UNIVERSIDADE TÉCNICA DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

Measurement of the inclusive photon production cross section and study of associated W-photon production in proton-proton collisions at the LHC

Pasquale Musella
(*Mestre*)

Dissertação para obtenção do Grau de Doutor em Física

Orientador: Doutor João Manuel Coelho Dos Santos Varela

Júri
Presidente: Presidente do Conselho Científico do IST
Vogais: Doutora Maria Teresa Rodrigo Anoro
Doutor Gustavo Da Fonseca Castelo-Branco
Doutor João Manuel Coelho Dos Santos Varela
Doutor João Carlos Carvalho De Sá Seixas
Doutor Michele Gallinaro

Dezembro de 2010

Abstract

The Large Hadron Collider (LHC) at CERN Geneva is presently colliding beams of protons at a centre-of-mass energy of 7 TeV. Operations started at the end of 2009 and it is expected to accumulate a total integrated luminosity of 1 fb^{-1} by the end of 2011, opening the first window on the Physics of fundamental interactions at the TeV scale. The Compact Muon Solenoid (CMS) is a general-purpose detector currently operating at the LHC. The CMS Electromagnetic Calorimeter (ECAL) has been designed to achieve excellent energy and position resolution for electron and photons and is composed by 76000 $PbWO_4$ crystals.

This thesis describes the original contributions of the author to the preparation of the CMS experiment and analysis of the first collision data.
The trigger and data acquisition software of the CMS ECAL is presented and its performance discussed, as well as the performance of the ECAL detector and trigger system.
The first results on the measurement of the inclusive isolated photon production cross section at $\sqrt{s} = 7$ TeV using 2.9 pb^{-1} of pp collisions are presented. Finally, a study of the CMS sensitivity to W-photon production is presented and perspectives for the total and differential cross-section measurements with early LHC data are discussed.

Keywords

isolated photons cross section electroweak physics gauge couplings QCD CMS

PACS Codes

14.70.Bh 14.70.Fm 07.05.Hd 29.40.Vj

Resumo

O grande colisionador de hadrões (Large Hadron Collider, LHC) do CERN em Genebra encontra-se presentemente a colidir feixes de protões com uma energia no centro de massa de 7 TeV. A sua entrada em operação deu-se no final de 2009 e espera-se que acumule uma luminosidade de 1 fb^{-1} até ao final de 2011, abrindo a primeira janela experimental sobre a Física das interações fundamentais à escala do TeV. O solenóide compacto de muões (Compact Muon Solenoid, CMS) é um detector de carácter generalista a operar no LHC. O calorímetro electromagnético de CMS (Electromagnetic Calorimeter, ECAL) foi dimensionado de forma a providenciar uma excelente resolução espacial e em energia na detecção de electrões e fotões, contando com 76000 cristais de $PbWO_4$.

Esta tese descreve as contribuições originais do autor na preparação de CMS e análise dos primeiros dados de colisões no LHC.
O software dos sistemas de disparo e de aquisição de dados do ECAL de CMS são apresentados e a sua performance discutida, assim como a performance do sistema de disparo do ECAL.
É apresentada a primeira medida da secção eficaz da produção inclusiva de fotões isolados utilizando dados do LHC em 2010. Finalmente, é apresentado um estudo da sensibilidade de CMS à produção associada W-fotão e são discutidas as perspectivas para a medida das secções eficazes total e diferencial com os primeiros dados do LHC.

Palavras Chave

fotões isolados secção eficaz física electro-fraca acoplamento de gauge QCD CMS

Códigos PACS

14.70.Bh 14.70.Fm 07.05.Hd 29.40.Vj

Acknowledgements

Preparing a Ph.D. thesis is a unique experience for everyone who would like to undertake a scientific career. Having the task to solve a specific research problem in the most fine details is an occasion of personal and professional growth of great value. However, such a growth could not be achieved, and the problem be solved, without the essential support of many people. The author of this manuscript owes much to several of the people with whom he had the pleasure to interact during the development of the work presented herein.

First of all, the author would like to thank Prof. João Varela for giving him the possibility of working in such a stimulating environment such as the CMS collaboration and CERN. His accurate, enlightening, still practical advices and suggestions, pointing to the core of the problems, have been a greatly valued guidance.

The author's gratitude goes further to the LIP/CMS group. The greatest obligation is due to Dr. André David, for sharing with the author the most difficult part of each step of the work and for being an always available source of advices and an interlocutor in illuminating discussions. He would like to further thank Dr. Michele Gallinaro, for his constant support and his indefatigable constancy in keeping the route towards the final destination and motivating a steady progress of research. A special acknowledgement also goes to Dr. Pedro Silva and to Pedro Parracho for the helpful and constructive discussions and for the pleasant team work developed through the period of this work. The author would also like to thank other people at LIP for their help and support: José Carlos Da Silva, Dr. Pedro Martins, Marcelo Jordão and Dr. Pedro Ribeiro.

The support of LIP is gratefully acknowledged and a special thank you goes to the LIP secretariat and, in particular, to João Pedro Santos for his help in many administrative matters.

The author would also like to thank all people in the CMS ECAL team and in the electroweak, e/gamma and QCD Physics groups with whom he had the pleasure to interact. By mere chronological order: Dr. Tiziano Camporesi, Dr. Reyes Alemany, Dr. Alessio Ghezzi, Dr. Giovanni Franzoni, Dr. Evgueni Valssov, Dr. Paolo Rumerio, Dr. Pascal Paganini, Dr. Alessandro Thea, Dr. Nicoló Cartiglia, Dr. Vuko Brigljevic, Dr. Marco Pieri, Dr. Paolo Meridiani, Prof. Shin-Shan Yu and Dr. Serguei Ganjour.

Looking further back in time, the author would also like to thank Prof. Tommaso Tabarelli De Fatis and Prof. Stefano Ragazzi, for their precious teachings and for providing him with the first contact with Physics research. A word of gratitude also goes to all people with whom the author had the pleasure to share the passion for Physics while he was an undergraduate student and, in particular, to Dr. Simone Alioli, Stefano Marelli and Filippo Resnati.

To conclude, the author thanks deeply his parents Salvatore and Graziella, who gave him the possibility of pursuing such high profile studies, as well as his brothers Giuseppe and Orsola for their support. Lastly, and most importantly, he wishes to express his great gratitude to his life companion Marta, who shared with him all the high and low moments that he went through during the development of this work, giving him the best support a man could wish.

The work described hereafter was supported by FCT through the scholarship SFRH/BD/30536/2006.

Contents

Chapter 1

Introduction

This manuscript is being written at a very exciting time for Physics. After almost two decades of design and construction, the Large Hadron Collider (LHC) at CERN has started its research program. Many of the open questions in fundamental Physics might receive important answers over the next few years. The work presented in this manuscript has been carried out in a unique period for the experimental particle Physics community. It started in the last phases of construction of the Compact Muon Solenoid (CMS) detector, it was developed during the phase of detector commissioning and preparation for data taking and it was concluded roughly one year after the first proton-proton collisions took place at the LHC. The author considers it a privilege to have been able to develop his Ph.D. work during such rich and intense times.

This thesis aims at describing the original contribution of the author to the preparation of the CMS experiment and to the analysis of the first collision data. Being able to provide an original and significant contribution to a collective enterprise as large as the CMS experiment, which totals more than 3000 collaborators, is a challenging task. This text documents the work carried out by the author during his Ph.D. program in the context of the CMS collaboration. By the end of the manuscript it should be clear whether the author succeeded in the task.

The route to the final answer passes through all the following chapters. The starting point is to define the "why" and the "how", to set a clear stage to the evaluation of the "what". The profound reason why Mankind constructed an apparatus like the LHC probably resides in the innate human inclination to search for knowledge and progress. This certainly goes beyond the scope of this manuscript, where it will be considered the more modest question of what answers the scientific community expects to be provided by the experiments at the LHC. The goal of Chapter 2 is precisely to address this question. The intimately related question on the "how" is the subject of Chapter 3, devoted to the description of the LHC machine and the CMS detector.

The subsequent chapters are dedicated to the "what", explaining in detail the work carried out and the results achieved. The latter began at the end of 2006: the CMS Electromagnetic Calorimeter (ECAL) construction was ending and the installation in the CMS experimental cavern was about to begin. In this phase of the experiment, the author contributed to the development of the ECAL Trigger and Data Acquisition system and in particular of the distributed software in charge of the configuration and monitoring of the detector. Throughout 2006 and 2007 the CMS detector was installed and commissioned and the author contributed to this effort which culminated in the 2008 CMS commissioning run with cosmic rays. From this run, the CMS collaboration published more than 20 articles on scientific journals describing the detector performance. Out of those, the author gave a direct contribution to the measurement of the CMS trigger performance.

The work developed in this context is describe in Chapters 4 and 5 and results from it were published in scientific journals and presented at international conferences:

- P. Musella et al., "The CMS Electromagnetic Calorimeter Data Acquisition System at the 2006 Test Beam",. Presented at 15th IEEE Real Time Conference 2007 (RT 07), Batavia, Illinois, 29 Apr - 4 May 2007

- P. Musella et al., "The CMS ECAL database services for detector control and monitoring", *J. Phys. Conf. Ser.* **219** (2010) 022016. doi:10.1088/1742-6596/219/2/022016

- P. Musella, "The ECAL online software in the commissioning of the CMS detector", *Nucl. Instr. Meth.* **A617** (2010), no. 1-3, 303 – 305. Proceedings of the 11th Pisa Meeting on Advanced Detectors. doi:10.1016/j.nima.2009.07.102

- CMS Collaboration, "Performance of the CMS Level-1 Trigger during Commissioning with Cosmic Ray Muons", *JINST* **5** (2010) T03002, arXiv:0911.5422. doi:10.1088/1748-0221/5/03/T03002

In the meanwhile, since 2008, the author started a feasibility study aimed at assessing the sensitivity of the CMS detector to associated Wγ production. The study, described in Chapter 8 eventually established the feasibility of the measurement of the integrated and differential production cross sections with early LHC data. In the scope of this measurement a technique aimed at statistically separate isolated photons from jet background was developed.
Since 2009, the work started to focus on the first collision data from the LHC. The author contributed to the development and commissioning of photon reconstruction and identification algorithms. The photon reconstruction algorithms were tested using LHC collision data from 2009 and 2010. The results are presented in Chapters 5 and 6 and were published by the CMS collaboration in the following reports:

- CMS Collaboration, P. Musella et al., "Electromagnetic physics objects commissioning with first LHC data". CMS-PAS-EGM-10-001, 2010
- CMS Collaboration, P. Musella et al., "Photon reconstruction and identification at $\sqrt{s} = 7$ TeV". CMS-PAS-EGM-10-005, 2010

The work has been presented in international conferences, such as:

- P. Musella, "Performance of the CMS Electromagnetic Calorimeter and first results on electromagnetic physics objects". Presented at the Physics at the LHC Conference 2010, 7-12 June 2010, DESY, Hamburg (Germany)

Finally, the author played a leading role in the first measurement of the differential isolated photon cross section in proton-proton collisions as $\sqrt{s} = 7$ TeV. The measurement was performed in the pseudorapidity region $|\eta| < 1.45$, using 2.9 pb^{-1} of pp collision data and covered the kinematic region $0.006 < x_{\mathrm{T}} = 2\ E_{\mathrm{T}}^{\gamma}/\sqrt{s} < 0.086$. This is the first measurement involving isolated final state photons performed at the LHC. The work is described in Chapter 7. The results presented here have been approved by the CMS collaboration and a scientific communication was submitted for publication on the journal *Physical Review Letters*:

- CMS Collaboration, "Measurement of the Isolated Prompt Photon Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV.", `arXiv:1012.0799`. CERN-PH-EP/2010-053

Chapter 2

The Physics topic

Contents

This chapter is devoted to the introduction of the Physics topics that will be developed in the following of this manuscript. First a brief introduction of the Standard Model (SM) of Particle Physics is given, the limitations of the theory are discussed and the Physics of proton-proton interaction is introduced. The last two sections are devoted to the discussion of the inclusive photon production and of the gauge bosons interactions.

2.1 The Standard Model of Particle Physics in brief

In the SM, the description of matter and its interactions is based on the concept of symmetry. The latter is an intuitive concept that can be experienced in everyday life and is related to that of invariance. An object is said to be symmetric if it is possible to perform some operation on it that brings it to a configuration equivalent to the starting one.
In Physics, the concept of symmetry is very powerful. The symmetries of a physical system are in fact connected to the conservation of physical quantities. In the language of classical field theory, this is expressed by Nother's theorem [9] which states that any continuous transformation that leaves the form of the Lagrangian of a physical system invariant corresponds to a conserved current. Symbolically[1]:

$$\begin{aligned} &\Phi \rightarrow \Phi + \Delta\Phi,\ \Delta\mathcal{L}[\Phi, \partial\Phi] = 0 \\ &\qquad\qquad \Rightarrow \partial_\mu\ j^\mu(x) = 0,\ j^\mu = \frac{\partial\mathcal{L}}{\partial\left(\partial_\mu\Phi\right)}\Delta\Phi \end{aligned} \tag{2.1}$$

where $\mathcal{L}$ is the Lagrangian density of the system (that will be referred to simply as Lagrangian in the following) and Φ symbolically refers to the fields' content of the system. The current $j^\mu(x)$ satisfies the continuity equation at all points of the space-time. The conservation law can also be expressed in terms of a charge which is constant in time:

$$Q = \int d^3x\ j^0(x),\quad \frac{d}{dt}\ Q = 0 \tag{2.2}$$

This is directly translated at the quantum level, where the conserved charges are represented by operators which commute with the Hamiltonian of the system.

The construction of the SM proceeds from the definition of the symmetry group of the theory. Matter fields are arranged in irreducible representations of the group. Experimentally, three generations of fermions (whose properties are reported in Table 2.1) with identical quantum numbers, but different masses are observed. Each of the three generations is arranged in representations of the symmetry group, which is:

$$G_{SM} = SU(3)_C \times SU(2)_L \times U(1)_Y \tag{2.3}$$

Three charges are associated to this symmetry group: the *colour*, linked to the $SU(3)_C$ sub-group; the *isospin*, linked to $SU(2)_L$; and the *hypercharge*, connected to $U(1)_Y$. The $SU(2)_L \times U(1)_Y$ subgroup is known as *Electroweak* subgroup. Each of the fermion generations is represented by the following five multiplets:

$$\Psi_f = \left(Q_L(3,2)_{1/6},\ U_R(3,1)_{2/3},\ D_R(3,1)_{-1/3},\ L(1,2)_{-1/2},\ E_R(1,1)_{-1}\ \right) \tag{2.4}$$

The components of Ψ_f are chiral spinor fields, labelled $F(C,L)_Y$, where C, L and Y indicate, respectively, the dimension of the $SU(3)_C$ and $SU(2)_L$ representation to which the field belongs and the associated

[1] Throughout this chapter the natural units system where $\hbar = c = 1$ is used. Sums over repeated indexes are implicit, Greek letters are employed for Lorentz indexes while Latin letters are used for internal symmetries.

Table 2.1: Summary of the properties of the SM fermions. The value of the masses are reproduced from [10].

