}
9.5÷11.5	1490	1520	2.0%
11.5÷12	862	878	2.1%
12÷12.6	1002	998	2.4%
12.6÷13.2	1033	1052	1.9%
13.2÷13.7	841	822	2.3%
13.7÷14.2	796	812	2.0%
14.2÷14.7	704	722	2.5%
14.7÷15.3	840	821	2.3%
15.3÷16.1	867	850	2.0%
16.1÷17	844	863	2.2%
17÷18	798	784	1.8%
18÷19.5	892	910	1.9%
19.5÷21.5	875	906	2.3%
21.5÷25.5	929	907	2.5%
25.5÷90	1003	1025	2.4%

Table 6.11: *Systematic error on $N^{reco}_{J/\Psi}$ for each $p_T^{J/\Psi}$ due to the fit dependence.*

in Section 6.4.2 it has been explained that the introduction of the pseudo-proper decay length instead of the usual proper decay length brings to a B-lifetime model which is a convolution of a gaussian with an exponential function. To estimate systematic effects, it has been considered only the exponential part in the fit and the gaussian function was removed to have a conservative estimate of the model used to describe the B-hadron lifetime distribution. Results on how those systematics affect the B fraction measurement are reported in Tab. 6.15.

Systematics due to the reconstruction and identification selections

Reconstruction and identification selections have been optimized to have a good reconstruction efficiency of J/Ψ's and at the same time a good signal-to-background ratio. The choice of the selection cuts affects the final number of selected signal and background events and so $N^{reco}_{J/\Psi}$ and f_B.

In Tab. 6.14 are reported the estimated values of $N^{reco}_{J/\Psi}$ and f_B when selection cuts are enlarged (see Tab. 6.13) (this brings to an higher number of collected signal events, but also to an higher background). The difference of the values are taken as the systematic error.

Bins (GeV/c)	**old** f_B **values**	**new** f_B **values**	ϵ_{sys}
9.5÷11.5	0.368	0.357	3%
11.5÷12	0.353	0.360	2%
12÷12.6	0.359	0.355	1%
12.6÷13.2	0.363	0.378	4%
13.2÷13.7	0.381	0.370	3%
13.7÷14.2	0.379	0.371	2%
14.2÷14.7	0.359	0.355	1%
14.7÷15.3	0.370	0.377	2%
15.3÷16.1	0.365	0.380	4%
16.1÷17	0.392	0.384	2%
17÷18	0.376	0.387	3%
18÷19.5	0.417	0.433	5%
19.5÷21.5	0.368	0.382	4%
21.5÷25.5	0.388	0.377	3%
25.5÷90	0.406	0.413	2%

Table 6.12: *Systematic error on f_B for each $p_T^{J/\Psi}$ due to the pseudo-proper decay length fit model dependence.*

Off-line Selections	**old values**	**new values**
E_T	5	4
9/25	0.85	0.7
H/E	0.2	0.3
$\lvert\eta_{SC}^{corr} - \eta_{trk}\rvert$	0.009	0.02
$\lvert\phi_{SC}^{corr} - \eta_{trk}\rvert$	0.02	0.05
$\frac{1}{E_T} - \frac{1}{p_T}$	[-0.04,0.02]	[-0.06,0.04]
DCA	100 μm	100 μm
cosδ	0.85	0.70

Table 6.13: *Reconstruction and identification off-line selection cuts enlarged (values for the barrel region).*

Bins (GeV/c)	**old** f_B	**new** f_B	ϵ_{sys}	**old** $N^{reco}_{J/\Psi}$	**new** $N^{reco}_{J/\Psi}$	ϵ_{sys}
9.5÷11.5	0.368	0.374	1.5%	1490	1480	0.7%
11.5÷12	0.353	0.359	1.7%	862	869	0.9%
12÷12.6	0.359	0.355	1.4%	1002	1006	0.4%
12.6÷13.2	0.363	0.358	1.3%	1033	1027	0.6%
13.2÷13.7	0.381	0.387	1.7%	841	845	0.5%
13.7÷14.2	0.379	0.372	1.9%	796	792	0.5%
14.2÷14.7	0.359	0.362	1.2%	704	709	0.6%
14.7÷15.3	0.370	0.366	1.0%	840	845	0.3%
15.3÷16.1	0.365	0.360	1.4%	867	862	0.4%
16.1÷17	0.392	0.397	1.6%	844	841	0.5%
17÷18	0.376	0.371	1.5%	798	792	0.3%
18÷19.5	0.417	0.409	1.9%	892	888	0.3%
19.5÷21.5	0.368	0.363	1.4%	875	869	0.7%
21.5÷25.5	0.388	0.381	1.7%	929	923	0.4%
25.5÷90	0.406	0.401	1.3%	1003	1010	0.5%

Table 6.14: *Systematic errors on f_B and $N^{reco}_{J/\Psi}$ for each $p_T^{J/\Psi}$ due to the reconstruction and identification selections dependence.*

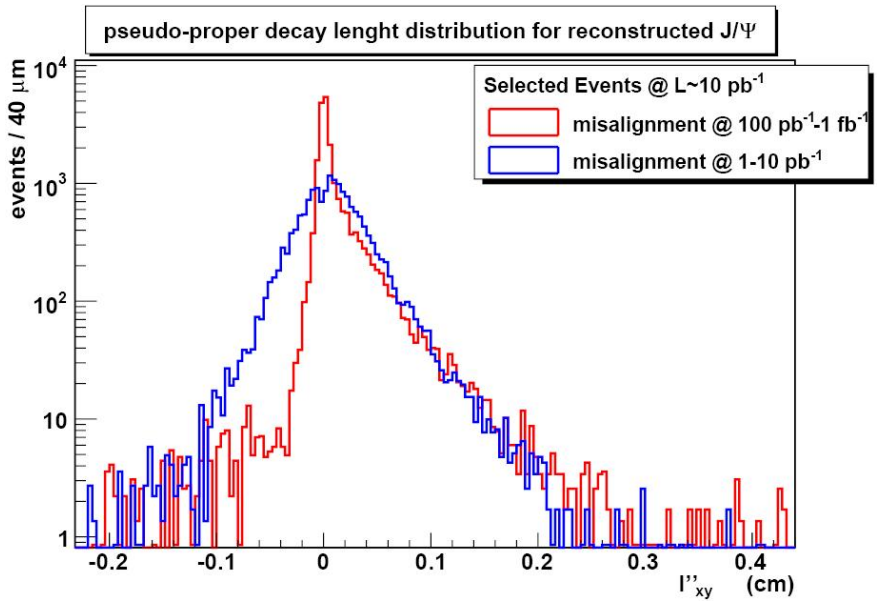

Figure 6.16: *Pseudo-proper decay length distribution of reconstructed J/Ψ candidates (KF tracks are considered) for the different alignment conditions.*

Misalignment effects

Misalignment affects the reconstructed decay length ($\vec{L}$) distributions, and as such could affect the fitted f_B parameter. In order to test any bias in the fitting procedure of the pseudo-proper decay length distribution, the unbinned maximum likelihood fit was repeated in two scenarios (i.e. "ideal" condition -alignment at 100 pb^{-1}÷1 fb^{-1}- and "start-up" condition -misalignment at 1÷10 pb^{-1}). The result of the compared pseudo-proper decay length distributions is shown in Fig. 6.16. In Tab. 6.15 the systematic errors of the f_B due to the alignment are reported.