Generation	Particle	Symbol	$SU(3)_C$ rep.	Q_{EM}	Mass	
Leptons						
1st	Electron	e	1	1	0.510998910	$\pm 1.3 \times 10^{-8}$ MeV/c^2
	El. neutrino	ν_e	1	0	< 2	eV/c^2
2nd	Muon	μ	1	1	105.658367	± 0.000004 MeV/c^2
	Muon neutrino	ν_μ	1	0	< 0.19	MeV/c^2
3rd	Tau	τ	1	1	1776.84	± 0.17 MeV/c^2
	Tau neutrino	ν_τ	1	0	< 18.2	MeV/c^2
Quarks						
1st	Up	u	3	2/3	2.55	$^{+0.75}_{-1.05}$ MeV/c^2
	Down	d	3	$-1/3$	5.04	$^{+0.96}_{-1.54}$ MeV/c^2
2nd	Charm	s	3	2/3	1.27	$^{+0.07}_{-0.11}$ GeV/c^2
	Strange	c	3	$-1/3$	105	$^{+25}_{-35}$ MeV/c^2
3rd	Top	t	3	2/3	171.3	$\pm 1.1 \pm 1.2$ GeV/c^2
	Bottom	b	3	$-1/3$	4.20	$^{+0.17}_{-0.07}$ GeV/c^2

value of the *hypercharge*. The Q_L multiplet contains the left-handed components of the quark fields, which transform according to the fundamental representation of both $SU(3)_C$ and $SU(2)_L$. U_R and D_R are the right-handed components of up- and down-type quarks, which belong to the fundamental representation of $SU(3)_C$ but are singlets under $SU(2)_L$. Similarly, L contains the left-handed components of the lepton fields and E_R the right-handed component of the charged leptons. All leptons are singlets under $SU(3)_C$, left-handed leptons are doublets under $SU(2)_L$, while right-handed leptons are singlets.

The symmetry group G_{SM} is intimately related to the way interactions between particles develop. Interactions are introduced by means of the requirement of *gauge invariance* of the theory under the symmetry group. Each of the generators of the group is associated to a vector field, called *gauge* field and the space-time derivative in the Dirac free Lagrangian is replaced with a *covariant derivative*. The fermionic part of the Lagrangian is therefore written as:

$$\mathcal{L}_{\text{Fermion}} = i\bar{\Psi}_f \gamma^\mu \left(\partial_\mu + ig_s G^a_\mu T^a_f + igW^b_\mu L^b_f + ig' B_\mu Y_f\right) \Psi_f \tag{2.5}$$

Here, the G^a_μ ($a = 1..8$) represent the gluon fields, associated to the $SU(3)_C$ group; W^b_μ ($b = 1..3$) are the weak fields, associated to $SU(2)_L$; and B_μ is the field associated to $U(1)_Y$. T^a_f and L^b_f are the generators of the $SU(3)_C$ and $SU(2)_L$ to which the component Ψ_f belongs, and Y_f is the associated *hypercharge*. The constants g_S, g and g' represent the couplings associated with each of the subgroups. The expression (2.5) is invariant under the *gauge transformation*:

$$\Psi_f \to U\Psi_f,\ \mathcal{A}_\mu \to U(\mathcal{A}_\mu + \partial_\mu)U^{-1},\ U = \exp\left(\ i\theta^i(x)Q^i\ \right) \tag{2.6}$$

where $Q = \left(T^a, L^b, Y\right)$ and $\mathcal{A}_\mu = G^a_\mu T^a + W^b_\mu L^b + B_\mu$. Such transformation can be thought of as a local phase-rotation of the fields of the theory. The concept of *gauge invariance* is an essential ingredient for the construction of the SM as it constrains the most general form of the theory. In fact, more gauge-invariant terms beyond equation (2.5) can be added to the Lagrangian like, in particular, the one describing the

propagation of the gauge fields:

$$\mathcal{L}_{\text{Gauge}} = -\frac{1}{4}\left(G^a_{\mu\nu}G^{a\ \mu\nu} + W^b_{\mu\nu}W^{b\ \mu\nu} + B_{\mu\nu}B^{\mu\nu}\right) \tag{2.7}$$

where the gauge field-strength tensors were introduced:

$$\begin{aligned} G^a_{\mu\nu} &= \left(\partial_\mu G^a_\nu - \partial_\nu G^a_\mu\right) \ + if^{abc}G^b_\mu G^c_\nu \\ W^a_{\mu\nu} &= \left(\partial_\mu W^a_\nu - \partial_\nu W^a_\mu\right) + i\varepsilon^{abc}W^b_\mu W^c_\nu \\ B_{\mu\nu} &= \left(\partial_\mu B_\nu - \partial_\nu B_\mu\right) \end{aligned} \tag{2.8}$$

Here, f^{abc} are the structure constants of the $SU(3)$ group and ε^{abc} is the completely antisymmetric symbol in three dimensions.
Equations (2.7) and (2.8) will be examined later more carefully. For the moment it should just be noted that the presence of quadratic terms in the Lagrangian involving the gauge fields and their derivatives lead to vertexes involving three and four gauge bosons.

The symmetry G_{SM} is not directly observed in nature. For the Electroweak gauge subgroup, equation (2.7) describes the propagation of four massless vector fields, thus predicting the observation of four massless vector bosons. In fact, only one such massless state is seen in nature: the photon. The other three states, the W^+, W^- and Z_0 bosons, are instead massive.
In the SM, this observation is accommodated by postulating that the G_{SM} symmetry is spontaneously broken. This is achieved by introducing an additional scalar field in the theory with that has a non-vanishing vacuum expectation value (VEV). A suitable choice of the quantum numbers of this field reproduces well the experimental observations: the $SU(2)_L \times U(1)_Y$ group is reduced to the electromagnetic gauge symmetry $U(1)_{EM}$, associated to the photon, and at the same time the remaining three vector fields acquire a mass. The fields corresponding to the observed particles, $(W^+_\mu, W^-_\mu, Z_\mu, A_\mu)$ are obtained from the canonical base through the following relations:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right), \quad \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \cdot \begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix} \tag{2.9}$$

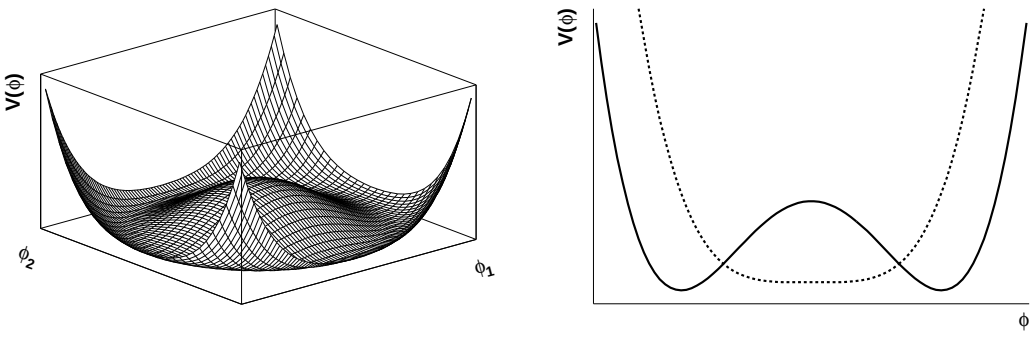


Figure 2.1: Illustration of the Higgs potential. On the left, the potential for a doublet of scalar fields is plotted as a function of the amplitudes of the fields. The one-dimensional set of minima can be seen. On the right, a unidiemensional projection is shown. The solid curve shows the behaviour of the potential for a non-vanishing value of μ, that gives raise to EWSB. The dashed curve shows the case of a vanishing value of μ, that does not break the Electroweak symmetry.

Table 2.2: Summary of the properties of the SM gauge bosons. The value of the masses are reproduced from [10]. In the case of the gluon, the value is theoretical. Experimentally, a gluon mass of few MeV/c^2 cannot be excluded.

Particle	Symbol	$SU(3)_C$ rep.	Q_{EM}		Mass
Photon	γ	1	$< 1 \times 10^{-35}$	< 1	$\times 10^{-18}$eV/c^2
W	$W^{\pm}$	1	1	80.398	± 0.025 GeV/c^2
Z	Z_0	1	0	91.1876	± 0.0021 GeV/c^2
Gluon	g	8	0	0	MeV/c^2

The linear combinations Z_μ and A_μ are associated to the Z_0 boson and the photon respectively and the $W_\mu^\pm$ fields to the charged W bosons. The parameter defining the mixing of the (W_μ^3, B_μ) system, θ_W, is known as Weinberg angle or weak mixing angle.
The mechanism of spontaneous symmetry breaking introduces precise relations between the Weinberg angle, the Electroweak couplings, the masses of the weak bosons and the VEV of the scalar field. In the simplest case, the scalar field added to the theory - known as the Higgs field - transforms as $\phi(1,2)_{1/2}$. In other words, it is a $SU(2)_L$ doublet with hypercharge $Y = 1/2$. The most general renormalisable gauge-invariant Lagrangian involving this field is of the form:

$$\mathcal{L}_{\text{Higgs}} = \left| \left(\partial_\mu - igT_2^i W_\mu^i - i\frac{g'}{2} \right) \phi \right|^2 + V(\phi),\ V(\phi) = \frac{\mu^2}{2}|\phi| + \frac{\lambda}{4}|\phi|^4 \tag{2.10}$$

For $\mu^2 < 0$ and $\lambda > 0$, the field acquires a non-null VEV $|\phi_0|^2 = -\frac{\mu^2}{\lambda} = v^2$ on an infinite number of configurations. The Higgs field can than be reparametrised as:

$$\phi = \frac{1}{\sqrt{2}} \exp\left(igT_2^i \theta^i(x)\right) \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.11}$$

The Electroweak symmetry group does not vanish completely: the vacuum is invariant under the Abelian gauge group generated by $Q_{EM} = T^3 + 2Y$, which can be identified with the electromagnetic gauge group $U(1)_{EM}$. Inserting the expression (2.11) into (2.10), and choosing $\sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}}$ the Higgs Lagrangian takes the form[2]:

$$\mathcal{L}_{\text{Higgs}} = \frac{1}{2}\partial_\mu H \partial^\mu H + \left(\frac{1}{4} g^2 W^{+\mu} W_\mu^- + \frac{g^2}{8\cos^2\theta_W} Z^\mu Z_\mu \right)(v+H)^2 - \frac{\mu^2}{2}(v+H)^2 + \frac{\lambda}{4}(v+H)^4 \tag{2.12}$$

The second term in equation (2.12) contains a mass term for the W and Z bosons which is proportional to the Higgs VEV, in particular the symmetry breaking mechanism predicts:

$$\begin{aligned} M_W &= \frac{1}{2} g v \\ M_Z &= \frac{1}{2} \frac{g}{\cos\theta_W} v \\ \rho &= \frac{m_W^2}{m_Z^2 \cos^2\theta_W} = 1 \end{aligned} \tag{2.13}$$

The measured values of the W and Z bosons masses are $M_W = 80.398$ GeV and $M_Z = 91.1876$ GeV, which yields $\sin\theta_W = 0.223$ [10]. The value of the Higgs VEV can be estimated considering that in the low

[2] in order to obtain equation (2.12), the exponent in (2.11) needs to be eliminated, by means of a gauge transformation. This form of the equations is known as the unitarity gauge.

energy limit the Electroweak interaction reduces to the Fermi 4-fermions interaction. This implies that $v^2 = 1/\sqrt{2}G_F$, where G_F is the Fermi constant, that can be measured in the muon decay. Experimentally, $G_F = 1.166367(5) \times 10^{-5}$ GeV^{-2}, thus $v = 245$ GeV. As it will be discussed in the following, the relations between M_W, M_Z and v have been verified experimentally to high precision using data from collider and low energy experiments [10, 11].
Through the Electroweak symmetry breaking mechanism (EWSB), part of the Higgs field degrees of freedom apparently disappear and are substituted by the longitudinal polarisation of the W^μ and Z^μ fields, thus allowing them to acquire a mass. However, a scalar field H still remains in the theory. From equation (2.12), the mass of the associated particle (Higgs boson) is given by $M_H^2 = 2\lambda v^2$. The EWSB mechanism thus predicts the observation of a spin-0 particle, coupling to fermions and weak bosons. The theory does not predict the mass of the particle, that remains unobserved while this manuscript is being written.

The EWSB describes the observed spectrum of gauge bosons. In fact, the same mechanism can also be exploited to generate the masses of the fermions. This is achieved by introducing a Yukawa contact interaction between the fermions and the Higgs field:

$$\mathcal{L}_{\text{Yukawa}} = -Y^d_{ij}\bar{Q}_{L_i}\phi D_{R_j} - Y^u_{ij}\bar{Q}_{L_i}\bar{\phi}D_{U_j} - Y^e_{ij}\bar{L}_i\phi E_j + h.c. \tag{2.14}$$

Where the matrices $Y^{d,u,e}_{ij}$ contain the couplings of the Yukawa interactions. After the introduction of equation (2.11), these interactions give raise to the mass matrices:

$$M^u = \frac{v}{\sqrt{2}}Y^u,\ M^d = \frac{v}{\sqrt{2}}Y^d,\ M^e = \frac{v}{\sqrt{2}}Y^e, \tag{2.15}$$

It can be shown that it is always possible to find unitary transformations that diagonalise the mass matrices, $V_L M V_R^\dagger = \frac{v}{\sqrt{2}}\text{diag}\{\lambda_i\}$. Without loss of generality, one can choose $V_L^d = V_R^d = V_R^u = V_L^e = V_R^e = \mathbf{1}$ and $V_L^u = V_{CKM}$. These transformations have no effect on the neutral current interactions associated to the Z_μ and A_μ fields, but in the quark sector it modifies the charged current interactions mediated by the W_μ fields inducing terms that mix the quark generations and violated the CP discrete symmetry. The quark mixing matrix V_{CKM} is known as Cabibbo-Kobayashi-Maskawa matrix.

Equations (2.5), (2.7), (2.10) and (2.14) summarise the content of the Standard Model of Particle Physics. The properties of the observed particles in the theory are reported in Tables 2.1 and 2.2.