Uncertainties in the total J/Ψ selection efficiencies due to the Monte Carlo

The total selection efficiencies values ($\mathcal{A}\epsilon_{trig}\epsilon_{reco}$) reported in Tab. 6.10 are obtained referring totally to the used Monte Carlo samples. Anyway, efficiencies obtained in this way are affected by the systematic due to the particular Monte Carlo model chosen as reference. For example, the used signal sample has been generated in the so called "high radiation scenario" during the DGLAP evolution of the charmonium state. In this scenario is expected a big hadronic amount around the generated prompt J/Ψ. If the "low radiation scenario" is considered, less hadronic amount is foreseen and so less particle per events are expected to cross the detector. This show how different models implemented in the reference Monte Carlo samples bring to different situations and so different trigger and reconstruction efficiencies. The correction factors in Eq. 6.6 are introduced properly to

Bins (GeV/c)	**old** f_B **values**	**new** f_B **values**	ϵ_{sys}
9.5÷11.5	0.368	0.365	0.9%
11.5÷12	0.353	0.357	1.1%
12÷12.6	0.359	0.355	0.8%
12.6÷13.2	0.363	0.359	0.6%
13.2÷13.7	0.381	0.384	0.4%
13.7÷14.2	0.379	0.376	0.5%
14.2÷14.7	0.359	0.357	0.3%
14.7÷15.3	0.370	0.368	0.4%
15.3÷16.1	0.365	0.368	0.3%
16.1÷17	0.392	0.389	0.3%
17÷18	0.376	0.380	0.4%
18÷19.5	0.417	0.421	0.5%
19.5÷21.5	0.368	0.364	0.3%
21.5÷25.5	0.388	0.384	0.4%
25.5÷90	0.406	0.413	0.6%

Table 6.15: *Systematic error on f_B for each $p_T^{J/\Psi}$ due to the pseudo-proper decay length fit model dependence.*

Bins (GeV/c)	$\mathcal{A}\epsilon_{trig}\epsilon_{reco}$	**LRS** $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$	**CSM** $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$	ϵ_{sys}
9.5÷11.5	0.02607	0.02689	0.02702	4%
11.5÷12	0.0496	0.05211	0.0479	6%
12÷12.6	0.0610	0.0643	0.0650	6%
12.6÷13.2	0.0722	0.0768	0.0771	6%
13.2÷13.7	0.0770	0.0800	0.0804	3%
13.7÷14.2	0.0846	0.0868	0.0895	5%
14.2÷14.7	0.1042	0.0954	0.0964	5%
14.7÷15.3	0.0882	0.0920	0.0932	5%
15.3÷16.1	0.1053	0.1103	0.1095	4%
16.1÷17	0.1091	0.1127	0.1139	4%
17÷18	0.1105	0.1115	0.1123	1%
18÷19.5	0.1191	0.1255	0.1222	5%
19.5÷21.5	0.1336	0.1359	0.1346	2%
21.5÷25.5	0.1403	0.1487	0.1498	6%
25.5÷90	0.167	0.175	0.173	4%

Table 6.16: *Systematic errors of total J/Ψ selection efficiency $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ due to the Monte Carlo sample of reference.*

decrease the dependence of the equation from the used Monte Carlo. Anyway, being this work based only on Monte Carlo simulations, the particular model considered in the Monte Carlo has been considered as a systematic. To estimate the systematic error referred to the considered Monte Carlo, other different model of signal generation are considered (the "low radiation scenario" -LRS- and the charmonium produced in CSM). New efficiencies are obtained for the two cases, compared to the original values. Differences give the systematic errors due to the Monte Carlo. Obtained results are reported in Tab. 6.16.

Systematic due to events of a $p_T^{J/\Psi}$ bin reconstructed in other $p_T^{J/\Psi}$ bins

The last systematic studied in the differential cross section measurement of prompt J/Ψ's in the J/Ψ →e⁺e⁻ decay channel is due to events in a certain $p_T^{J/\Psi}$ bin reconstructed in the contiguous bins. Indeed, detector resolutions could bring to reconstructed J/Ψ's with a reconstructed p_T different from the Monte Carlo truth p_T. If those J/Ψ's are in the upper or lower limit of each bin, they can be reconstructed in the immediately contiguous upper or lower bin. This brings to a systematic effect which affects the number of reconstructed prompt J/Ψ's $N^{reco}_{promptJ/\Psi}$. To quantify the "contamination" of each $p_T^{J/\Psi}$

bin from the contiguous bins, the following equation is defined:

$$N^{reco}_{promptJ/\Psi \quad i} = \sum_{j=1}^{15} \epsilon_{ij} \cdot N^{MC}_{promptJ/\Psi \quad j} \tag{6.36}$$

Eq. 6.36 gives the number of reconstructed events $N^{reco}_{promptJ/\Psi}$ in the i-th bin: the sum is made on the 15 $p_T^{J/\Psi}$ bins, $N^{MC}_{promptJ/\Psi}$ is the number of events for each bin at Monte Carlo truth level, and ϵ_{ij} are coefficients which take into account the total selection efficiency for each bin and the "mixing" effect of events of each bin into the i-th bin. This allows to define the matrix A with ϵ_{ij} as matrix elements

$$\vec{N}^{reco}_{promptJ/\Psi} = A\vec{N}^{MC}_{promptJ/\Psi} \tag{6.37}$$

where $\vec{N}^{reco}_{promptJ/\Psi}$ is the vector with the number of reconstructed J/Ψ's for each $p_T^{J/\Psi}$ bin as elements, and $\vec{N}^{MC}_{promptJ/\Psi}$ is the vector with the number of events for each bin at Monte Carlo truth level as elements.

The matrix elements ϵ_{ij} can be obtained looking in which $p_T^{J/\Psi}$ bin a certain event with a given $p_T^{J/\Psi}$ bin at Monte Carlo truth level is reconstructed. In this way, the i-th raw is referred to the i-th $p_T^{J/\Psi}$ bin at reconstruction level and the j-th column is referred to the j-th $p_T^{J/\Psi}$ bin at Monte Carlo truth level. Each element counts how many events included in the j-th $p_T^{J/\Psi}$ bin at Monte Carlo truth level are reconstructed in the i-th $p_T^{J/\Psi}$ bin. Each element of the j-th column is divided for the number $N^{MC}_{promptJ/\Psi}$ of events included in the j-th $p_T^{J/\Psi}$ bin at Monte Carlo truth level. In this way the matrix elements ϵ_{ij} are obtained normalized to unitary, to have the physical meaning of an efficiency. In Tab. 6.17 is reported the obtained A matrix elements.

In Eq. 6.6, the ratio $N^{reco}_{promptJ/\Psi}/\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ gives the "real" number of prompt J/Ψ's generated within the $p_T^{J/\Psi}$ bin, neglecting the "mixing" effect of the $p_T^{J/\Psi}$ bins for the reconstructed candidate J/Ψ's; indeed, in this case the diagonal elements of matrix A coincide to the $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ values reported in Tab. 6.10.

The systematic due to this "mixing" effect was estimated in this way: the ratio $N^{reco}_{promptJ/\Psi}/\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ of Eq. 6.6 has been compared with the "real" number of prompt J/Ψ's generated within the $p_T^{J/\Psi}$ bin which is the i-th component of the vector $\vec{N}^{MC}_{promptJ/\Psi}$, and so it is obtainable inverting the A matrix in Eq. 6.37:

$$\vec{N}^{MC}_{promptJ/\Psi} = A^{-1}\vec{N}^{reco}_{promptJ/\Psi} \tag{6.38}$$

In Tab. 6.18 there are reported the number of prompt J/Ψ's expected for each $p_T^{J/\Psi}$ bin obtained with the $N^{reco}_{promptJ/\Psi}/\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ ratio and with the components of the vector

1.35%	0.88%	0.15%	0.09%	0.17%	0.05%	0.05%	0.15%	0.03%	0.12%	0.03%	0.17%	0.09%	0.07%
0.48%	1.71%	0.78%	0.15%	0.06%	0.02%	0%	0.10%	0.05%	0.11%	0.03%	0.03%	0.06%	0%
0.20%	1.47%	2.29%	1.14%	0.29%	0.09%	0.03%	0.08%	0.05%	0%	0.03%	0%	0.03%	0.03%
0.06%	0.41%	2.05%	2.75%	1.41%	0.38%	0.24%	0.10%	0%	0.03%	0.03%	0.11%	0%	0.06%	0.03%
0.06%	0.06%	0.32%	1.86%	2.49%	1.50%	0.45%	0.15%	0.05%	0.03%	0.06%	0.03%	0.03%	0.03%	0%
0.05%	0.04%	0.17%	0.77%	2.10%	2.73%	1.33%	0.51%	0.15%	0.03%	0.03%	0.05%	0.10%	0.03%	0.14%
0.03%	0.04%	0.05%	0.13%	0.77%	2.35%	3.65%	1.60%	0.28%	0.11%	0.03%	0%	0%	0.03%	0.03%
0.07%	0.08%	0.05%	0.04%	0.19%	0.96%	3.57%	2.86%	1.70%	0.19%	0%	0.03%	0.10%	0%	0%
0.03%	0.03%	0.04%	0.03%	0.06%	0.09%	0.88%	3.09%	4.52%	1.81%	0.21%	0.11%	0%	0%	0%
0.05%	0.12%	0.04%	0.04%	0.04%	0.09%	0.11%	0.13%	3.00%	5.18%	1.47%	0.19%	0.03%	0.06%	0.03%
0.03%	0.04%	0.04%	0.07%	0%	0%	0%	0%	0.43%	3.10%	5.88%	1.61%	0.13%	0.03%	0%
0.05%	0%	0.01%	0.01%	0%	0.02%	0.03%	0.10%	0%	0.25%	2.94%	7.12%	0.97%	0%	0%
0.04%	0.06%	0.05%	0.03%	0.02%	0.09%	0.03%	0.08%	0.05%	0%	0.09%	2.52%	10.17%	1.05%	0%
0.06%	0.01%	0.04%	0.04%	0.06%	0.05%	0%	0.05%	0%	0.05%	0.06%	0.11%	1.58%	11.52%	0.66%
0.04%	0.01%	0.03%	0.06%	0.04%	0.02%	0.05%	0.03%	0%	0.03%	0.06%	0%	0%	1.02%	15.67%

Table 6.17: *The obtained A matrix elements 15×15 (15 $p_T^{J/\Psi}$ bins considered in the analysis) as defined in Eq. 6.37. Rows refer to the $p_T^{J/\Psi}$ bin at reconstruction level and columns refer to the $p_T^{J/\Psi}$ bin at Monte Carlo truth level.*

Bins (GeV/c)	$N^{reco}_{promptJ/\Psi}/\mathcal{A}\epsilon_{trig}\epsilon_{reco}$	$\vec{N}^{MC}_{promptJ/\Psi}$	ϵ_{sys}
9.5÷11.5	36134	39243	7%
11.5÷12	11250	12150	8%
12÷12.6	10557	11190	6%
12.6÷13.2	9114	9752	7%
13.2÷13.7	6766	7307	8%
13.7÷14.2	5839	6365	9%
14.2÷14.7	4328	4544	5%
14.7÷15.3	5998	6358	6%
15.3÷16.1	5233	5599	7%
16.1÷17	4702	4984	6%
17÷18	4507	4777	6%
18÷19.5	4366	4715	8%
19.5÷21.5	4139	4387	6%
21.5÷25.5	4048	4331	7%
25.5÷90	3569	3819	7%

Table 6.18: *$N^{reco}_{J/\Psi}$, f_B and $N^{reco}_{promptJ/\Psi}$ for each $p_T^{J/\Psi}$ bin obtained by a fit on invariant mass distribution. Errors are obtained by statistical errors on fitted parameters.*

Source	**Factor affected**	ϵ_{sys}
luminosity	$\int \mathcal{L}dt$	5%
J/Ψ mass fit	$N^{reco}_{J/\Psi}$	1.8÷2.5%
l''_{xy} resolution and B-life time models	f_B	1÷5%
e-ID & reco sel.	$N^{reco}_{J/\Psi}$, f_B	1.0÷1.9%, 0.3÷0.9%
misalignment	f_B	0.3÷1.1%
Monte Carlo	$\mathcal{A}\epsilon_{trig}\epsilon_{reco}$	1÷6%
$p_T^{J/\Psi}$ bin mixing	$N^{reco}_{J/\Psi}$	5÷9%

Table 6.19: *Summary of systematic uncertainties considered in the differential cross section measurement of prompt J/Ψ's production. All the uncertainties are $p_T^{J/\Psi}$ dependent, except for the uncertainty from the integrated luminosity.*

$\vec{N}^{MC}_{promptJ/\Psi}$ computed with Eq. 6.38. The differences are considered to estimate the systematic error of the cross section measurement due to the $p_T^{J/\Psi}$ "mixing" effect during the J/Ψ candidates reconstruction. This systematic error is assumed to affect the number of reconstructed J/Ψ ($N^{reco}_{J/\Psi}$).

In Tab. 6.19 there are reported all the sources of systematic errors studied in this work, which measured factor of Eq. 6.6 is affected, and the ranges within errors are included. Systematic errors should affect the cross section measurement by a factor 10−25%.

6.6 Differential Cross Section Measurement of prompt J/Ψ Production in the J/Ψ →e⁺e⁻ Decay Channel

The differential cross section for each $p_T^{J/\Psi}$ is obtainable by Eq. 6.6 written as follows:

$$\frac{d\sigma}{dp_T}(J/\Psi) \cdot Br_{J/\Psi \to e^+e^-} = \frac{(1 - f_B) \cdot N^{reco}_{J/\Psi}}{\mathcal{A}\epsilon_{trig}\epsilon_{reco} \cdot \int \mathcal{L}dt \cdot \Delta p_T} \tag{6.39}$$

In Tab. 6.20 there are reported the differential cross section values obtained for each $p_T^{J/\Psi}$ bin with Eq. 6.39. There are also reported each factor used in the computation with the respective errors, both statistical and systematic ones. Statistical and Systematic errors are then propagated in the cross section computation. The final error of the differential cross section for each $p_T^{J/\Psi}$ bin is given by the squaring sum of the statistical and systematic errors ($\epsilon_{stat} \oplus \epsilon_{sys}$).

6.6.1 Comparison of measured data with theoretical predictions: discrimination of the various phenomenological models.

Finally, in Fig. 6.17 (a) case) there are shown the predicted theoretical distribution of the differential cross section for prompt J/Ψ's production *in the* $|\eta| < 2.5$ *region* as a function of the p_T of prompt J/Ψ, compared with the "measured" data (i.e. the measurement based on Monte Carlo simulation). The theoretical models considered are prompt production in the colour singlet model (CSM), and the production following the NRQCD theory, considering two different scenario (the *high radiation* and the *low radiation*) in the DGLAP evolution of the colour-octet $c\bar{c}_{(8)}$ states. Fig. 6.17 shows how it is possible to reconstruct the differential cross section of the produced prompt J/Ψ in the J/Ψ →e⁺e⁻ decay channel and to use it as a good observable of the theory to discriminate the various

Bins (GeV/c)	$N^{reco}_{J/\Psi}$	f_B	$\mathcal{A}\epsilon_{trig}\epsilon_{reco}$	Δp_T (GeV/c)	$\int \mathcal{L}dt$ ($\mathbf{pb^{-1}}$)	$\frac{d\sigma}{dp_T} \cdot BR_{J/\Psi\to e^+e^-}$ **nb/GeVc**$^{-1}$
9.5÷11.5	1490±(62,145)	0.368±(0.029,0.020)	2.607±(0.092,0.104)%	2	10.0±(-,0.5)	1.05±(0.10,0.13)
11.5÷12	862±(34,95)	0.353±(0.032,0.017)	4.96±(0.31,0.30)%	0.5	10.0±(-,0.5)	1.23±(0.14,0.17)
12÷12.6	1002±(41,88)	0.359±(0.029,0.011)	6.10±(0.35,0.37)%	0.6	10.0±(-,0.5)	0.98±(0.10,0.11)
12.6÷13.2	1033±(47,98)	0.363±(0.025,0.022)	7.22±(0.41,0.43)%	0.6	10.0±(-,0.5)	0.87±(0.08,0.11)
13.2÷13.7	841±(34,91)	0.381±(0.015,0.019)	7.70±(0.51,0.23)%	0.5	10.0±(-,0.5)	0.83±(0.07,0.11)
13.7÷14.2	796±(33,88)	0.379±(0.035,0.017)	8.46±(0.58,0.42)%	0.5	10.0±(-,0.5)	0.713±(0.087,0.098)
14.2÷14.7	704±(28,57)	0.359±(0.033,0.010)	10.42±(0.70,0.52)%	0.5	10.0±(-,0.5)	0.485±(0.058,0.053)
14.7÷15.3	840±(30,72)	0.370±(0.030,0.013)	8.82±(0.61,0.44)%	0.6	10.0±(-,0.5)	0.587±(0.066,0.068)
15.3÷16.1	867±(35,81)	0.365±(0.025,0.021)	10.53±(0.68,0.42)%	0.8	10.0±(-,0.5)	0.376±(0.038,0.047)
16.1÷17	844±(36,73)	0.392±(0.027,0.015)	10.91±(0.73,0.44)%	0.9	10.0±(-,0.5)	0.337±(0.035,0.038)
17÷18	798±(30,65)	0.376±(0.023,0.018)	11.05±(0.77,0.11)%	1	10.0±(-,0.5)	0.272±(0.027,0.029)
18÷19.5	892±(36,91)	0.417±(0.035,0.031)	11.91±(0.76,0.60)%	1.5	10.0±(-,0.5)	0.208±(0.023,0.030)
19.5÷21.5	875±(36,79)	0.368±(0.029,0.021)	13.36±(0.91,0.27)%	2	10.0±(-,0.5)	0.121±(0.013,0.014)
21.5÷25.5	929±(37,92)	0.388±(0.022,0.020)	14.03±(0.91,0.84)%	4	10.0±(-,0.5)	0.0642±(0.0060,0.0087)
25.5÷90	1033±(38,102)	0.406±(0.029,0.016)	16.7±(1.1,0.7)%	64.5	10.0±(-,0.5)	0.00389±(0.00040,0.00048)