2.2 Physics beyond the Standard Model

The Standard Model is a mathematically consistent renormalisable quantum field theory, which is in agreement with all the available laboratory observations. The formulation presented in the previous section assumes that neutrino are massless, however the theory can be extended [12] to account for the evidence of massive neutrinos provided by the observed oscillations patterns (see the review of neutrino Physics in [10]). With this addition, the SM can be considered the correct approximation of the description of matter and its interactions down to 10^{-16} cm, modulo the unobserved Higgs boson and additional weakly interacting particles. Once combined with the theory of gravitation, the theory would account for all the available observations and experiments, except for the so-called dark energy and dark matter, which emerge from astrophysical observations [13, and references therein].
The predictions of the theory have been verified experimentally through an extensive campaign of high precision measurements. The result of this effort, which includes results from low energy experiments, neutrino experiments as well as collider experiments (carried out at SLAC, LEP and the Tevatron) verified the predictions of the SM with a 0.1% precision [11, 10]. All SM predictions can be parametrised in terms of 6 quantities, which can be chosen to be $\alpha_s, \alpha_{EM}, G_F, M_Z, M_H, m_t$. A simultaneous fit to more than 40 different parameters measured in low energy and collider experiments yields an excellent

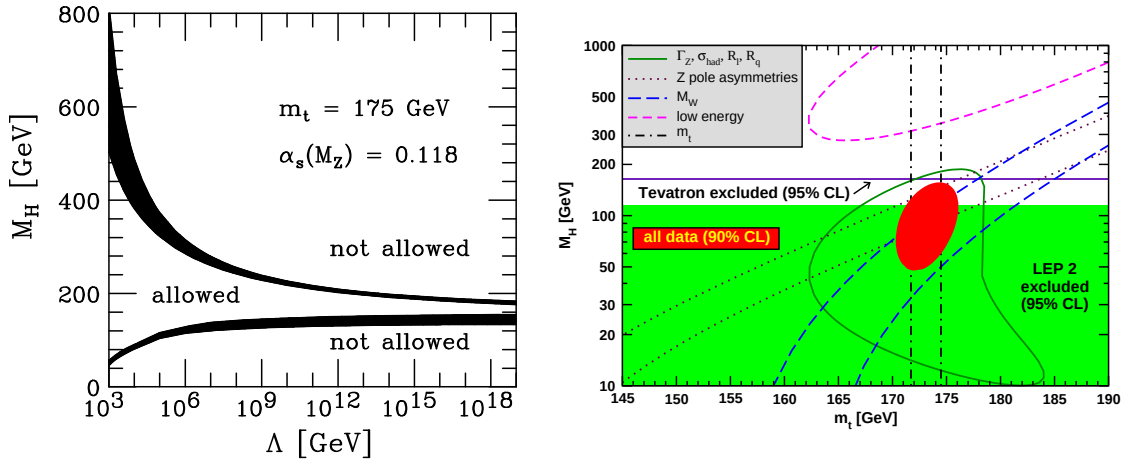


Figure 2.2: (left) Theoretical limits on the SM Higgs boson mass. The upper limit represents the *triviality* bound, while the lower limit comes from the requirement of vacuum stability. Figure reproduced from [14]. (right) Experimental determination of the SM Higgs mass. The bounds from precision measurements and direct exclusion limits are shown. Figure reproduced from [10].

agreement of experimental data with the predictions. To-date, the most important missing experimental confirmation of the theory is the observation of the Higgs boson, associated with the EWSB. Using the precision data, the value of M_H can be predicted through the dependence of radiative corrections on the mass. The combined fit yields a value of $M_H = 77^{+28}_{-22}$ GeV, while the lower limit from direct searches is $M_H > 114$ GeV. Even though there is no conflict, due to the large uncertainty of the predictions, the central value lies in the excluded region, as depicted in Figure 2.2 (right).

Despite the great experimental success, the theory is not believed to be fundamental. Some of the unsatisfactory aspects (an extensive review of such issues can be found, e.g., in [15]) are the large number of free parameters (20 in the minimal case, 27 or 29 if neutrino masses are allowed), the unexplained chiral asymmetry of the gauge group and charge quantisation. Also, the existence of three identical families of fermions and the very large differences between the masses of different families has no explanation within the SM. Other problems are related to the fact that the CP violation mechanisms predicted by the theory (and observed in the laboratory) are not sufficient to explain the large baryon asymmetry observed in the Universe. Furthermore, some aspects of the theory require a very precise fine-tuning of the theory parameters in order to account for the observations and make the theory self-consistent. One such aspect is the so-called hierarchy problem, which arises in the context of radiative corrections to the Higgs mass. The self-consistency of the model requires the Higgs mass to be of the order of the W an Z masses. On the other hand, in the presence of additional degrees of freedom, M_H receives radiative corrections proportional to the scale of the new Physics. As a result, a very precise tuning (up 30 decimal places) of these corrections is needed in order to make the Higgs mass of the order of the Electroweak scale. Other examples of fine-tuning regard the unobserved CP violation in the strong sector and the so-called cosmological constant problem, where the Higgs vacuum energy contribution has to be precisely cancelled by some counter-term up to 50 decimal places.

For the reasons mentioned above, it is widely believed that new Physics not described by the Standard Model should exist. Surely new interactions are expected to take place at the so-called Plank mass, the scale where the gravitational interaction becomes strong $\Lambda_{Plank} = G_N^{-1} \sim 10^{19}$ GeV. However, a few arguments suggest that new phenomena might be observed at energies closer to the Electroweak scale. Some of these arguments are related to cosmological observations, where the evidence for weakly interacting matter might suggest the existence of stable particles with masses in the range $10^2 - 10^3$ GeV.

Several SM extensions, such as supersymmetric theories (SUSY) or theories predicting additional space-time dimensions, provide candidates for this new type of particles.
Another approach is to analyse the SM and ask if there is a scale at which the theory becomes inconsistent. An indication can be found examining the behaviour of the theory for different values of M_H. Even though the SM does not predict the value of the parameter λ, theoretical considerations limit the range in which the theory is self-consistent. The *triviality* bound provides an upper limit on M_H: examining the running of λ for $M_H > 350$ GeV, a Landau pole is observed for large values of the interaction scale. It is natural to avoid this problem, requiring the theory to be valid up to a scale Λ below the pole. This gives $M_H < \frac{2}{3\pi^2}\frac{2v^2}{\log \Lambda/v}$. A second bound on the value of M_H is obtained from the requirement that $V(\phi)$ is bounded from below: this constraint give $M_H > \frac{3m_t^4}{\pi^2 v^2}\log \Lambda/v$. The allowed M_H region as a function of Λ is shown in Figure 2.2 (left). There it can be seen that for $M_H \sim 130 - 180$ GeV, the SM remains self-consistent up to the Plank scale, while higher or lower values would indicate lower scales at which new Physics is expected.
Additional information can be retrieved from the study of elastic W^+W^- scattering. In the SM, the amplitude for the scattering of the longitudinal component is, in the limit $s, M_H >> M_W$:

$$M = -i\sqrt{2}G_F M_H^2 \left(\frac{s}{s - M_H^2} + \frac{t}{t - M_H^2}\right) \tag{2.16}$$

It is evident that for $M_H \to \infty$, which is equivalent to removing the Higgs from the theory, the scattering amplitude grows proportionally to the square of the centre-of-mass energy. This behaviour would violate the unitarity constraint, which translates the conservation of the total current in a scattering process and can be expressed by the requirement that the no amplitude in a partial wave expansion of the cross section exceeds its kinematic limit. Since the latter goes as $1/s$, it is clear that no scattering amplitudes can indefinitely grow with energy. In the absence of any Higgs field, unitarity is violated for $\sqrt{2}\ \sqrt{32\pi v} > 2.4\ TeV$. In the presence of the Higgs field, the unitarity requirement constrains M_H:

$$M = M_H < \sqrt{\frac{8\pi}{3}}v \sim 700 \text{ GeV} \tag{2.17}$$

The above limit can be interpreted as the condition on M_H that makes the Higgs and gauge sector of the SM a weakly coupled theory. If the above condition would not be satisfied, it would imply that the EWSB is due to a strongly coupled Higgs sector or some other strongly interacting degrees of freedom. In both cases, an enhancement of the WW cross section or the observation of WW bound states is expected for $\sqrt{s} \sim \mathcal{O}(\text{TeV})$.

2.3 Quantum Chromo-Dynamics and pp interactions

In section 2.1, the Standard Model of Particle Physics was introduced. The main focus was put on the Electroweak sector of the theory. Since the subject of this thesis concerns the Physics of the LHC, it is very important to also briefly discuss the strong sector of the theory.
In the SM, strong interactions are governed by the $SU(3)_C$ gauge group. The resulting theory [16], known as Quantum Chromo-Dynamics (QCD), provides a very successful model of strong interaction phenomena, such as the properties of hadronic matter, Deep Inelastic Scattering experiments and hadronic matter production in collider experiments.
The experimental evidence shows that hadronic matter is made by strongly interacting particles, generically called partons, that can be identified with the quarks and gluons of QCD. In order to study the elementary interactions between partons in hadronic collisions, it is mandatory to understand the structure of the hadrons and the influence of the binding forces on the state of their constituents. QCD provides the tools to tackle this problem.

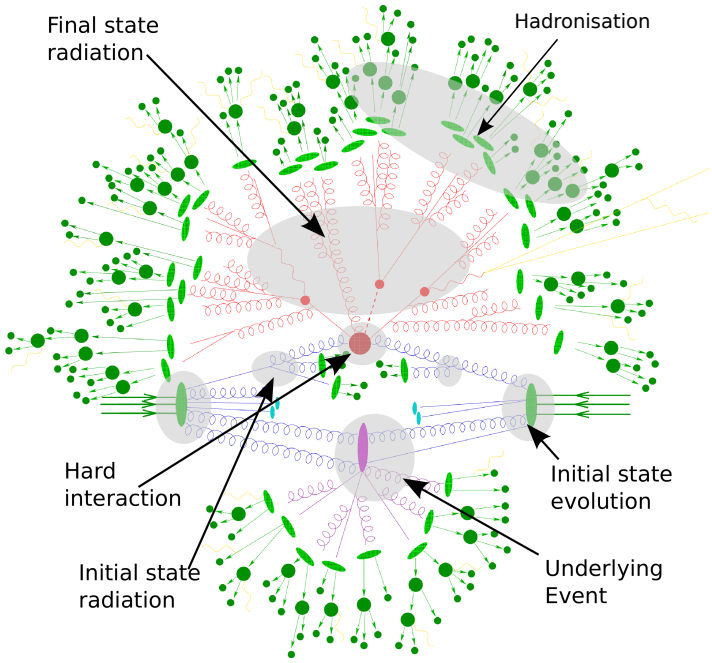


Figure 2.3: Schematic representation of an hadronic collision, illustrating the processes contributing to the system evolution.

An important property of QCD, which determines much of the phenomenology of strong interactions is the so-called *asymptotic freedom*: the evolution of the running coupling constant of the theory is such that the latter is weakly coupled at high energy, while it becomes strongly coupled at low energy. At leading order (LO) in perturbation theory, the renormalisation group equation for the coupling constant yields:

$$\alpha_s(Q^2) = \frac{g_s^2(Q^2)}{\pi} = \frac{1}{b_0 \log(Q^2/\Lambda_s^2)} \tag{2.18}$$

Here Q is the energy scale at which the process is calculated, b_0 is a numerical coefficient that depends on the number of quark flavours in the theory, while Λ_s is a constant with the dimensions of an energy, experimentally determined to be of the order of 200 MeV. The presence of this scale signals the fact that the theory is strongly coupled at low energy where it cannot be treated perturbatively. It is worth noticing that all the low-energy phenomenology of strong interactions is characterised by this scale: typical hadronic masses are of the order of Λ_S, hadrons' lifetimes are $\sim 1/\Lambda_s$ and typical cross sections of the order of $1/\Lambda_S^2$. On the other hand, since at high energy the coupling becomes weak, hadrons' constituents can be though to behave as free particles at short distances.

Using asymptotic freedom as an heuristic guide, it is possible to identify four phases in an hadronic collisions, each involving different energy scales. The overall picture is summarised in figure 2.3. For definiteness, the specific case of proton-proton collisions will be considered. In the first phase of the interaction, partons are bound inside the initial state protons, the typical scale of the interaction is of the order of the proton mass m_p and non-perturbative effects dominate. In the second phase, two partons from the incoming hadrons interact at a scale $Q >> m_p$: in this regime partons behave as free particles and the interaction can be calculated using perturbative QCD (pQCD). In a third phase, characterised by a scale $Q > \mu > \Lambda_s$, partons emitted by the hard interaction themselves emit radiation. Finally, as

the scale of the interaction becomes comparable with Λ_s, non-perturbative Physics starts to dominate and the hadronisation process takes over. This intuitive picture can be stated quantitatively through the factorisation theorem. The cross section for the production of a final state X in a pp collision can be calculated as:

$$\sigma(pp \to X; p_1, p_2) = \sum_{ij} \int dx_1 dx_2 \; f_i^p(x_1, \mu_F) \; f_i^p(x_2, \mu_F) \; \hat{\sigma}_{ij}(x_1 p_1, x_2 p_2, \hat{x}, \mu_F^2, \mu_f^2) \; D(\hat{x}, X, \mu_f^2) \quad (2.19)$$

where:

- the scales μ_F and μ_f represent the scales at which the hard process, calculated with pQCD, is separated from the initial and final state evolutions respectively;
- the sum over i and j extends over the partons of the incoming protons;
- x_1 and x_2 are the fractions of the proton momenta carried by the colliding partons;
- the functions $f_i^p(x, \mu_F)$ are called Parton Distribution Functions (PDF) and represent the probability that a parton i carries a fraction x of the proton momentum at a scale μ_F;
- the quantity $\sigma_{ij}(x_1 p_1, x_2 p_2, \hat{x}, \mu_F^2, \mu_f^2)$ represents the partonic-level cross section for the production of a state $\hat{x}$ from a pair of partons (i, j) ;
- the function $D(\hat{x}, X, \mu_f^2)$ represent the probability that the partonic state $\hat{x}$ evolves into the final state X. These type of functions are known as fragmentation functions.

Using expression (2.19) it is possible to clearly factorise the perturbative and non-perturbative part when computing the properties of a physical process involving hadrons. Due to the complexity of the calculations, Monte Carlo (MC) simulation programs are often employed. Several MC codes are available, which can perform the different steps of QCD calculations with different accuracies. Two main groups of MC programs can be identified. The first one comprises the so-called *shower* MC programs, that perform the full calculation in equation (2.19), yielding realistic events where the final state is populated with the result of the hadronisation process. Examples of this first group are the PYTHIA [17], HERWIG [18] and SHERPA [19] event generators. These programs are ubiquitously used in High Energy Physics, because they provide a detailed description of the whole production process. However, the accuracy of the calculation that is implemented in these programs usually doesn't exceed LO in pQCD. New shower MC programs are being developed in order to overcome this limitation and provide more accurate calculations [20]. A second category of MC generators comprises the so-called *integrators*, which perform the computations up to the outgoing partons, without calculating the hadronisation part. Even if in these case the description of the final states is less detailed, this kind of programs implements more accurate calculations, usually up to Next to Leading Order (NLO) in pQCD, or better.