Table 6.20: *$N^{reco}_{J/\Psi}$, f_B, $\mathcal{A}\epsilon_{trig}\epsilon_{reco}$ and Δp_T for each $p_T^{J/\Psi}$ bin, with the respective statistical and systematic errors (in brackets, the statistical ones in the left and the systematic ones in the right). Then the computed cross section values are also reported, with the propagation of the errors.*

Bins (GeV/c)	$N^{reco}_{promptJ/\Psi}$	$\int \mathcal{L}dt$ **(pb^{-1})**	Δp_T **(GeV/c)**	$\frac{d\sigma}{dp_T} \cdot BR_{J/\Psi\to e^+e^-}$ **(nb/GeVc^{-1})**
9.5÷11.5	39243	10	2	1.13±(0.10,0.13)
11.5÷12	12150	10	0.5	1.33±(0.14,0.17)
12÷12.6	11190	10	0.6	1.04±(0.10,0.11)
12.6÷13.2	9752	10	0.6	0.94±(0.08,0.11)
13.2÷13.7	7307	10	0.5	0.90±(0.07,0.11)
13.7÷14.2	6365	10	0.5	0.774±(0.087,0.098)
14.2÷14.7	4544	10	0.5	0.515±(0.058,0.053)
14.7÷15.3	6358	10	0.6	0.625±(0.066,0.068)
15.3÷16.1	5599	10	0.8	0.396±(0.038,0.047)
16.1÷17	4984	10	0.9	0.350±(0.035,0.038)
17÷18	4777	10	1	0.286±(0.027,0.029)
18÷19.5	4715	10	1.5	0.227±(0.023,0.030)
19.5÷21.5	4387	10	2	0.127±(0.013,0.014)
21.5÷25.5	4331	10	4	0.0642±(0.0060,0.0087)
25.5÷90	3819	10	64.5	0.0428±(0.00040,0.00048)

Table 6.21: *Differential cross sections measured values (Eq. 6.40) corrected by the deconvolution based on matrix A. The same errors obtained above are associated to the values.*

phenomenological models starting already from 10 pb^{-1} of LHC data. For high p_T values, it seems to be possible also to discriminate the various scenario foreseen in the evolution of the colour-octet state.

It has to be noted that the reconstruction in the electrons channel does not allow the measurement of the differential cross section for low p_T values due to the triggers thresholds and to the huge hadronic background which affects this decay channel, but it allows anyway to obtain a good observable, given that phenomenological models foresee differences starting only from relatively high p_T ($\sim \mathcal{O}$(8-10 GeV/c)).

In Fig. 6.17 there are discrepancies at low J/Ψ p_T regions ($\sim \mathcal{O}$(10 GeV/c)), where the measured cross section seems to be systematically under-estimated. This is due to the $p_T^{J/\Psi}$ "mixing" effect which is particularly present for low-p_T values. In Fig. 6.17 (the b) case) it is possible to see that if correct values of $N_{promptJ/\Psi}$ (the third column in Tab. 6.18) for each bin are considered and the differential cross sections for each bin are re-calculated as

$$\frac{d\sigma}{dp_T}(J/\Psi) \cdot Br_{J/\Psi\to e^+e^-} = \frac{N_{promptJ/\Psi}}{\int \mathcal{L}dt \cdot \Delta p_T} \tag{6.40}$$

then discrepancies vanish. In Tab. 6.21 is reported the differential cross section computed

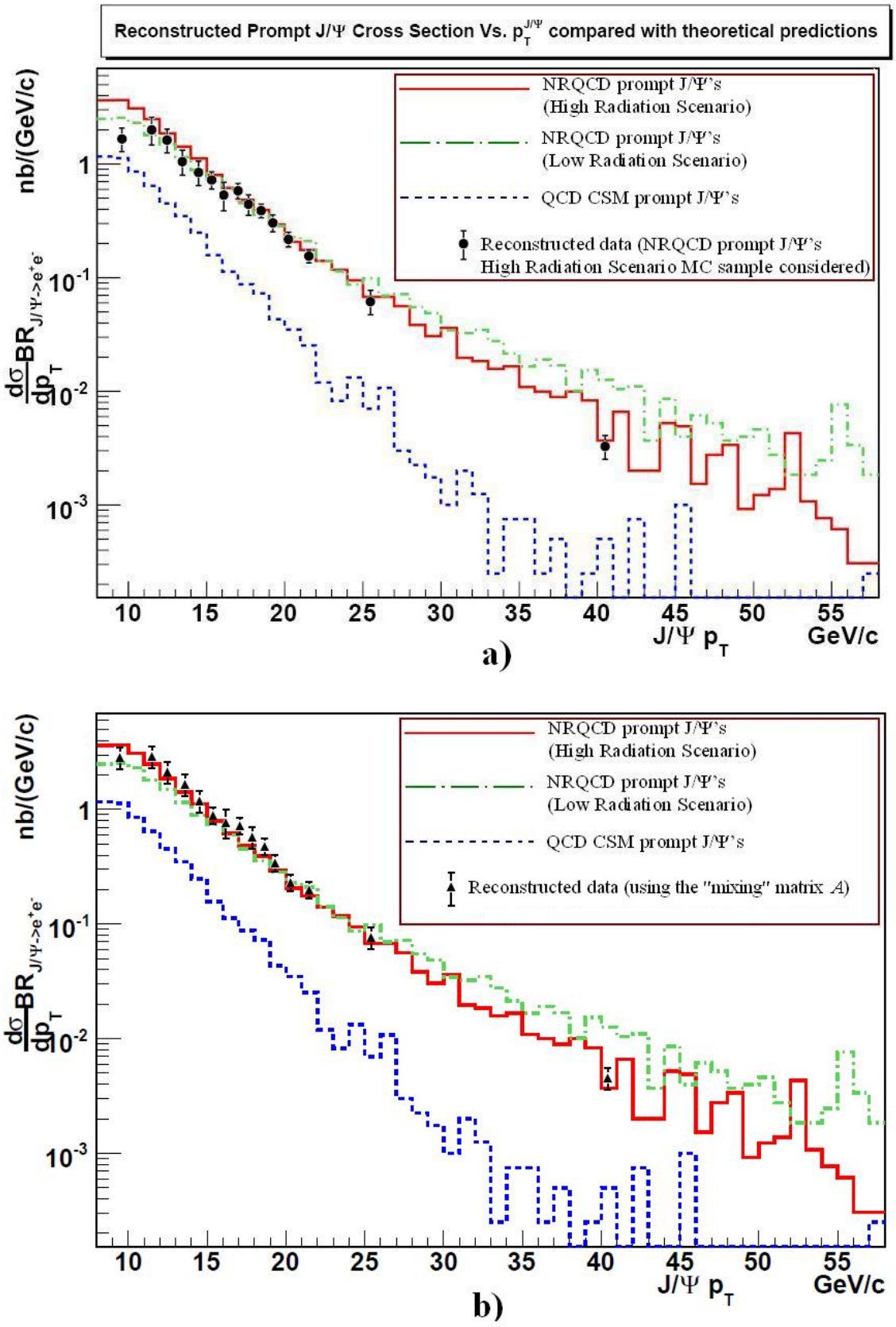

Figure 6.17: *a) Differential cross section of prompt J/Ψ's production (in the $|\eta| < 2.5$ region) as a function of the p_T of the J/Ψ: comparison between various theoretical predictions (simulation made with* **Pythia***) -**lines**- and measured values considering the J/Ψ $\rightarrow e^+e^-$ decay channel (reconstruction made starting from the Monte Carlo sample) -**dots**. b) The same distributions, obtained with the number of prompt J/Ψ's for each $p_T^{J/\Psi}$ bin obtained from the "mixing" matrix A.*

with the deconvolution of the measured values using the matrix A.

Conclusions

The study of CMS detector performance in the measurement of prompt J/Ψ's production in $\sqrt{s}$ = 10 TeV proton-proton collisions at the LHC in the J/Ψ $\rightarrow e^+e^-$ decay channel with the CMS detector was presented.

This measurement is very important in the LHC start-up phase because it can be made after the first inverse picobarns of data collected. It can solve some questions about the heavy quarkonia production in p-p collisions which are yet not completely understood. The reconstruction of J/Ψ's in the J/Ψ $\rightarrow e^+e^-$ decay channel allows to improve the understanding of the detector, in particular is a "reference candle" to monitor the ECAL calibration in low-E_T regions.

In this work it was shown that it is possible to use the J/Ψ $\rightarrow e^+e^-$ decay channel to measure the J/Ψ p_T production differential cross section, yet starting from the first $\int \mathcal{L}dt$ =10 pb^{-1} collisions at 10 TeV.

Anyway, the big challenge in electrons reconstruction at hadronic colliders is the huge hadronic background which requires high trigger thresholds, making very difficult physics studies in the low-mass region. The problem is to ensure a statistically significant signal sample keeping low contamination from background processes.

The low instantaneous luminosities foreseen in the LHC start-up phase allow to adopt trigger thresholds of the order of 5 GeV for the transverse electron energy, which is the mean value expected from prompt J/Ψ decays. This has given the motivation to study the feasibility of the measurement of prompt J/Ψ's production with the CMS detector in the e^+e^- decay channel.

The study has started with the Monte Carlo simulation of the production of prompt J/Ψ's, studying the various phenomenological model implemented in the Monte Carlo generator `Pythia`, as exposed in Chapter 3. Then, it has been presented a low energy-adapted electron reconstruction algorithm as an alternative to the standard electron reconstruction used by the CMS experiment, which was developed and optimized for high energy electrons (such as pairs decaying from a Z^0 boson). In this way, it has been possible to increase the J/Ψ candidates statistics, while keeping the same signal-to background ratio

with respect to the standard electron reconstruction.

A trigger for low-mass resonances decaying in e^+e^- pairs has been officially included in the CMS software. In this work has been studied how this trigger allows the collection of signal events for various luminosities scenario. Additional trigger paths are used to optimize the total signal efficiency.

To make possible a good measurement of the prompt J/Ψ's production, all possible background sources have been studied. In particular, the hadronic background topology has been studied to understand which off-line selections could be useful to improve the background rejection in the reconstruction. A measurement can be performed with the first data collected by the CMS detector for an integrated luminosity of 10 pb^{-1}. The differential cross sections are determined as a function of the J/Ψ transverse momentum, integrated over the pseudorapidity range $|\eta| < 2.4$. The non-prompt J/Ψ's (deriving from B-hadrons decays, which represent an irreducible background) have been disantangled by using the large B-hadron lifetime. The most important systematic uncertainties to the measurement were addressed. The experimental error is dominated by the systematics, even with the limited integrated luminosity that was considered. Some of the systematics will, of course, benefit from the larger data set that will be collected. Also a better alignment could help to better disantangle the B-hadron component and reject the hadronic background, respectively.

The adopted electron reconstruction allows to use robust tracking algorithms, keeping good efficiency even with tracks showing high impact parameter from the interaction point. The standard electron reconstruction, instead, suffers from a neat efficiency decrease for high impact parameter tracks, posing a serious bias in lifetime-based studies, like the disentanglement of flying B-decay and non-flying prompt components of J/Ψ production.

By the end of 2010, when hopefully an integrated luminosity of order of ∼100 pb^{-1} will be collected, the J/Ψ differential cross section measurement are expected up to a transverse momentum of ∼90÷100 GeV/c and possibly beyond. Around 90000 prompt J/Ψ's (∼130000 total J/Ψ's) with a p_T higher than ∼9-10 GeV/c might be collected at that moment.

Anyway, yet starting from 10 pb^{-1} of collected data it is possible to reconstruct around 9000 prompt J/Ψ's up to a transverse momentum of ∼40÷50 GeV/c: starting from them, the prompt J/Ψ differential cross section $\frac{d\sigma}{dp_T}$ can be measured with an uncertainty of $\mathcal{O}(10\%)$, which is sufficient to disentangle NRQCD and CSM models from prompt J/Ψ production.

Appendix A

Invariant Mass Measurement

Consider the two-body decay

$$A \rightarrow 1 + 2 \tag{A.1}$$

where we are interested to the reconstruction of the *invariant mass* of A. Special relativity says that if $p_1^\mu = (E_1, c\vec{p}_1)$ and $p_2^\mu = (E_2, c\vec{p}_2)$ are respectively the four-momentum of particle 1 and particle 2, than the invariant obtained by the scalar product in the Minkowsky metrics (called *invariant mass*) is given by

$$(p_1 + p_2)^\mu (p_1 + p_2)_\mu = (E_1 + E_2)^2 - c^2|\vec{p}_1 + \vec{p}_2|^2 \tag{A.2}$$

gives the square of the module of the four-momentum of particle A in the rest frame, i.e. the square of its mass M^2. In the relativistic approximation $E \sim c|\vec{p}|$ we can obtain

$$\begin{aligned} M^2c^4 &= E_1^2 + E_2^2 + 2E_1E_2 - c^2|\vec{p}_1|^2 - c^2|\vec{p}_2|^2 - 2c^2\vec{p}_1 \cdot \vec{p}_2 = \\ &= 2E_1E_2 - 2c^2|\vec{p}_1||\vec{p}_2|\cos\delta \end{aligned} \tag{A.3}$$

where δ is the angle between particles 1 and 2.

Experimentally, the measured variables of 1 and 2 to reconstruct the invariant mass of A are the transverse momentum/energy (the same things in the relativistic approximation) E_T, the pseudo-rapidity η and the azimuthal angle ϕ. From them, we have to reconstruct each component of the four-momentum of particles 1 and 2, i.e. $p^0 = E$, $p^1 = cp_x$, $p^2 = cp_y$ and $p^3 = cp_z$, where $\vec{p} = (p_x, p_y, p_z)$ is the three-vector of the particle momentum. Remembering the definition of η (Chapter 2, Eq. 2.6) the following relations can be obtained:

$$\begin{aligned} \cosh\eta &= \frac{1}{\sin\theta} \\ \sinh\eta &= \mathrm{cotan}\theta \end{aligned} \tag{A.4}$$

where θ is the polar angle respect to the z axis. Remembering the transformation matrix from cartesian to spherical coordinates (and assuming the natural units used in this work, where $c = 1$), the four-momentum of particle $1, 2$ is

$$p^{\mu}_{1,2} = (E_{T1,2} \cdot \cosh\eta_{1,2}, E_{T1,2} \cdot \cos\phi_{1,2}, E_{T1,2} \cdot \sin\phi_{1,2}, E_{T1,2} \cdot \sinh\eta_{1,2}) \tag{A.5}$$

So the invariant mass can be computed as $M^2 = (p_1 + p_2)^{\mu}(p_1 + p_2)_{\mu}$, as written in Eq. A.2.

In `CMSSW` is implemented a class from `ROOT` libraries (`TLorentzVector`) with a member which returns the four-momentum passing directly E_T, η, ϕ (in the cases of ultra-relativistic particles it is not required the particle mass m).