The non-perturbative effects related to hadronic structure are parametrised by the PDFs. Even though such functions cannot be calculated using pQCD, equation 2.19 implies that they do not depend on the particular process under consideration; it is therefore possible to measure them in dedicated experiments and then use the same distributions in the computation of other processes. Also, pQCD provides the tools to calculate $f^p(x, Q)$ provided that $f^p(x, Q_0)$ is known. This is achieved through the DGLAP evolution equations [22]. From the practical point of view, PDFs are mainly obtained from the fit to data from Deep Inelastic Scattering experiments. Few analyses can be found in the literature which use slightly different parametrisations. Some examples are the CTEQ6.6 [23], MRSTW08 [24](shown in Figure 2.4) and HERAPDF [25] sets of PDFs.
The partons not involved in the hard interaction constitute the so-called underlying event. In the MC generators, the underlying event is modelled based on the hadron multiplicities and transverse momentum

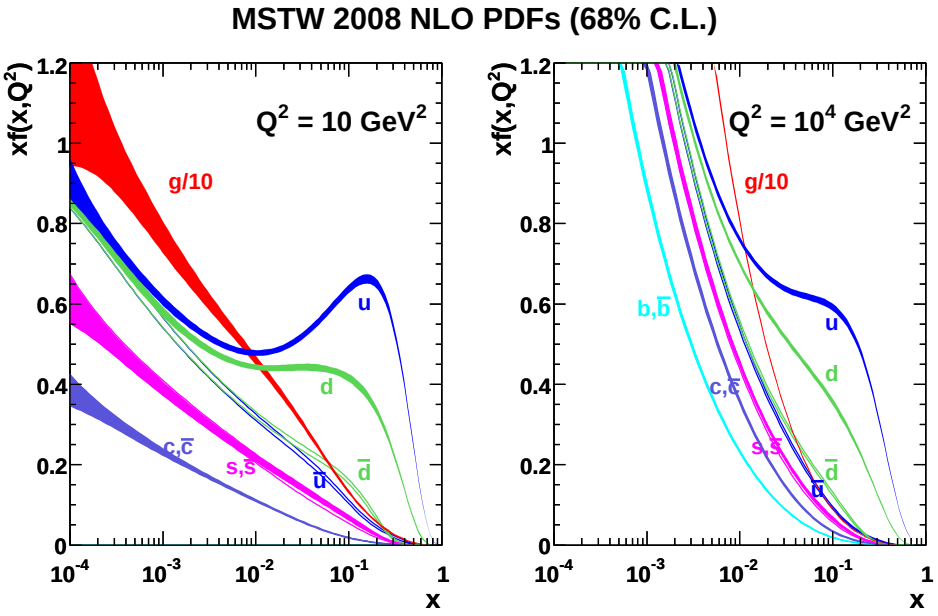


Figure 2.4: Proton Parton Distribution Functions. (left) Proton PDF calculated for a factorisation scale $Q^2 = 10$ GeV^2 (right) Same for $Q^2 = 10^4$ GeV^2. Figure reproduced from [21].

spectra measured in soft inelastic hadron collisions (also known as minimum bias events). The description of this part of the hadronic interactions relies on phenomenological models, which need to be tuned on experimental data [26].
Once the hard process calculation has been performed and the final state radiation included, a final non-perturbative step, usually called *hadronisation*, is necessary in order to map the partonic final state into an hadronic final state that can be experimentally measured. Two different approaches can be followed in this step, depending on the quantity that is being calculated. In the case of the calculation of the inclusive production of a particular final state, a set of fragmentation functions analogous to the PDF can be used. Such functions cannot be calculated in pQCD, but their evolution follows the DGLAP equations. In the case of high multiplicity final states, however, this approach is not viable and MC simulations use phenomenological models whose free parameters are fitted to experimental data. The two most used models are the so-called *cluster model* and the *string model*.

2.4 Inclusive γ production in pp collisions

It was seen in the previous section that the understanding of Physics of hadronic collisions heavily relies on QCD calculations. It was also noted that such calculations need experimental inputs to parametrise the non-perturbative effects, such as the initial state evolution of the system. It this thus clear that testing the pQCD predictions and providing constraints to the non-perturbative evolution parametrisation is essential for the understanding of Physics at the LHC.
This section is dedicated to the discussion of the Physics of inclusive photon production. The study of photon production at hadron colliders is motivated by several reasons. First of all, the process provides a stringent test of pQCD predictions over several order of magnitudes. The use of photons as probes of the hard interactions has several advantages over purely hadronic states: photons do not hadronise so their state is not altered by the final state QCD dynamics; also, the performance of detectors in determining

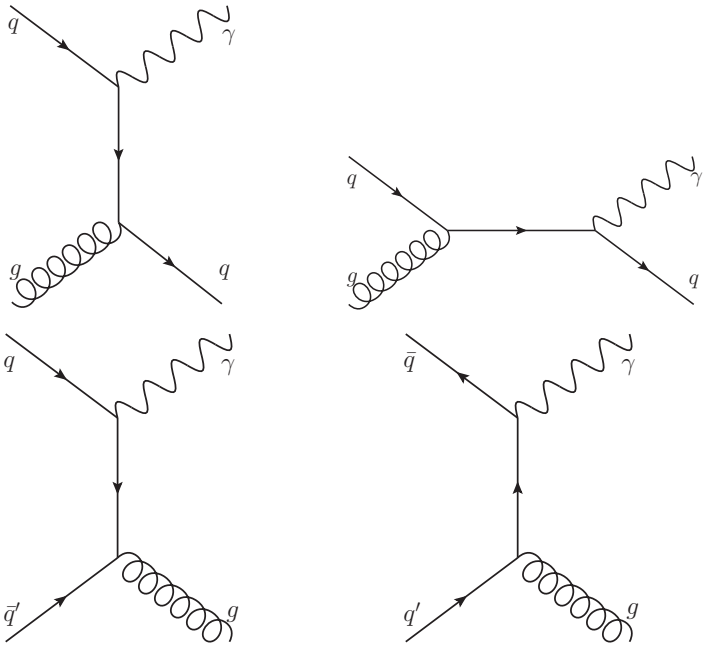


Figure 2.5: Feynman diagrams contributing to direct photon production at Leading Order: (top) QCD Compton scattering. (bottom) Quark-antiquark annihilation.

the position and energy is better for photons than for jets.
The measurement of the differential photon production cross section has the potential of constraining the gluon PDF. Further insight on the proton structure can also be obtained through the study of photon/jet correlations or photon production in association with a particular jet flavour [27].
A further motivation for the study of photon production is the fact that it represents an irreducible background for many searches, such as $H \to \gamma\gamma$, SUSY and extra-dimensions.

At LO in perturbative QCD, prompt photons are produced through the quark-gluon Compton scattering $qg \to q\gamma$ and through quark pair annihilation $q\bar{q}' \to g\gamma$. The corresponding Feynman diagrams are displayed in Figure 2.5. An additional contribution to inclusive photon production arises from the fragmentation of coloured partons from other QCD processes. This second production mechanism can in fact be considered as the re-summation of higher order corrections to the direct production and parametrised in terms of a set of fragmentation functions analogous to those introduced in equation (2.19) for hadronic final states. Symbolically, the differential photon production cross section can be written as [28]:

$$\frac{d\sigma}{dp_T^\gamma d\eta^\gamma} = \frac{d\sigma^D(\mu_R, \mu_F, \mu_f)}{dp_T^\gamma d\eta^\gamma} + \sum_{k=q,\bar{q},g} \frac{d\sigma_k^F(\mu_R, \mu_F, \mu_f)}{dp_T^k d\eta^k} \otimes D_{\gamma/k}(\mu_F) \tag{2.20}$$

Here, μ_R, μ_F and μ_f represent, respectively the renormalisation, factorisation and fragmentation scales, and the symbol $\otimes$ stands for the integral over the fraction z of the parton transverse momentum carried by the photon. The expression $d\sigma^D/dp_T d\eta$ describes the direct production mechanism, while $d\sigma^F/dp_T d\eta$ accounts for the production if coloured partons that fragment to a photon. The functions $D_{\gamma/k}$ can be thought of as the probability for an outgoing parton k to fragment into a photon. It should be noted that the distinction between direct and fragmentation contributions beyond the LO approximation is arbitrary and depends on the choice of the renormalisation scheme. This is the reason why equation (2.20) shows a dependence of $d\sigma^D/dp_T d\eta$ on the fragmentation scale. As already discussed in section 2.3, the

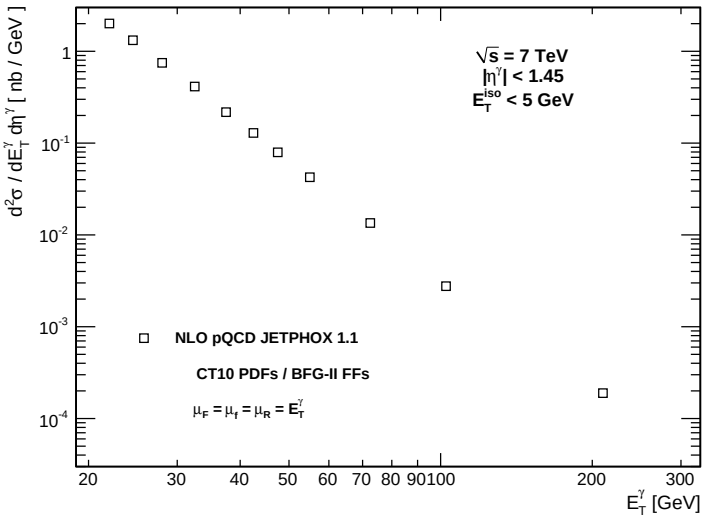


Figure 2.6: Prediction for isolated photon production cross section in the central ($|\eta| < 1.45$) rapidity region. The calculation was performed with the JETPHOX 1.1 [29] integrator using the CT10 PDFs [30] and the BFG set II of fragmentation functions (FFs) [31].

fragmentation functions cannot be calculated in pQCD and need to be measured in data. The photon fragmentation functions $D_{\gamma/k}$ are mostly determined from LEP data using the $e^+e^- \to qq\gamma$ process [32].

From the experimental point of view, the challenge of measuring the inclusive photon cross section is the identification of the photon signal, which has to be disentangled from the severe background due to the electromagnetic decay of light mesons such as the π^0 and the η. At high energies, the opening angle between the two photons from the meson decay is so small that an event-by-event discrimination solely based on the electromagnetic shower properties becomes very difficult. To reject this type of background, the photon candidates are usually requested to fulfil an isolation criterion, which limits the total amount of energy detected around the candidate.
In order to match the experimental constraints, recent NLO MC programs are able to calculate the isolated part of the photon production cross section. The implemented algorithms sum the hadronic transverse energy[3] E_T^{had} in a cone:

$$\mathcal{R} = \sqrt{(\phi_\gamma - \phi_{had})^2 + (\eta_\gamma - \eta_{had})^2} \tag{2.21}$$

The isolated photon cross section is then calculated as $\sigma(E_T^{had} < E_T^{max})$. Besides having the advantage of allowing an easy match between the theoretical calculations and the experimental measurements, the isolation requirement also reduces the sensitivity to the knowledge of the fragmentation functions. The integration over the photon momentum fraction is effectively limited to the phase space $z > (1 + E_T^{max}/E_T^\gamma)$.

Figure 2.6 shows the prediction for the differential cross section for inclusive photon production at the LHC, obtained with the JETPHOX NLO integrator [28]. The large production rate makes the observation possible already at the LHC start-up. Such characteristic makes the channel an ideal playground for the commissioning of experimental techniques required for photon reconstruction and identification.

[3] In this chapter a cylindrical coordinate system is used, with the z axis parallel to the direction of the incoming protons. The polar angle is denoted as θ, while the azimuthal is ϕ. The pseudorapidity is defined as $\eta = -\log\tan(\theta/2)$.

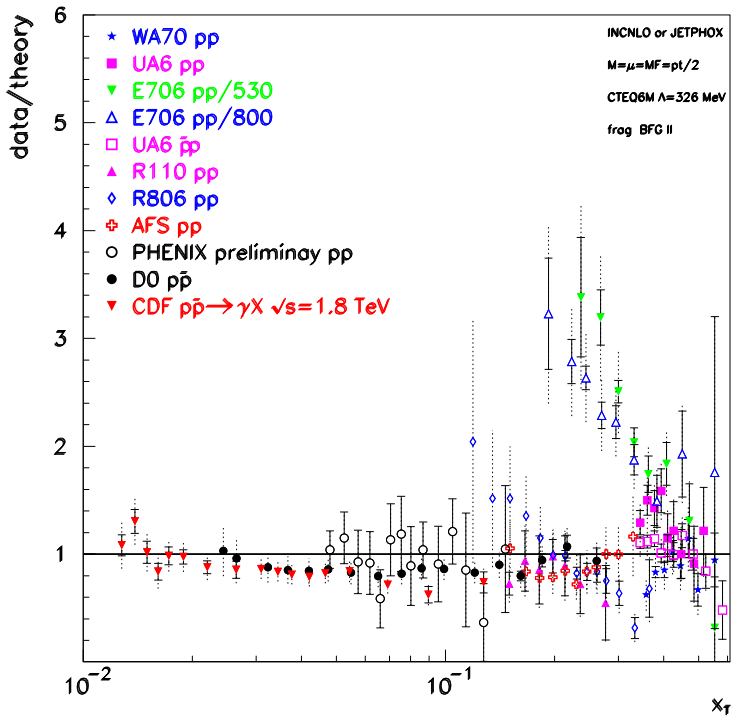


Figure 2.7: Comparison of photon production measurement to theoretical NLO predictions. The *data/theory* ratio is plotted as a function of the x_T variable, analogous to the Bjorken x variable in Deep Inelastic Scattering experiments. Figure reproduced from [35].

2.4.1 Previous measurements and new perspectives

Inclusive and isolated photon cross sections in hadronic collisions have been measured, over the last three decades, by several experiments at several different centre of mass energies. A comparison of the existing experimental measurements to the theoretical predictions is shown in Figure 2.7. Despite the great effort in improving the theoretical understanding of the process, the phenomenology of photon production in low energy experiments [33] still needs a firmer understanding. In particular, a debate on the size of non-perturbative effects (parametrised by the so called *intrinsic* k_T parameter) has been triggered in the 1990s by the measurement of the E706 experiment at the Tevatron, which data show large disagreement with respect to the NLO prediction[34].

More recent measurements [36, 35], taken at the Tevatron and RHIC colliders show an overall good agreement with theoretical predictions. Small increases in the lower energy part of the spectrum ($x_T = 2E_T^\gamma/\sqrt{s} \lesssim 0.02$) are observed by the D0 and CDF collaborations, even though data and theory agree within twice the total experimental uncertainty for all points.

The lack of understanding of low energy data determined the exclusion of photon production data from global PDF fits. Given the relatively large uncertainties in the determination of the gluon PDF in the LHC domain, the inclusion of further constraints, complementing deep inelastic scattering, Drell-Yan pair production and jet production measurements would be highly beneficial.

In particular [37], LHC data would allow to access the kinematic domain $x \sim 10^{-3} - 10^{-4}$, which is currently constrained by a relatively small number of experimental points. The kinematic domain accessible at the LHC is shown in Figure 2.8 and compared to the one probed by previous experiments. It can be seen that the LHC will be able to provide information on this kinematic region.

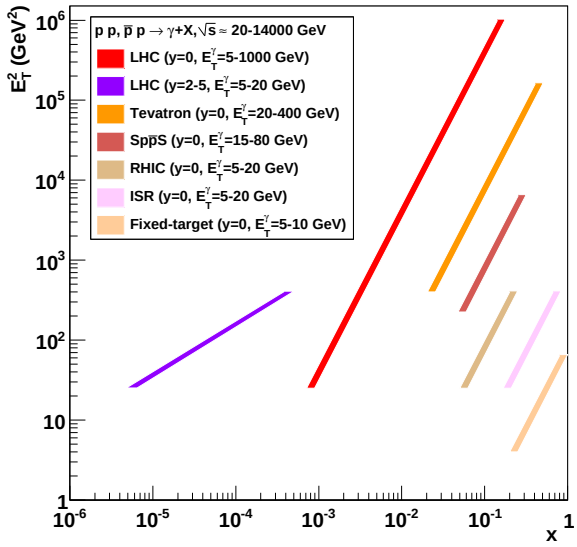


Figure 2.8: Kinematic domain in the (x, E_T^2) plane probed through the measurement of prompt photon production at the LHC, compared to lower energy experiments. The LHC curves correspond to a centre of mass energy of 14 TeV. Figure reproduced from [37].