Appendix B

The *Maximum Likelihood* Fitting Method

Given the probability density function B.1

$$x \rightarrow f(x|\theta) \tag{B.1}$$

where x is the aleatory variable and θ is the parameter which describes the B.1, we can define the *likelihood function* as the function which research the statistical estimator of the parameter θ:

$$\theta \rightarrow f(x_{i=1}^{N}|\theta) \tag{B.2}$$

where $x_{i=1}^{N}$ is the ensemble of N observations of aleatory variable x and θ is the parameter which have to be estimated.

Usually the likelihood function is written as

$$\mathfrak{L}(\theta|x_{i=1}^{N}) \tag{B.3}$$

and, in the case the N observation are independent, it can be written as the mathematical product of each single observation

$$\mathfrak{L}(\theta|x_{i=1}^{N}) = \prod_{i=1}^{N} f(\theta|x_i) \tag{B.4}$$

The estimator of the θ parameter is obtained maximizing the likelihood function (the *maximum likelihood method*). To simplify the computation, it is often preferred to maximize the function $\ln\mathfrak{L}$:

$$\ln\mathfrak{L}(\theta|x_{i=1}^{N}) = \sum_{i=1}^{N} \ln f(\theta|x_i) \tag{B.5}$$

Usually, in the fitting algorithms the *maximum likelihood method* is used to estimate the parameters of the observables of phenomenological model predicted by the theory comparing with experimental data. There are two methodologies of fitting using the maximum likelihood: the *binned maximum likelihood* and the *unbinned maximum likelihood*.

The Unbinned Maximum Likelihood

If the number of observations of the aleatory variable is low to fill histograms with enough statistic of events per bin, or when the probability density function is described by more than one aleatory variable (i.e. multidimensional probability density functions) it is preferred to use the *unbinned maximum likelihood* in the fitting algorithms. This kind of fits simply adopt the original definition of maximum likelihood method exposed on top, directly on the whole ensemble of observations $x_{i=1}^{N}$ and retrieve the parameters estimations.

The Binned Maximum Likelihood

The *binned maximum likelihood* is used in fitting algorithms when the number of observations of the aleatory variable is enough high to make difficult the technique used for the maximization of the unbinned maximum likelihood function. So it is preferred to collect the observations in an histogram of n bins (where generally $n \neq N$). Given the probability density function $f(x|\theta)$, the mean values $\nu_{i=1}^{n}$ of the numbers of entries filling the n bins is given by:

$$\nu_i(\theta) = N \cdot \int_{x_i^{min}}^{x_i^{max}} f(x|\theta)dx \tag{B.6}$$

where x_i^{min} and x_i^{max} are the minimum value and the maximum value relative to each i-th bin, and N is the total number of observations. The histogram can be seen as a single observation of a n-dimensional variable $\vec{k} = (k_1, k_2, ...k_n)$ with the probability density function given by the multinomial distribution

$$f(\vec{k}|\nu(\theta)_i^n) = \frac{n!}{k_1!...k_n!}\Big(\frac{\nu_1(\theta)}{n}\Big)^{k_1}...\Big(\frac{\nu_n(\theta)}{n}\Big)^{k_n} \tag{B.7}$$

where k_i is the number of observations filling the i-th bin of the n bins.

The likelihood function is obtained passing to the logarithm of Eq. (B.7):

$$\ln\mathfrak{L}(\theta) = \sum_{i=1}^{N} k_i \ln(\nu_i(\theta)) \tag{B.8}$$

The estimator of θ is obtained by the maximization of the $\ln\mathfrak{L}(\theta)$ function[103].

Bibliography

[1] `https://twiki.cern.ch/twiki/bin/view/CMS/SUSY`.

[2] `https://twiki.cern.ch/twiki/bin/view/CMS/EXOTICA`.

[3] See e.g. F. Halzen and A. Martin, *Quarks and leptons: an introductory course in modern Particle Physics*, Wiley, NY USA, 1984.

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

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

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

[7] See e.g. S. Dawson, *The Standard Model Intermediate mass Higgs Boson*, ed. Gordon L. Kane, World Scientific, 1997.

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

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

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

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

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

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

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

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

[16] S. Catani, M. Grazzini, *Foundamental Interactions Theory and Quantum Chromo-Dynamics*, lectures given at PhD course at *Scuola di Dottorato in Scienze di Firenze*, 2007.

[17] B. Willard, *Computation of NRQCD Parameters in Potential Models*, August 2004.

[18] G. A. Scheler, *Quarkonium production and decays*, CERN-TH 7170/94.

[19] CDF Collaboration, Phys. Rev.Lett. 79: 578-583, 1997.

[20] V. D. Barger, W. Y. Keung and R. J. Phillips, Phys. Lett. B 91 (1980) 253.

[21] W. E. Caswell and G. P. Lepage, Phys. Lett. B 167 (1986) 437.

[22] J. w. Qiu and G. Sterman, Int. J. Mod. Phys. E 12 (2003) 149, `arXiv:hep-ph/0111002`.

[23] A. V. Manohar, *The HQET/NRQCD Lagrangian to order* α_s/m^3, UCSD/PTH 97-01, `arXiv:hep-ph/9701294`.

[24] G. T. Bodwin, E. Braaten and G. P. Lepage, Phys. Rev. D 51, 1125 (1995), `arXiv:hep-ph/9407339`.

[25] R. Casalbuoni, *Quantum Field Theory*, Lectures given at Geneva University, 1995.

[26] M. Neubert, Phys. Rept. 245, 259 (1994), `arXiv:hep-ph/9306320`.

[27] M. Bargiotti, V. Vagnoni, *Heavy Quarkonia Sector in PYTHIA 6.324: Tuning, Validation and perspectives at LHC(b)* LHCb-2007-042, CERN June 2007.

[28] N. Brambilla, A. Vairo et al., *Heavy Quarkonium Physics*, CERN-2005-005 20 June 2005.

[29] The CDF Collaboration, Phys. Rev.Lett.79:578-583, 1997.

[30] `http://home.thep.lu.se/~torbjorn/Pythia.html`.

[31] T. Affolder et al. (CDF Collaboration), Phys. Rev. Lett. 85 (2000) 2886, `arXiv:hep-ex/0004027`.

[32] http://www.deasy.de/qwg07/talks/Kraan_071007.pdf

[33] A.C.Kraan, *Experimental Aspects of Heavy Quarkonium Production at the LHC*, July 2008, arXiv: hep-ex/08073123v1.

[34] H. Fritzsch, Phys. Lett. B 67 (1977) 217.

[35] F. Halzen, Phys. Lett. B 69 (1977) 105.

[36] M. Kramer, Nucl. Phys. B 459 (1996) 3, ArXiv: hep-ph/9508409.

[37] Zongchang Yang, *Study of a J/Ψ Cross Section Measurement with Early Data at CMS*, Beijing May 2009 (PhD Thesis).

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

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

[40] S. Chatrchyan et al. (The CMS Collaboration), JINST 3 (2008) S08004.

[41] G. Aad et al. (The ATLAS Collaboration), JINST 3 (2008) S08003.

[42] A. Augusto Alves Jr et al. (The LHCb Collaboration), JINST 3 (2008) S08005.

[43] K. Aamodt et al. (The ALICE Collaboration), JINST 3 (2008) S08002.

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

[45] http://cdsweb.cern.ch/record/1190469

[46] http://cdsweb.cern.ch/collection/sLHC%20Project%20Reports?ln=it&as=1

[47] CERN Workshop on Standard Model Physics (and more) at the LHC, CERN Yellow Report 2000-004.

[48] The CMS Collaboration, *The Magnet Project: Technical Design Report*, CERN/LHCC 97-10, CMS TDR Vol. 1 (1997).

[49] T. Virdee, *CMS Status*, presentation given in CMS Week, 16 March 2009.

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

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

[52] *The Electromagnetic Calorimeter Project Technical Design Report*, CERN/LHCC 97-33, CMS TDR 4, December 15, 1997.

[53] E. Auffray et al., *Development of a Uniformisation Procedure for the $PbWO_4$ Crystals of the CMS Electromagnetic Calorimeter*, CMS NOTE 2001/004.

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

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

[56] C. Rovelli, *The CMS electromagnetic calorimeter and the search for the Higgs boson in the decay channel $H \rightarrow WW^{(*)} \rightarrow 2e2\nu$*, PhD Thesis, Ecole Polytechnique De Palaiseau, 2006.

[57] P. Lenzi, *Study of Z^0/γ^* + jets production with the CMS detector using matched Matrix Element-Parton Shower event generators*, PhD Thesis, Università degli Studi di Firenze, 2008.

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

[59] M. Dasgupta, *Parton shower Monte Carlos vs resummed calculations for interjet energy how observables*, `arXiv:0706.2630 (hep-ph)`.

[60] T. Sjostrand et al., Eur. Phys. J. C39 (2005) 129.

[61] M. Bargiotti and V. Vagnoni, LHCb-2007-042 (2007).

[62] E. Levin, *An introduction to Pomerons*, arxiv:`http://arxiv.org/abs/hep-ph/9808486`.

[63] Particle Data Group, *Review of particle physics*, Physics Letters B 667 Issues 1-5, Junnary 2009.

[64] T. Sjöstrand, L. Lönnblad, S. Mrenna, P. Skands, *PYTHIA: Physics and Manual*, hep-ph/0308153 LU TP 03-38, August 2003.

[65] A.C. Kraan, *Prompt J/Ψ production mechanism studies with the CMS detector*, CMS AN 2007/042, 2007.

[66] https://twiki.cern.ch/twiki/bin/view/CMS/MBUETuning.

[67] Andrey Pozdnyakov (CMS experiment, QCD Low p_T POG), private comunication.

[68] http://cms.cern.ch/iCMS/jsp/page.jsp?mode=cms&action=url&urlkey=CMS_OFFLINE.

[69] H. Schildt, *C++ (Fourth Ed.)*, McGrawHill, 2006.

[70] M. Lutz, *Programming Python*, Sebastopol, CA: OŔeilly, 1996.

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

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

[73] S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. A506 (2003) 250−303.

[74] E. Meschi et al., *Electron Reconstruction in the CMS Electromagnetic Calorimeter*, CMS NOTE 2001/034.

[75] R. Ranieri, *Trigger Selection of WH→ $\mu\nu b\bar{b}$ with CMS*, PhD Thesis, Università degli studi di Firenze, 2003.

[76] P. Billoir and S. Qian, *Fast Vertex Fitting with a Local Parametrization of Tracks*, Nucl. Instrum. Meth., A311, (1992) 132−150.

[77] R. Fruhwirth, Nucl.Inst. & Methods A 262, 1987, 444.

[78] H. Bethe and W. Heitler, Proc.R.Soc.London A 146 (1934) 83.

[79] W. Adam et al., *Reconstruction of electrons with the Gaussian Sum Filter in the CMS tracker at LHC*, CMS NOTE 2005/001.

[80] R. Fruhwirth et al., Comp.Phys.Comm.110 (1998) 80.

[81] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideEgamma.

[82] http://indico.cern.ch/getFile.py/access?contribId=2 &resId=0&materialId=slides&confId=43511.

[83] http://indico.cern.ch/getFile.py/access?contribId=2 &resId=1&materialId=slides&confId=53939.

[84] https://twiki.cern.ch/twiki/bin/view/CMS/WorkBookOfflinePrimaryVertexFinding.

[85] CMS Collaboration, *Physics and Technical Design Report Vol. I*, CERN, February 2006.

[86] https://cmsweb.cern.ch/dbs_discovery/index?userMode=expert.

[87] Chris Seez and Paolo Meridiani (CMS experiment, EGamma POG), private comunication.

[88] The CMS Collaboration, *The TriDAS Project Technical Design Report, Volume 1: The Trigger Systems*, CERN/LHCC 2000-038, CMS TDR 6.1 (2000).

[89] The CMS Collaboration, *The TriDAS Project Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger*, CERN/LHCC 2002-26, CMS TDR 6.2 (2002).

[90] https://twiki.cern.ch/twiki/bin/view/CMS/TriggerMenuDescription8E29Devel.

[91] https://twiki.cern.ch/twiki/bin/view/CMS/TriggerMenuDescription1E31Devel.

[92] A. Ghezzi, R. Covarelli, CERN, private comunication.

[93] A. Ghezzi, R. Covarelli et al., presentation given in the EGAMMA POG at CERN, http://indico.cern.ch/getFile.py/access?contribId=2&resId=1&materialId=slides&confId=53939.

[94] A. Ghezzi, presentation given in the EGAMMA HLT Meeting at CERN, http://indico.cern.ch/getFile.py/access?contribId=1&resId=0&materialId=slides&confId=44057.

[95] https://twiki.cern.ch/twiki/bin/view/CMS/Quarkonia.

[96] https://twiki.cern.ch/twiki/bin/view/CMS/WorkBookBTagging.

[97] http://cmslxr.fnal.gov/lxr/source/DataFormats/TrackReco/interface/TrackBase.h.

[98] D. Acosta et al. (CDF Collaboration), Phys. Rev. Lett. 88 (2002) 161802.

[99] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideKinematicVertexFit/.

[100] The CMS Collaboration, *The CMS Offline Workbook - Vertex Reconstruction Tools*, https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideVertexTools.

[101] D. Acosta et al. (CDF Collaboration), Phys. Rev. D 71 (2005) 032001.

[102] A. Lucaroni et al., *Updates on the alignment of CRAFT reprocessed data*, presentation given at the Alignment Meeting at CERN, http://cms.pg.infn.it/sites/default/files/andrea26:02.pdf.

[103] G. Cowan, *Statistical data analysis*, Oxford, UK: Clarendon (1998) 197 p.

[104] http://mathworld.wolfram.com/Convolution.html.

[105] The RooFit Manual, http://roofit.sourceforge.net/docs/RooFit_Users_Manual_2.07-29.pdf.

[106] https://twiki.cern.ch/twiki/bin/view/CMS/ElectronTagAndProbe/.

[107] G. Anelli et al. (The TOTEM Collaboration), JINST 3 (2008) S08007.

[108] https://twiki.cern.ch/twiki/bin/view/CMS/LumiWikiHome.

[109] CDF Collaboration, Phys. Rev. D71 (2005) 032001.

Acknowledgements

There are many people of the CMS collaboration I would like to thank.

First of all, I would like to thank my supervisor Simone Paoletti, which followed me in this three years of PhD in Florence and has shown constant dedication in following my work.

Then, I want to thank the other CMS people in Florence: Marco Meschini, Raffaello D'Alessandro, Carlo Civinini, Vitaliano Ciulli, Giacomo Sguazzoni, Chiara Genta, Ettore Focardi, Elisabetta Gallo and Giuseppe Barbagli (the list is sorted in totally random way), which have helped me many times for the organization, improvement and the support of this work.

I have to thank also many people at CERN. I'm very grateful to Chris Seez and Paolo Meridiani (CMS Egamma POG conveners) for many useful discussions about electrons reconstruction and background study.

I sincerely appreciate Aafke Kraan for her great help on the comprehension of the quarkonia production physics and the proper tuning of the `Pythia` generator for the simulations.

Thanks also should be given to Roberto Covarelli and Alessio Ghezzi for their support on trigger studies in $J/\Psi \rightarrow e^+e^-$ events.

I would like to thank also Roberta Arcidiacono and Chiara Rovelli for their collaboration in the background study and for many other "little" helps on software code comprehension.

Finally, I thank the ex-members of the CMS team in Florence, now at CERN: Matteo Sani, Riccardo Ranieri, Stefano Mersi, Nicolò Magini, Andrea Bocci. They have teached me all "secrets" of the detector, of the software code, and more in general of analysis in high energy physics.

Ringraziamenti

Eccoci...

La fine di un lavoro non è mai merito di un singolo. Per quanto chi gli è stato vicino non sia mai stato in grado di aiutarlo "tecnicamente" in quel lavoro, comunque sarebbe stato indubbiamente più difficile arrivare al fine.

Per questo è più che doveroso ringraziare in questa sede tutti coloro che hanno contribuito e mi sono stati a fianco in questi *cortissimi* tre anni di dottorato.