2.5 Triple Gauge Couplings and associated Wγ production in pp collisions

The SM Lagrangian that describes the propagation of the gauge fields is given by equations (2.7) and (2.8). It was already noted that these expressions lead to vertexes involving three and four gauge bosons. Interactions of this kind are known as Triple and Quartic Gauge Couplings (TGC and QGC). In the SM, the form of these vertexes is completely determined by the structure of G_{SM}.

Experimentally, the structure of the Electroweak symmetry group[4] can be probed studying final states sensitive to triple and quartic gauge vertexes. Studies done at LEP [38] showed that the process $e^+e^- \rightarrow W^+W^-$, which receives contributions from WWZ and WWγ vertexes, nicely agrees with the predictions of an $SU(2)_L \times U(1)_Y$ theory. In proton-proton collisions, the study of associated di-boson production is sensitive to the TGC; in particular, the study of associated Wγ production in pp collisions is sensitive to the WWγ vertex.

Deviations of the TGC vertexes from the SM predictions would be a signal of new Physics, which could arise from loop corrections involving new particles or from unexpected internal structure (compositeness) of the gauge bosons. If new Physics occurs at a scale significantly larger than the one directly probed experimentally, its effects may become visible only through the measurement of anomalous interactions. In this scenario, the new degrees of freedom can be integrated out and their effects expressed in terms of a set of anomalous couplings. The most general Lagrangian giving rise to WWZ and WWγ interactions

[4] Triple and quartic vertexes are also present in the strong sector but they will not be discussed here.

can be written as [39]:

$$\begin{aligned}\frac{i\mathcal{L}_{WWV}}{g_{WWV}} &= g_V^1 V^\mu \left(W^-_{\mu\nu}W^{+\nu} - W^+_{\mu\nu}W^{-\nu}\right) \\ &+ \kappa_V W^+_\mu W^-_\nu V^{\mu\nu} + \frac{\lambda_V}{M_W^2} V^{\mu\nu} W^{+\rho}_\nu W^-_{\rho\mu} \\ &+ i g_V^4 W^+_\mu W^-_\nu \left(\partial^\mu V^\nu + \partial^\nu V^\mu\right) \\ &+ i g_V^5 \varepsilon_{\mu\nu\rho\sigma} \left[(\partial^\rho W^{-\mu})W^{+\nu} - W^{-\mu}(\partial^\rho W^{+\nu})\right] V^\sigma \\ &+ \frac{\tilde{\kappa}_V}{2} W^+_\mu W^-_\nu \varepsilon^{\mu\nu\rho\sigma} V_{\rho\sigma} - \frac{\tilde{\lambda}_V}{2M_W^2} W^-_{\rho\mu} W^{+\mu}_\nu \varepsilon^{\nu\rho\tau\omega} V_{\tau\omega}\end{aligned} \tag{2.22}$$

where $V = \gamma, Z$, M_W is the W boson mass and the normalisation constants g_{WWV} are chosen as $g_{WW\gamma} = e$ and $g_{WWZ} = e\cot\theta_W$. From a comparison with equations (2.7) and (2.8), the SM prediction for the coupling constants can be found to be:

$$\begin{aligned} g_1^\gamma = g_Z^1 = \kappa_\gamma = \kappa_Z = 1, \\ \lambda_\gamma = \lambda_Z = 0, \\ g_V^4 = g_V^5 = \tilde{\lambda}_V = 0,\ \tilde{\kappa}_V = 1, (V = \gamma, Z)\end{aligned} \tag{2.23}$$

The most general form of the WWV vertexes is thus parametrised by 14 constants (7 for each vertex). The requirement of electromagnetic gauge invariance, constrains the value of g_γ^1, g_γ^4 and g_γ^5 to their SM values. The operators associated to g_V^5 violate the charge (C) and parity (P) discrete symmetries but preserve CP, while those associated to g_V^4, $\tilde{\kappa}_V$ and $\tilde{\lambda}_V$ parametrise CP violation in the gauge sector. Assuming that C and P are conserved separately in the gauge sector, the number of anomalous TGC (aTGC) couplings is restricted to 5:

$$a_i = \left(\Delta g_Z^1,\ \Delta\kappa_Z,\ \Delta\kappa_\gamma,\ \lambda_\gamma,\ \lambda_Z\right) \tag{2.24}$$

where $\Delta g_Z^1 = 1 - g_Z^1$ and $\Delta\kappa_V = 1 - \kappa_V$. These 5 parameters are energy-dependent as be discussed shortly. The further constraint of $SU(2)_L \times U(1)_Y$ global invariance reduces the number of aTGC to 3, by introducing the following relations:

$$\kappa_Z = g_Z^1 - (\kappa_\gamma - 1)\tan^2\theta_W,\ \lambda_Z = \lambda_\gamma \tag{2.25}$$

The WWγ vertex can be interpreted as a multipole expansion of the W electromagnetic coupling and the parameters g_γ^1, κ_γ and λ_γ terms related a multipole expansion of the W-photon electromagnetic interaction:

$$\begin{aligned} q_W &= e g_\gamma^1 \\ \mu_W &= \frac{e}{2M_W}(1 + \kappa_\gamma + \lambda_\gamma) \\ Q_W &= -\frac{e}{M_W^2}(\kappa_\gamma - \lambda_\gamma)\end{aligned} \tag{2.26}$$

where q_W, μ_W and Q_W are respectively the charge, magnetic dipole and electric quadrupole moment of the W boson.

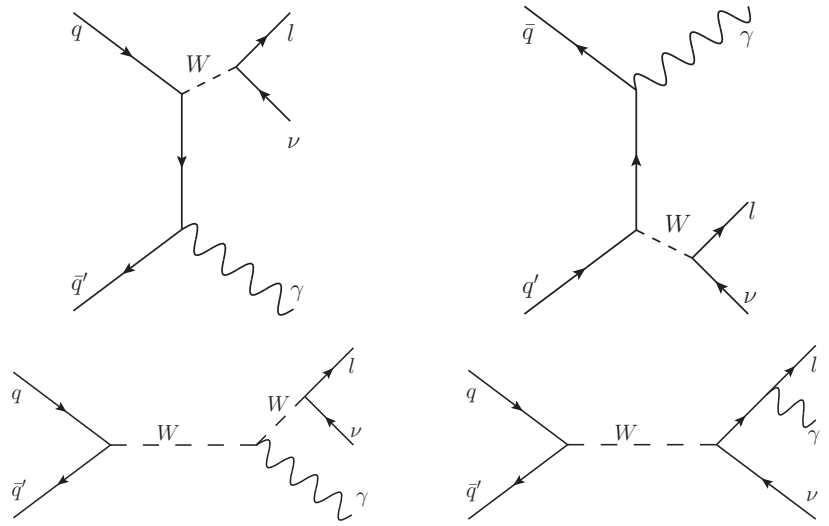


Figure 2.9: Born-level Feynman diagrams contributing to Wγ production in pp collisions. From left to right and top to bottom: the t- and u-channels, the s-channel, and radiative W decay.

2.5.1 Effect of the anomalous couplings

Wγ final states are produced, at LO in proton-proton collisions, through the partonic process $q\bar{q} \to W\gamma$. The contributing Feynman diagrams are shown in Figure 2.9: t-, u-, s-channel production and radiative W decay. Of these diagrams, only the s-channel production contains the triple WWγ vertex, which is sensitive to aTGC.
The modification of the scattering amplitudes due to the anomalous couplings can be understood in terms of the contributions to the helicity[5] amplitudes. The helicities of the incoming $q\bar{q}$ pair is fixed to $(-\frac{1}{2}, \frac{1}{2})$ by the $V - A$ structure of the weak charged currents. Thus the anomalous contribution to the s-channel interaction depends only on the helicities of the outgoing photon and W $(\lambda_\gamma, \lambda_W)$ and can be expressed as [40]:

$$\begin{aligned} \Delta M_{\pm,0} &= \frac{e^2}{\sin\theta_W} \frac{\sqrt{\hat{s}}}{M_W} \left(\Delta\kappa_\gamma + \lambda_\gamma\right) \frac{1}{2} \left(1 \mp \cos\theta^*_\gamma\right) \\ \Delta M_{\pm,\mp} &= \frac{e^2}{\sin\theta_W} \frac{1}{2} \left(\frac{\hat{s}}{M_W^2}\lambda_\gamma + \Delta\kappa_\gamma\right) \frac{1}{\sqrt{2}} \sin\theta^*_\gamma \end{aligned} \tag{2.27}$$

where the notation $\Delta M_{\lambda_\gamma,\lambda_W}$ was used, $\sqrt{\hat{s}}$ is the partonic centre of mass energy and θ^*_γ is the angle between the photon and the incoming quark in the partonic centre of mass system. Since the aTGC contribution enters only in the s-channel exchange, only the helicity states in equation (2.27) receive contributions from the anomalous interaction. No other helicity combinations are reachable though the s-channel exchange of a spin-1 particle, as it is forbidden by angular momentum conservation.

Examining Equation (2.27), two important points should be noted. The first concerns the energy-evolution of the anomalous coupling contribution: λ-type terms lead to enhancements of the scattering amplitudes that grow with the partonic centre of mass energy squared while κ-type contributions grow with the centre of mass energy. The different behaviours correspond to the different dimensionality of the corresponding operators in equation (2.22). For both λ- and κ-type couplings, at sufficiently large values of $\hat{s}/M_W^2$ the terms in equation (2.27) become larger than the SM contributions. Because of the

[5]The helicity of a quantum system is defined as the projection of the spin on the momentum $h = \frac{\vec{S}\cdot\vec{p}}{|\vec{p}|}$

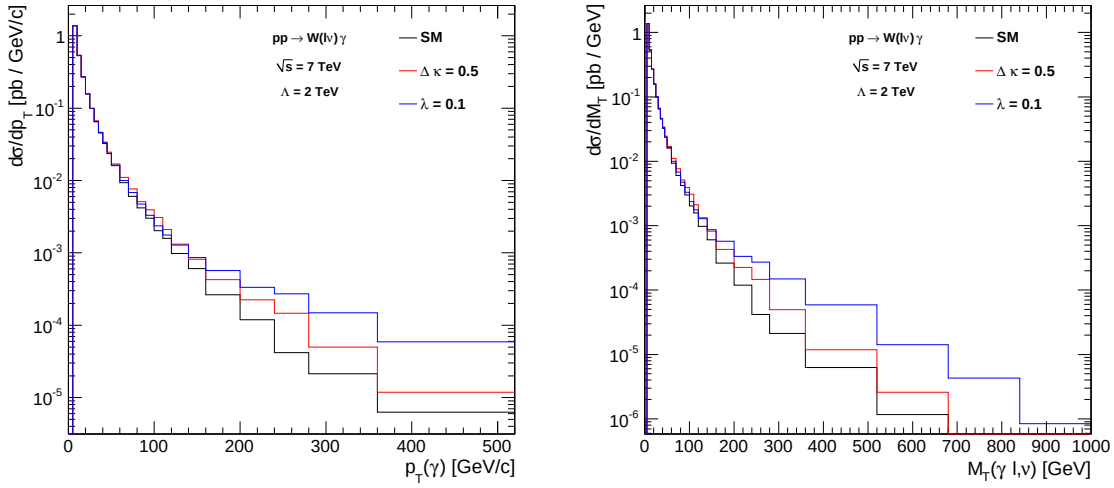


Figure 2.10: Effect of anomalous WWγ coupling of on the kinematics of leptonic Wγ final states. (left) Photon transverse momentum distribution. (right) Wγ transverse mass distribution. The enhancement in the high energy domain due to aTGC is visible. The plots were obtained with the generator WGAMMA_NLO [41] at NLO in pQCD, at parton level. A dipole form factor with $n = 2$ and $\Lambda = 2$ TeV was used. Some kinematic selections were applied to the final state particles, the details of which do not affect the overall picture. Details are reported in chapter 8.

energy-squared behaviour, the sensitivity to λ-type couplings increases rapidly with increasing collider energy, thus providing an advantage of the LHC over the Tevatron collider. The second point to notice is the angular behaviour of the anomalous coupling terms. The λ-type couplings give rise to an increase proportional to $\sin\theta^*_\gamma$, implying that the photon will more likely be reconstructed in the central part of the detector, which has better performances.

From the experimental point of view, the detection of Wγ production is more easily performed in the case of a leptonic decay of the W. The signature of this decay is the presence of a high transverse momentum lepton in association with missing energy due to the escaping neutrino. Such a clean signature allows to reduce the background due to jet production. From the discussion in the previous paragraph, it is clear that the aTGCs give rise to observable effects in both the event rate and the kinematics of the final state. In particular, an enhancement of events produced at large angles and high energy is expected. Figure 2.10 shows two examples of kinematic variables sensitive to aTGCs: the first is the photon momentum in the plane transverse to the incoming beam partons, while the second is the transverse mass of the Wγ system, defined as:

$$M_T(l\gamma;\nu) = \left[\left(\sqrt{M(\gamma,l)+p_T^2(\gamma,l)}+p_T(\nu)\right)^2 - p_T^2(\gamma,l)\right]^{1/2} \tag{2.28}$$

where $M(\gamma,l)$ and $p_T(\gamma,l)$ are respectively the mass and transverse momentum of the photon-lepton system, while $p_T(\nu)$ is the neutrino transverse momentum. The distribution of both variables shows an enhancement in the high energy regime in the presence of aTGCs. This type of observables can thus be used in order to probe the gauge couplings. Figure 2.11 shows the ratio between the anomalous and SM cross section for Wγ production with a photon transverse momentum above 80 GeV as a function of the aTGC. The enhancement due to the anomalous coupling is clearly visible.

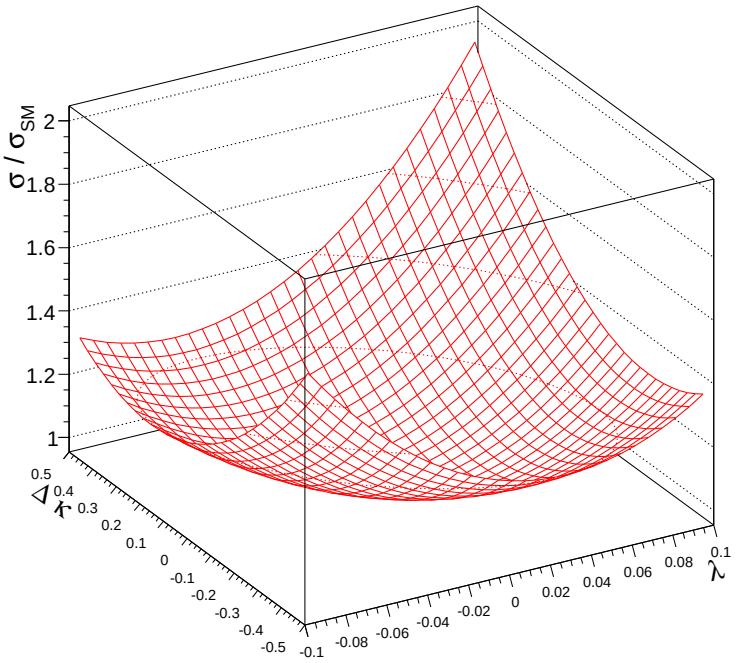


Figure 2.11: Ratio between integrated cross section for Wγ production with photon transverse momentum above 80 GeV in the presence of aTGC and the SM expectation. The enhancement due to anomalous coupling is visible.

2.5.2 High-energy limit

The energy behaviour of the anomalous coupling contribution to the scattering amplitudes represents a distinctive feature that makes the effect observable. However, a linear growth of the scattering amplitudes with the energy or the energy-squared leads to a violation of unitarity at high enough energies.

The pathological behaviour of the anomalous interactions at high energies is simply a consequence of the non-renormalisability of the effective Lagrangian in equation (2.22). This is of no-surprise as this effective theory is only supposed to be valid up to some scale Λ, related to the energy at which new degrees of freedom will manifest themselves. Expression (2.22) can be regularised by replacing the aTGC parameters with energy-dependent form factors. The functional form of the latter is arbitrary. The simplest choice is to introduce a sharp cut-off at the scale Λ and setting $a_i(\hat{s}) = a_i\theta\left(1-\sqrt{\hat{s}}/\Lambda\right)$, where $\theta(x)$ is the Heaviside step function. In this case, the unitarity constraint leads to the following bounds [42]:

$$\begin{aligned}\Lambda^2 &< 0.99\ \text{TeV}^2/|\lambda_\gamma| \\ \Lambda^2 &< 1.86\ \text{TeV}^2/|\Delta\kappa_\gamma| \\ \Lambda^2 &< 0.87\ \text{TeV}^2/|\Delta g_Z^1|\end{aligned} \tag{2.29}$$

A more sophisticated approach that can be found in the literature is to use the so-called dipole form

factors, which is motivated by the nucleon elastic form factor used in low energy scattering experiments:

$$a_i(\hat{s}) = \frac{a_i}{\left(1 + \frac{\hat{s}}{\Lambda^2}\right)^n} \tag{2.30}$$

It is sufficiency to have $n > 1$ to regularise both λ- and κ-type aTGCs. The choice of the form factor and the regularisation scale is unimportant when the anomalous coupling limits are expressed as a function of $\hat{s}$. This is usually the case for analyses performed at leptonic machines, while it can be more challenging for experiments at hadronic machines, especially in the presence of missing energy in the final state. Such situation is reflected in the choices taken by the LEP and Tevatron collaborations. In the former case, TGC analyses are performed without the introduction of form-factors, while in the latter case a dipole form factor is chosen, with $n = 2$ [43].

2.5.3 Experimental limits on anomalous couplings

Two types of experiments are sensitive to anomalous gauge couplings. The first category comprises low energy experiments that are sensitive to anomalous couplings through loop diagrams. However, limits set by low energy experiments, even though supported by high precision measurements, are model-dependent and cannot replace direct measurements based on the observation of di-boson final states. This kind of experiments falls in the second category, which comprises two main groups: lepton collider and hadron collider experiments.

The most stringent limit on $\Delta\kappa_\gamma$ from low energy experiments has been obtained by the CLEO collaboration through the measurement of the $b \to s\gamma$ branching ratio [44]. For $\lambda_\gamma = 0$, the 95% confidence level interval for $\Delta\kappa_\gamma$ is:

$$-2.6 < \Delta\kappa_\gamma < -1.2, \text{ or } -0.6 < \Delta\kappa_\gamma < 0.4 \tag{2.31}$$

A direct measurement of the anomalous coupling parameters performed at the $Sp\bar{p}S$ collider by the UA2 collaboration through a maximum-likelihood fit of the photon transverse momentum spectrum in Wγ events yields [45]:

$$-3.6 < \Delta\kappa_\gamma < 3.5, \quad -4.5 < \lambda_\gamma < 4.9 \tag{2.32}$$

To-date, the most stringent limits on aTGC are the ones provided by the LEP and Tevatron experiments. A recent combination of all data from the LEP experiments [10, 46], including WW, single W and $\bar{\nu}\nu\gamma$ production data gives the following 68% confidence level limits:

$$\Delta g_Z^1 = 0.016^{+0.022}_{-0.019}, \ \Delta\kappa_\gamma = 0.027^{+0.044}_{-0.045}, \ \lambda_\gamma = 0.028^{+0.020}_{-0.021} \tag{2.33}$$

No form factor was used in the analysis of the data, and equation (2.25) was assumed.
The CDF and D0 collaborations published the first results of limits on anomalous couplings using Tevatron Run-II data [43]. Both experiments assume a dipole form factor for the aTGCs. The CDF analysis combines data from Wγ and WZ channels assuming $\lambda_\gamma = \lambda_Z$ and $\Delta\kappa_\gamma = \Delta\kappa_Z$ to obtain the following 95% confidence level intervals for $\Lambda = 1.5$ TeV:

$$-0.46 < \Delta\kappa_\gamma < 0.39, \quad -0.18 < \lambda_\gamma < 0.17 \tag{2.34}$$

The D0 analysis is based on the study of Wγ production and obtains the following limits, at 95% confidence level and for $\Lambda = 2$ TeV:

$$-0.49 < \Delta\kappa_\gamma < 0.51, \quad -0.12 < \lambda_\gamma < 0.13 \tag{2.35}$$

Due to its large centre of mass energy, the LHC is expected to provide stringent limits especially on the λ_γ coupling. With large integrated luminosities, in particular, it expected to improve the LEP results by a factor of 2–3 [47].

2.6 Conclusions

This chapter was devoted to an introduction to LHC Physics, in particular to those aspects that are most relevant to the topic of this thesis. The main concepts that lead to the construction of the Standard Model of Particle Physics were briefly introduced. Three main ingredients concur to the definition of the theory: the choice of a symmetry group, the representation of fermions and scalars and the mechanism of symmetry breaking. The experimental measurements of the SM predictions were briefly illustrated, the fundamental limitations of the model and some of the arguments suggesting that new phenomena are expected to be observed at the LHC were introduced. The Physics of photon production in hadronic collisions and its potential to test pQCD predictions and provide input to the measurement of PDFs, was discussed. Finally, the Physics of di-boson production and its sensitivity to anomalous gauge couplings was also discussed and the current experimental constraints on aTGCs were reviewed.

Chapter 3

Experimental apparatus

Contents

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a superconducting hadron accelerator and collider with two rings, installed in the 26.7 km tunnel that was constructed for the CERN LEP collider.

The machine, currently in operation, is composed by two rings that accelerate counter-rotating beams of protons and heavy ions. The design centre-of-mass energy is $\sqrt{s} = 14$ TeV for proton-proton collisions and $\sqrt{s} = 5.52$ TeV/n for Pb-Pb collisions. In the 2010 run, the proton-proton centre-of-mass energy has been of 7 TeV and it will possibly be raised to 8 or 10 TeV for the 2011 run.

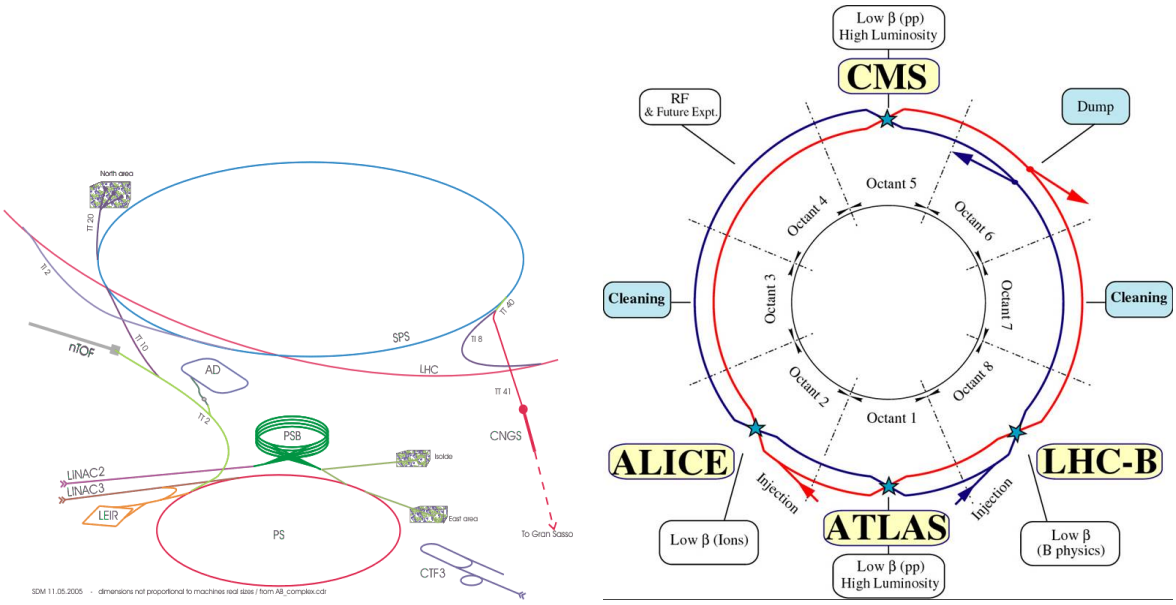


Figure 3.1: CERN accelerator complex (left) and layout of the LHC accelerator and its eight insertion points (right). Figures reproduced from [48].

Two transfer tunnels link the LHC to the CERN accelerator complex, depicted in Figure 3.1 (left), that acts as its injector. The injection chain for protons (heavy ions) starts with LINAC2 (LINAC3) after which follows the Proton Synchrotron Booster (PSB) (the Low Energy Ion Ring, LEIR, for ions), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS).

Details on the CERN accelerator complex can be found in reference [48], while an accurate description of the machine design and functioning can be found in reference [49]. Here, a brief description of the LHC will be presented.

The design of the LHC was partially determined by the pre-existing infrastructure. The cost-effective decision to reuse the LEP 3.7 m-diameter tunnel "as built" determined the adoption of the twin-bore magnet design that characterises the machine. The tunnel is geometrically organised in eight crossing points, flanked by long straight sections, and eight arcs. Each straight section has a length of 528 m and can serve as an experimental or utility insertion. A schematic representation of the LHC tunnel is reported in Figure 3.1 (right). The insertion points are labelled with integer numbers increasing counter-clockwise.

Four of the crossing points are instrumented with detectors: the ATLAS, ALICE, CMS and LHCb experiments are located at Points 1, 2, 5 and 8 respectively. The insertions in Points 3 and 7 host collimators used to remove halo particles with large longitudinal and transverse oscillation amplitudes. Two independent Radio Frequency (RF) superconducting cavities are located in Point 4 with the purpose of

capturing, confining and accelerating the beams. Point 6 contains the beam dump insertion, where the beams can be extracted from the machine by a combination of horizontally deflecting fast-pulsed magnets ("kickers") and vertically deflecting septum magnets.

The arcs of the LHC lattice are made of 23 rectangular arc cells, which contain different types of magnets. The magnet system of the accelerator is based on NbTi superconductors cooled to a temperature below 2 K and generates fields above 8 T. Roughly 1 600 superconducting magnets are employed in total.

The LHC beams are subdivided in bunches containing 10^{11} protons in nominal conditions and distributed in the machine with a 25 ns spacing between them. The orbit length is 89.1 μs, so up to 3564 bunches can be hosted in the machine. The number of filled bunches however depends also on the structure of the injection chain. As a result, 2808 bunches will be filled in design conditions.

The rate at which events are produced in collisions is directly proportional to the particles' interaction cross section and to the luminosity of the machine. For a Gaussian beam distribution, the machine luminosity can be expressed as:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi\varepsilon_n \beta^*} F$$

where N_b is the number of bunches in each beam, n_b the number of particles per bunch, f_{rev} the revolution frequency of the beams, γ_r the Lorentz relativistic factor, ϵ_n the normalised emittance [50] of the beams and β^* the value of the machine beta function [50] at the interaction point. F is a geometric reduction factor defined as:

$$F = \left[1 + \left(\frac{\theta_c \sigma_z}{2\sigma_*}\right)^2\right]^{-1/2}$$

here θ_c is the beam crossing angle and σ_z (σ_*) the longitudinal (transverse) bunch size RMS.
The amount of data delivered to the experiments, which determines the sensitivity to rare Physics processes is measured by the integrated machine luminosity defined as $\int_0^t \epsilon L dt$, where t is the machine operation time and ϵ the machine availability.
The ϵ term accounts for the time in which the machine is offline and the ratio between the total turnaround time (time between two successive fills) and the duration of Physics runs. During the first part of the 2010 run, the LHC machine availability has been on average $\sim 40\%$ [51].
The design peak luminosity of the LHC is $10^{34} cm^{-2} s^{-1}$. During the 2010 run, a peak luminosity of $10^{32} cm^{-2} s^{-1}$ has been attained, according to expectations. The instantaneous luminosity is not constant over a Physics run, but decays due to the degradation of the intensity and emittance of the beams. During nominal operation, the main reason for luminosity decay is beam lost in collisions. However, other effects, like beam-beam or beam-gas interactions can also contribute to the luminosity degradation. The luminosity life-time can be expressed as:

$$\frac{1}{\tau} = \frac{1}{\tau_{Nuc}} \cdot \frac{1}{\tau_{beam-gas}} \cdot \frac{1}{\tau_{beam-beam}}$$

where τ_{Nuc}, $\tau_{beam-gas}$ and $\tau_{beam-beam}$ parametrise the losses due to collisions, beam-gas and beam-beam interactions, respectively. The instantaneous luminosity can then be expressed as

$$L(t) = L_{peak} \cdot \mathrm{e}^{-t/\tau}$$

The instantaneous luminosity is measured by the experiments using detectors located at small angles with respect to the interaction point. In the case of CMS, such measurement is performed by means of a the quartz-fibre calorimeter located at 11.2 m from the interaction point [52]. During the 2010 run, the CMS experiment recorded an integrated luminosity of 43 ± 5 pb^{-1} in proton-proton collisions, as shown in Figure 3.2.

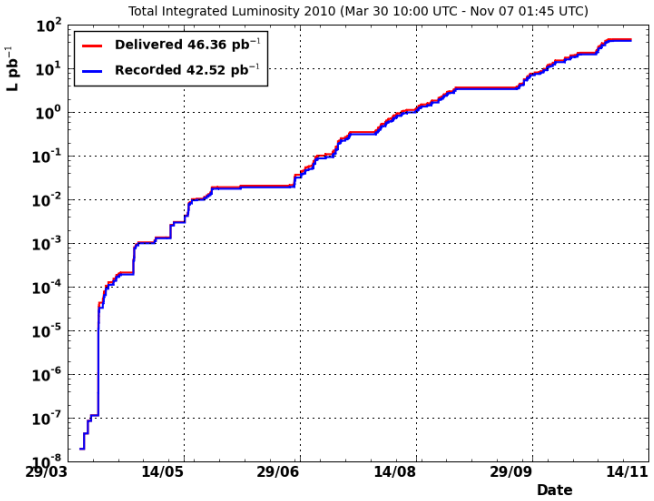


Figure 3.2: Integrated luminosity recorded by the CMS experiment in the 2010 proton-proton run. The roughly exponential increase corresponds to successive increases in the instantaneous luminosity. Figure reproduced from [53].