Desidero partire più di tutti dalla mia famiglia: mia sorella e i miei genitori, che hanno sempre sostenuto la mia idea di intraprendere l'attività di ricerca senza farmi gravare sulle spalle licenziamenti, crisi e quant'altro, e che anzi hanno dovuto subire pure la "scelleratezza" di un figlio che si spende i soldi dietro una moto da oltre 12000 euro. A loro vanno i più sentiti ringraziamenti.

Ringrazio poi Erica e la sua famiglia: loro ancor di più hanno subito il peso delle mie scelte, e nonostante tutto mi sono stati vicini anche in questo lavoro, incoraggiandomi là dove io più volte volevo mollare. Senza anche il loro supporto, non sarei a scrivere queste quattro righe. Ery, comunque sia andata, sarai una persona importante per me, una delle più importanti della mia vita...

Devo ringraziare poi coloro che in passato mi hanno dato la giusta convinzione per andare avanti: Bruno, Patrizio, Tommasino, Stefano (viaggiava il Ninja, eh? Chissà cosa pensi ora che IO ho un CBR...)...tutte persone che ho ammirato tantissimo nella mia vita e che hanno ammirato ME, al contrario di quanto spesso faccio io stesso...tutte persone che non ci sono più a farlo...

Ringrazio poi i "fratelli" del Cantiere, amici ritrovati dopo anni, figli della stessa sorte, cresciuti dalla stessa strada, col sangue dello stesso Sud, e che dunque chiamo "fratelli". Carlo, Matteo e Davide (*u' cunnutu sinn'è fujutu 'nta ll'America*); le chiaccherate alle 4.00 del mattino davanti al lampredottaro...nessuno di loro ha proseguito con gli studi (che siano stati più furbi di me?), tutti che però mi hanno "obbligato" a farlo a suon di

pedate, quando volevo tirarmi indietro.

Samuele, compagno di studio e "di vita" come lo chiamo io. Sempre pronto ad ascoltarmi, sempre pronto a non giudicare mai in maniera distruttiva le mie "scelleratezze". Spesso non ho mai contraccambiato abbastanza, e ancora non mi ha mai mandato a fan...un altro che ha abbandonato a mezzo il dottorato, ma quando gli dicevo che volevo farlo io m'ha spronato a non tirarmi indietro, perché secondo lui la mia passione "si vede distintamente ad occhio nudo".

Un grazie a Marco Carbone. Padre di famiglia impeccabile e ottimo amico, che risate in quelle serate di studio matto, la "gita" a Empoli, il "seminario" su LHC a Pratolino. L'essermi stato vicino nei momenti di sconforto. grazie.

Un grazie affettuoso anche a Selenia e Letizia, ex-compagne di facoltà che conosco oramai da anni, che lungo tutto il percorso di studi mi hanno sempre incoraggiato a guardare in avanti, che spesso mi perdo a guardare indietro. E così è stato anche durante il dottorato...grazie Selè per i pranzi e le cene offerte!!!!!!!

Un grazie di cuore ai tecnici (che ne hanno sempre "saputo" più dei fisici alla fine) con cui ho collaborato su al CERN durante la mia esperienza di test sui cavi di alimentazione del tracciatore di CMS. Come dimenticare Nunzio Giudice con le sue canzoni napoletane in vetta a CMS, Nunzio Guardone (che ne pensi di testare altri quattro cavi??) e Denis Kostylev, amico anche al di fuori del CERN. Grazie anche a loro ho avuto la giusta fiducia per affrontare questi anni di dottorato in allegria.

E infine i compagni di Università, coi quali alla fine ho passato dalle 8 alle 10 ore al giorno assieme a volte. Come posso pensare che non mi abbiano aiutato in questo dottorato??

Mi "tocca" iniziare con Giulio...a parte gli scherzi, ho avuto la fortuna di trovare in un collega un *amico*, fortuna che non tocca a tutti...quasi dieci anni a fianco! Quante cose sa di me, ha visto di me, anche le "scelleratezze", mi ha sempre consigliato bene e ad evitare (diciamolo, spesso purtroppo inutilmente...) alcuni miei comportamenti "sconvenienti".

Poi c'e' Chiara, peccato che quest'anno (proprio l'ultimo, quello in cui mi sono comportato peggio: lei che mi rimetteva sempre in riga) sia partita. Anche lei dieci anni fianco a fianco, anche lei che mi ha sempre ripreso dagli sbagli, che m'ha preso a pedate ogni qual volta dicevo "io mollo...". Andrea, quest'anno un pó entrambi "in difficoltà", spiazzati dal fatto che a volte la vita cambia, e vuole che cambiamo anche noi, e che mi ha aiutato a ingoiare rospi grossi come case, nella parte finale della tesi, grossi da impedirmi quasi di scriverla.

Un ringraziamento fortissimo poi va ai miei colleghi dottorandi di CMS Firenze: Antonio, a cui voglio un bene spropositato, quante chiaccherate su ogni argomento. Simone, mia guida perenne a cui chiedo scusa tutte le volte che gli ho mancato di rispetto. Cristina, che mi ha saputo più di una volta consigliare, consolare e soprattutto riprendermi là dove sbagliavo (anche in questo caso, spesso inutilmente)...

Un sentito grazie anche a Franco Capello. In missione perenne al CERN, si fa precedere dalle sue imprecazioni. E' stato molto utile il suo aiuto nella mia tesi riguardo la comprensione dell'allineamento del rivelatore.

E poi c'è Giacomo...Giacomino!! Compagno di stanza, di "elettroni" e di *amarezza*. Quanto ci siamo amareggiati a vicenda per via del lavoro e delle donne che non abbiamo!! Ma soprattutto, intenditore sovraffino di musica: "...e la sedia del babbo che adesso è la mia...come è bella la vita...". Mi voleva scrivere pure una dedica pensando che fossi morto in moto...un mito!! Sandro Gonzi, amico in questi anni di università, e adesso finalmente "collega". Fantastiche le sue perle di saggezza. Tommy Braz, vicini di stanza, ha mostrato una grande sensibilità, per questo lo stimo. Vanessa, per tutto quello che mi ha detto.

Poi ci sono i compagni più recenti...non per questo, che mi abbiano aiutato meno in quest'ultimo anno che io ho considerato il più duro. Inizio dai teorici: Giovanni V., compagno di tante chiaccherate davanti a un buon bicchiere quando si poteva, Giancarlo che mi mette sempre allegria con le sue massime su Checco Zalone, l'insubre Damiano (alooorha!!!!) che ha saputo mostrarmi una sincera vicinanza nei peggiori momenti di quest'anno ("amore, poi morte, poi...chi ne ha saputo più nulla?"), una sensibilità che dovrebbe mostrare di più agli altri, al posto di tante (diciamolo) caz...Grazie a Korhad, per i suoi aiuti a comprendere cose prettamente teoriche che hanno riguardato il mio lavoro, futuro premio Nobel per la Fisica...Pucci, compagno delle lezioni infinite di Relatività, immersi nei manifolds Lagrangiani...

E come dimenticare Stefano Carboni, che alla prossima storia agricola che mi racconta lo rincorro...Giovanni Sorry, maestro d'amarezza e di scommessa, grazie per avermi fatto ridere con le storie marchegiane...

Poi ci sono pure gli strutturisti, diffusi ovunque come le malattie...Marco Schioppo, compagno di stanza e di moto (che non ha) che mi fece capire che dovevo ripartire da me stesso quando non funzionava più niente, Marco Tarallo, per i pranzi assieme. Baront, vicino di casa, vicini di stanza, ma lontani col cuore, ed è meglio cosi (ricorda i due pacchetti di sigarette). Vera, che ha saputo starmi vicino là dove c'era bisogno, anche se troppo occupata dalle faccende del laboratorio a volte...

Eleonora...anzi, Ele.....hai visto? Ti ricordi cosa mi dicesti a maggio

mentre ero al CERN? Alla fine le “ho tirate fuori”, e l’ho scritte, queste quattro righe di tesi. Almeno in questa occasione, l’ho fatto...

A tutti voi, e ad altri che non sono stati inclusi in questa “lista” ma a cui “dal vivo” continuamente mostro il mio affetto...

GRAZIE.