3.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) detector is a multi-purpose apparatus currently operating at the LHC. As seen in the previous section, the LHC is designed to collide proton beams at the centre of mass energy of 14 TeV and instantaneous luminosity of $10^{34} cm^{-2} s^{-1}$. The large proton-proton total cross section, the high beam intensity and the short spacing between bunches pose very challenging requirements on the detector in terms of radiation-tolerance, on-line data reduction and time resolution. Also, given the large multiplicity of the interactions (up to 20 proton-proton collisions per bunch-crossing on average are expected at nominal luminosity, leading to the production of more than 1000 charged particles), very high detector granularities are required.

The design of the CMS detector meets the requirements dictated by the LHC Physics program, while the same time coping with the demanding environmental conditions. The detector provides good muon identification and momentum resolution ($\sim 1\%$ at 100 GeV/c) and efficient charge discrimination up to momenta of $\mathcal{O}(\text{ TeV}/c)$; good charged particle momentum resolution in the inner tracker and good vertex resolution; good electromagnetic energy resolution, good di-photon and di-electron mass resolution, π^0 rejection and wide geometric coverage; good missing-transverse-energy resolution and di-jet mass resolution.

The central feature of the CMS apparatus is a superconducting solenoid, of 6 m internal diameter, providing a field of 3.8 T. Within the field volume are the silicon pixel and strip tracker, the crystal electromagnetic calorimeter (ECAL) and the brass/scintillator hadron calorimeter (HCAL). Muons are measured in gas-ionization detectors embedded in the steel return yoke. In addition to the barrel and endcap detectors, CMS has extensive forward calorimetry. The apparatus has an overall length of 22 m, a diameter of 15 m, and weighs 12 500 tonnes.

A detailed description of CMS can be found in reference [54]. Here, the main features of the detector components will be reviewed, with particular focus on the parts that are most relevant for the subject of this thesis.

CMS uses a right-handed coordinate system, with the origin at the nominal interaction point, the x-axis

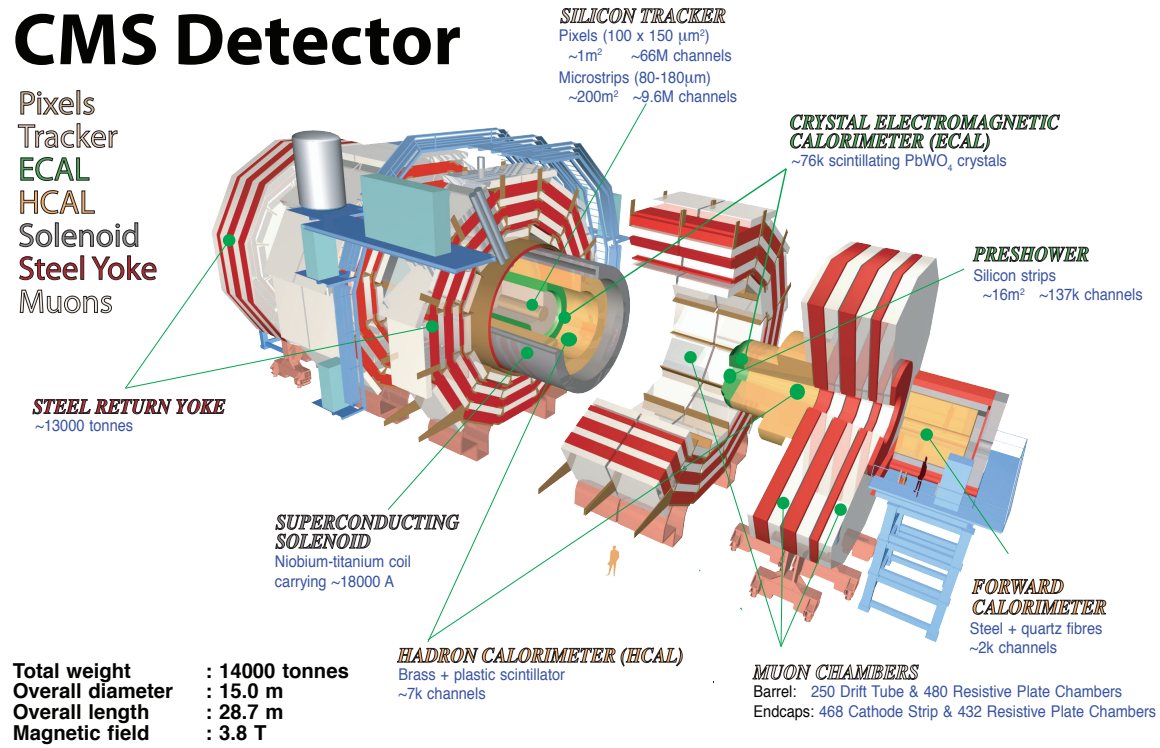


Figure 3.3: Representation of the CMS detector and its major components.

pointing to the centre of the LHC, the y-axis pointing up (perpendicular to the LHC plane), and the z-axis along the anticlockwise-beam direction. The polar angle, θ, is measured from the positive z-axis and the azimuthal angle, ϕ, is measured in the x-y plane. The pseudorapidity is defined as $\eta = -\log\tan(\theta/2)$.

3.2.1 Muon System

The CMS muon system is capable of reconstructing muons in the kinematic range relevant for the LHC Physics programme. It has been designed to provide good muon identification, momentum resolution and efficient trigger on muons with $|\eta| < 2.4$ and 10 GeV/$c \lesssim p_T \lesssim 1$ TeV/c.

The muon detectors are hosted in the magnet flux-return yoke. Besides conveying the high field generated by the solenoidal magnet, the iron yoke serves as hadron absorber for muon identification. As the system comprises roughly 25000 m^2 of gaseous detectors, inexpensive, reliable and robust technologies have been chosen.

In the barrel region, where the muon flux is low and the magnetic field uniform, drift chambers with standard rectangular drift cells have been chosen. The CMS drift tube system (DT) covers the pseudorapidity region $|\eta| < 1.2$. The system is made by 4 stations interspersed in the barrel return yoke. The three innermost stations are made by 4 pairs of chambers measuring the $\phi - r$ or z directions, while the outermost station measures only the z-view. The drift cells of each chamber are offset by a half-cell width with respect to their neighbour to eliminate dead spots in the efficiency.

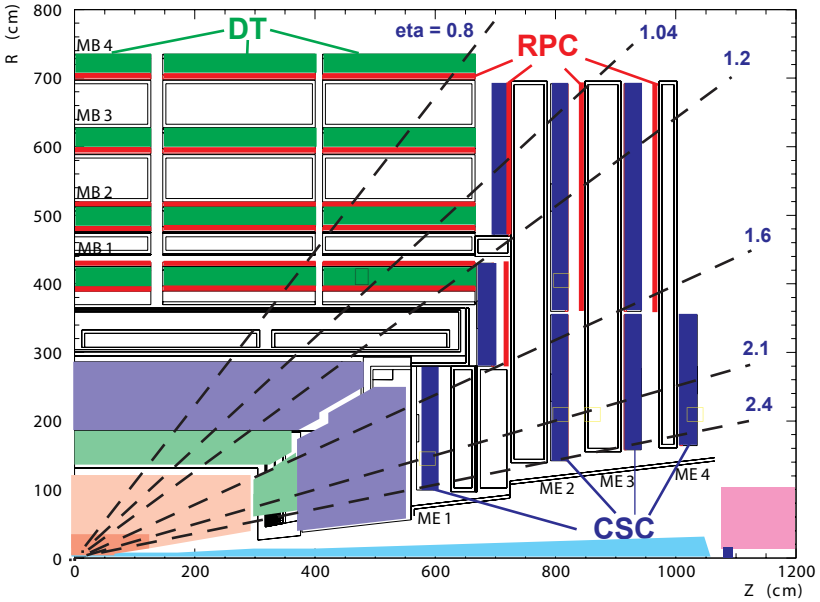


Figure 3.4: Schematic showing the three main systems used for muon detection in CMS; while RPC detectors provide accurate time information for triggering, DT (CSC) detectors also provide accurate position information in the barrel (endcaps).

In the endcap regions, where the muon fluxes are larger and the magnetic filed is large and non-uniform, cathode strip chambers (CSC) are used. Such detectors have a fast response time, are finely segmented and highly tolerant to radiation, and cover the region $0.9 < |\eta| < 2.4$. The technology is that of multi-wire proportional chambers, with segmented cathode to achieve good spatial resolution. The CSC system comprises 468 trapezoidal chambers covering 10 or 20° in the ϕ direction.

In order to improve the performance of the muon trigger, an additional system of resistive plate chamber detectors (RPC) is installed in the region $|\eta| < 2.1$. The RPC system provides a fast, independent, and highly segmented trigger system providing good time resolution, even if with coarser position resolution than the DT and CSC systems. A double-gap technology has been chosen to allow the use of lower gains per gap and provide larger detection efficiency. Two rectangular-section RPCs per DT chamber are installed in the barrel region and two trapezoidal ones per CSC chamber in the endcap regions.

A diagram showing the mechanical layout of the three muon detection systems can be found in Figure 3.4. The muon system was extensively exercised prior to the LHC start-up, using cosmic ray events. The alignment of the system prior to the LHC 2009 run was estimated to be as good as originally foreseen after 10 pb^{-1} of 14 TeV proton-proton collisions [55]. The resolution of the momentum measurement has been measured using cosmic rays traversing the full detector. Comparing the momentum measurement in the upper and lower part of the spectrometer, the momentum resolution was found to vary between $\sim 1\%$ and $\lesssim 10\%$ for momenta between 10 GeV/c and 1 TeV/c, as shown in Figure 3.5 (left). The di-muon invariant mass resolution in Z events observed [56] in proton-proton collisions after 200 pb^{-1} at $\sqrt{s} = 7$ TeV has also been found to bee in good agreement with the expected start-up performances as shown in Figure 3.5 (right).

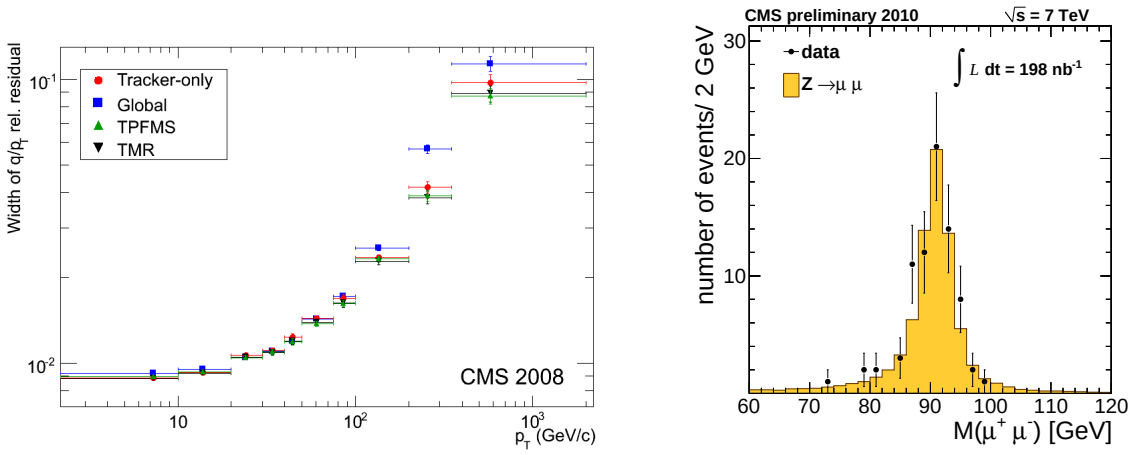


Figure 3.5: (left) Resolution of the muon momentum measurement using cosmic rays traversing the full detector. (right) Di-muon invariant mass in Z events observed with first proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$. Figures reproduced from [55] and [56] respectively.

3.2.2 Calorimeter system

The calorimeter system is in charge of measuring the energy of electrons, photons and hadrons produced in collisions. The CMS calorimeter is divided into electromagnetic and hadronic sections. The electromagnetic section surrounds the silicon tracker and is made by an homogeneous $PbWO_4$ crystal calorimeter covering the pseudorapidity region up to $|\eta| < 3.0$, complemented, in the forward part ($1.653 < |\eta| < 3.0$) by a Si-Pb preshower. The hadronic section surrounds the electromagnetic one and uses a brass-scintillator sampling technology in the region $|\eta| < 3.0$. A tail-catcher, located outside of the magnet yoke, is used to increase the calorimeter thickness in the region $|\eta| < 1.3$. A Cherenkov steel-quartz calorimeter, located at 11.2 m from the nominal interaction point covers the region $3.0 < |\eta| < 5.0$.

Electromagnetic Calorimeter

The characteristics of $PbWO_4$ crystals make them an appropriate choice for calorimetry at the LHC. The large density ($8.28\ g/\,\text{cm}^3$), small radiation length (0.89 cm) and Molière radius (2.2 cm), allow the realisation of a compact and high granularity calorimeter. The fast response of the medium to incoming particles (80% of the scintillation light is emitted within 25 ns) matches well the requirements posed by the event pile-up rate foreseen during high-luminosity operation of the LHC. The crystals have a relatively low light output that depends on the temperature (2.1%/° C at 18° C). This requires the use of efficient photo-detectors and a very good temperature stability.
The radiation tolerance of the crystals is sufficiently high to ensure good performances throughout the LHC operation. Radiation has been observed to have no effect on the scintillation mechanism, while it affects the crystal transmission properties through the formation of colour centres in the crystal lattice. The mechanism has been observed to be reversible and the transparency to be recovered in the absence of radiation. In order to preserve the detector resolution, however, the crystal transparency needs to be closely monitored during data taking.
The crystal detector is mechanically organised in two regions: a barrel section (EB) covers the region $|\eta| < 1.479$ and two endcap disks (EE) cover the region $1.479 < |\eta| < 3.0$. In the EB region, the light is read out by means of avalanche photo diodes (APD), while in EE vacuum photo triodes (VPT) are used. The barrel detector, shown in Figure 3.6 (left), has an internal radius of 1.29 m. Its granularity is 360-fold in ϕ and 2×85-fold in η, resulting in a total of 61,200 crystals. The crystals have tapered shapes and are mounted in a quasi-projective geometry, so that they make a small angle with respect to the direction

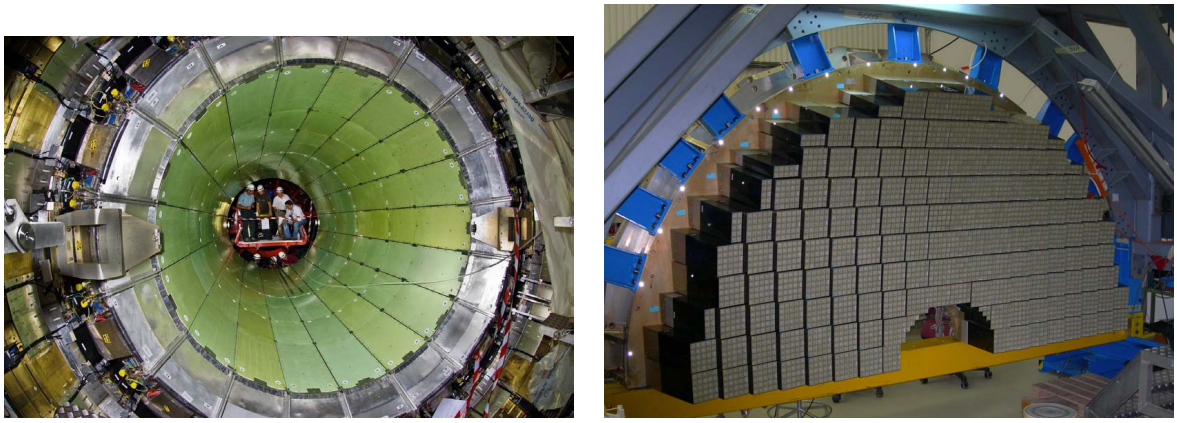


Figure 3.6: The electromagnetic calorimeter: barrel section after installation in CMS (left) and half of one endcap section during the assembly (right). Figures are copyright of CERN and the CMS collaboration.

of the nominal interaction point, in order to maximise the detector hermeticity. The crystal front faces measure $2.2 \times 2.2\,\mathrm{cm}^2$ and the length is 23.0 cm, corresponding to roughly 25.8 radiation lengths. The crystals are mechanically organised in 36 *supermodules*, each covering a 10° sector in ϕ direction in the positive or negative η side.
The light readout is performed by a pair of APDs, each with an active area of $5 \times 5\,\mathrm{mm}^2$, operated at gain 50, at a temperature of 18° C and read out in parallel. The gain stability directly affects the ECAL energy resolution. Variations of the bias voltage affect the APD gain (1/M dM/dV $\sim$ 3.1%/V at gain 50), as well as temperature variations (1/T dM/dT $\sim -2.4\%/^\circ$ C). In order to keep these contributions to the resolution at the level of per mille, a stability of the voltage within 10 mV and of the temperature within 0.05° C are required.
The EE detector is located at a distance of 3.15 m from the nominal interaction point along the beam axis. The detector is mechanically divided in four *dees*, one of which shown in Figure 3.6 (right). Each dee covers a 180° sector in the ϕ direction. The four dees contain a total of 14 648 tapered crystals, arranged in a rectangular x-y grid and grouped in squared 5×5 mechanical structures called super-crystals. The front face of the crystals measures 2.86×2.86 cm^2, for a length of 22.0 cm, corresponding to roughly 24.7 radiation lengths.
Each crystal is read out by one VPT with a diameter of 25 mm and an active area of $\sim 280\,\mathrm{mm}^2$. These VPTs have an anode of very fine copper mesh (10 μm pitch) allowing them to operate in the CMS magnetic field. The anode sensitivity of a VPT may show a dependence on count rate (anode current) under certain conditions. The magnitude of the effect may vary from a few percent to a few tens of percent in the absence of a magnetic field. Even though the effect is strongly suppressed in the presence of a strong magnetic field, it has been judged prudent to incorporate a light pulser system on the ECAL endcaps. This is able to deliver a constant background rate of pulses to the VPT in order to keep the count rate constant even in the absence of LHC interactions.
In the region $1.653 < |\eta| < 2.5$, the crystal calorimeter is complemented by a Pb-Si preshower (ES). The function of the latter is to improve the discrimination between isolated photons and neutral meson decays. The Preshower is a sampling calorimeter made by lead radiators and silicon sensors. Two pairs of silicon sensors planes (sampling the x and y shower profile) are interleaved with 2 (1) radiation-length radiators. A total of 137 000 strips are organised in more than 4 000 sensor, each with 32 strips and measuring $63 \times 63\,\mathrm{mm}^2$.
The ECAL energy resolution can be parametrised as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2$$

where S is the intrinsic stochastic term, N the noise term and C the constant term. The resolution of the barrel detector was measured in test-beams, using electrons with energies between 20 and 250 GeV. For electrons impinging in the barrel inside an area of $4 \times 4\,\mathrm{mm}^2$ around the crystal centres, the coefficients above were found to be $S = 2.8\%$, $N = 0.12$ and $C = 0.3\%$, as shown in Figure 3.7 (left).
In the environment of CMS, for unconverted photons with energies in the range of interest for physics analyses, $\mathrm{E}_{shower} \sim 100$ GeV, the energy resolution will be dominated by the constant term. Thus, the performance of the CMS ECAL at the LHC strongly depends on the quality of the calibration and monitoring. A calibration strategy which exploits Physics events to achieve an ultimate calibration precision better than 0.5% has been elaborated by the CMS collaboration [57]. The precision of the channel-to-channel calibration obtained after $250\,\mathrm{nb}^{-1}$ of collisions at $\sqrt{s} = 7$ TeV was 0.6% for the central barrel, in good agreement with expectations from MC simulations. The global energy scale of the detector, after the same luminosity, was estimated to be known with a precision of 1 (3)% in EB (EE).
As said, although radiation resistant, ECAL $PbWO_4$ crystals show a limited but rapid loss of optical transmission under irradiation. At the ECAL working temperature (18° C) a dose-rate dependent equilibrium of the optical transmission is achieved. In order to preserve the detector resolution, the evolution of the crystal transparency is measured, during data taking, using laser pulses injected into the crystals via optical fibres. Laser pulses are generated with a rate of ~ 100 Hz during the empty section at the end of the LHC orbit. Different part of the detector are monitored in sequence.
The performance of the detector in terms of time resolution was studied in-situ prior to the LHC start-up using cosmic ray events [58], and using collision events [59]. For events depositing energies above 10 to 20 GeV, the time resolution of the detector was found to be better than 100 ps, as shown in Figure 3.7 (right).

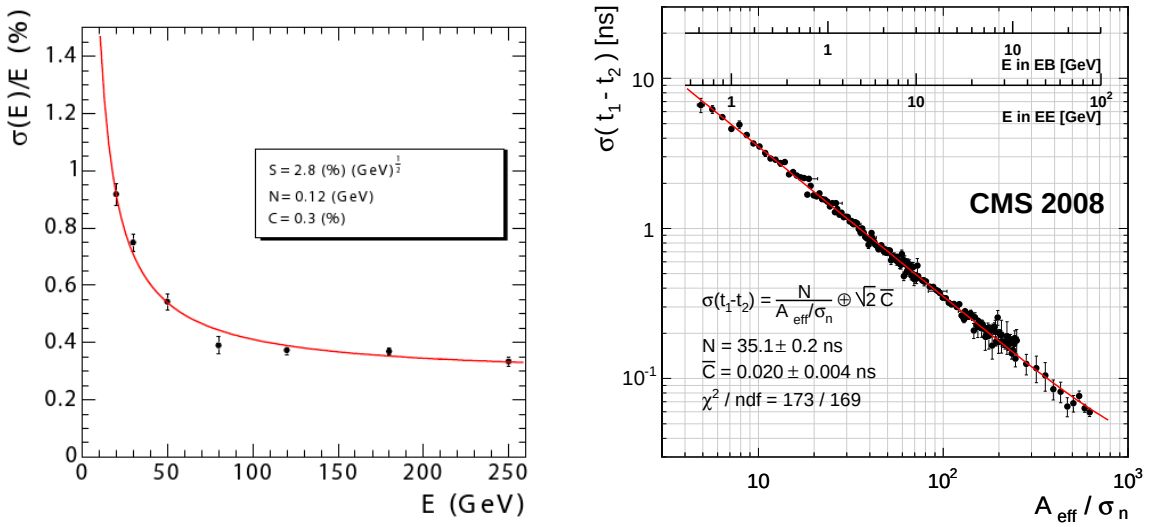


Figure 3.7: (left) ECAL barrel energy resolution as measured in electron test-beams. (right) ECAL time resolution measured with cosmic ray data. Figures reproduced from [54] and [58] respectively.

Hadron Calorimeter

The design of the hadronic section of the CMS calorimeter has been constrained by the volume available to install the detector. The barrel part of the calorimeter (HB), is inserted between the ECAL and the solenoid, for a total radial width of 1.18 m. To ensure adequate sampling of the hadronic showers, the calorimeter was extended outside of the solenoid with a tail catcher called HO, which uses the solenoid

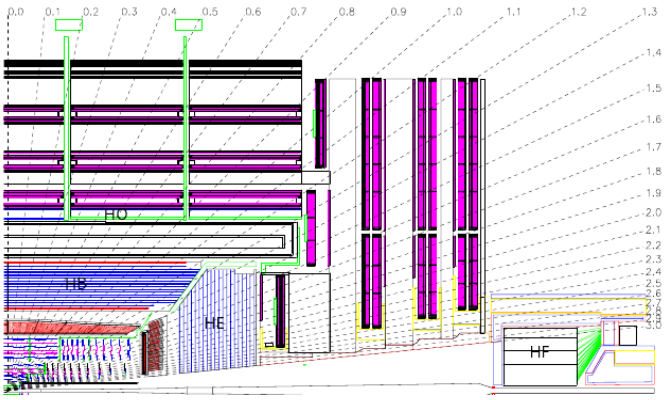


Figure 3.8: Schematic view of the HCAL components in CMS, including the barrel section (HB), the endcap section (HE), the tail catcher outside the solenoid (HO) and the forward section (HF).

as an additional absorber layer of $1.4/\sin\theta$ interaction lengths. The total thickness of the calorimeter is thus extended to a minimum of 11.8 interaction lengths.
The HB detector is a sampling calorimeter consisting of 36 identical azimuthal wedges covering the region $|\eta| < 1.3$. Each wedge is segmented into 16 azimuthal plates, bolted together in such a way that there is no projective dead material. Brass (70% Cu and 30% Zn) is used as the absorber, except for the first and last layers which are made of stainless steel for structural strength. The endcap sector (HE) covers the region $1.3 < |\eta| < 3.0$.
The active medium uses the well known tile and wavelength shifting fibre concept to bring out the light, which is then read out by means of hybrid photo diodes (HPDs). The granularity of HCAL is an integer multiple of the ECAL one. Up to $|\eta| < 1.6$, HCAL towers have a size of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$, while for $|\eta| \geq 1.6$, the size increases to $\Delta\eta \times \Delta\phi \sim 0.17 \times 0.17$.
The combined CMS calorimeter system is under-compensating, i.e. its response to electrons is larger than the one to hadrons of the same energy [60]. The pion-to-electron response ratio π/e was measured in test beams [61] and is related to the h/e ratio by the formula:

$$\frac{e}{h} = \frac{1 - f_{em}}{\pi/e - f_{em}}$$

where f_{em} is the electromagnetic fraction of the shower energy. For test-beam particles with energies above ~ 8 GeV, the h/e ratio was derived to be 1.4 ± 0.1. An event-by-event energy correction scheme was developed. A linear response (within 1.3%) to hadrons of momenta between 5 and 350 GeV/c was achieved.
The reconstruction resolution of the energy of hadronic jets was measured using data from pp collisions at $\sqrt{s} = 7$ TeV [62]. The best performances were obtained using particle-flow techniques, which correlate the information from the silicon tracker and the calorimeter system. A transverse energy resolution better than 10% for jets of transverse energy above 40 GeV was estimated, as shown in Figure 3.9 (left).
Also the performances in terms of missing energy reconstruction were measured in 7 TeV collisions [62]. Using the particle-flow algorithm, a resolution between 5% and 10% was estimated, as shown in Figure 3.9 (right).

3.2.3 Inner tracker

The tracker system is in charge of reconstructing the trajectories of charged particles and primary and secondary vertexes. At high luminosity, 20 overlapping pp collision at the LHC will produce more than

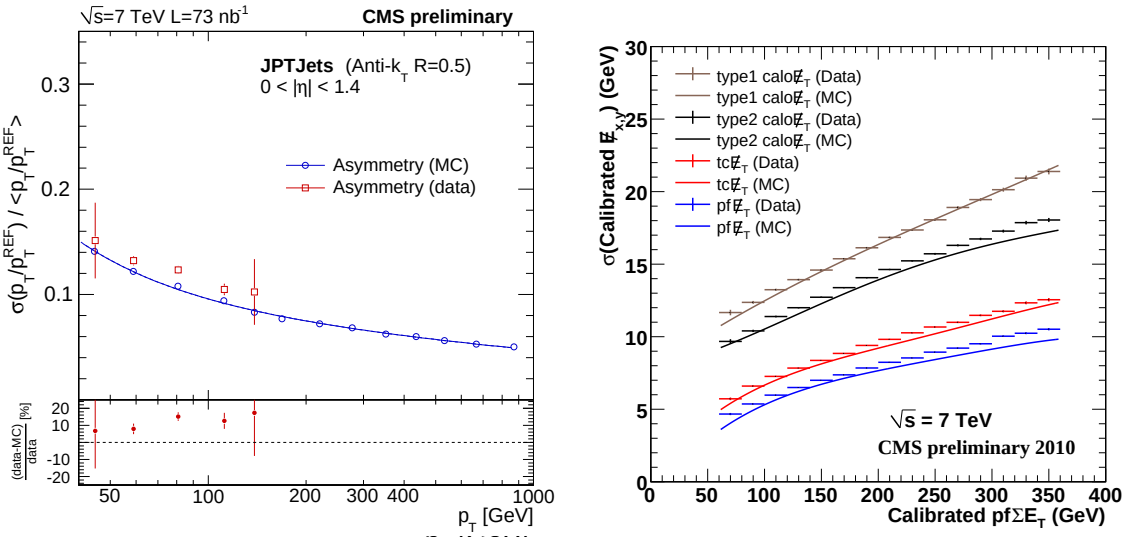


Figure 3.9: (left) Reconstructed jet energy resolution and a function of the jet transverse momentum. (right) Resolution of the missing transverse energy as a function of the visible transverse energy. Figures reproduced from [62].

1000 charged particles every 25 ns. The design of the tracker system has thus to cope with large levels of occupancy and radiation.

The CMS collaboration built a fine granularity tracking detector covering the region $|\eta| < 2.5$ and employing more than $200\,\mathrm{m}^2$ of active Si sensors. In the inner region 66 million Si pixels, with a pitch of $100 \times 150\ \mu\mathrm{m}^2$ in $r-\phi$ and z respectively, are used in order to attain an occupancy level of to $\sim 10^{-1}$/(pixel $\times$ bunchcrossing). For distances larger than 20 cm from the nominal interaction point the flux is greatly reduced allowing the use of Si strips. The dimension of the strips increases with increasing distance from the interaction point such that the occupancy level is kept to $\sim 1\%$. Thus, for radii between 20 (55) and 55 (120) cm, the pitch of the strips is typically 10 cm $\times$ 80 μm (25 cm $\times$ 180 μm).
The detector is operated at $-10°C$ in order to reduce the leakage currents induced by radiation damage and consequent increase in noise and power consumption. The operation below $0°C$ also reduces possible interactions of radiation-induced defects in the Si sensors (an effect known as reverse-annealing).
The Pixel detector is composed by 2 forward disks with 672 modules and 3 cylindrical layers with 768 modules. In each module, the pixelated sensors are bump-bonded to a radiation-hard readout chip able to perform automatic zero suppression after the amplification of the charge signal. The strips are distributed in four regions: the inner barrel (4 layers) and inner disks (3 pairs) and the outer barrel (6 layers) and end-cap disks (9 pairs). The signals from the strip sensors are amplified, shaped and pipelined for several μs by a custom integrated circuit.
A characteristic of the tracker detector which affects the performances of the calorimeter system is the amount of material that particles have to traverse before reaching the inner surface of the ECAL. The interaction with the tracker material leads to radiative effects which degrade the energy resolution. The thickness of the CMS tracker is ~ 0.4 radiation lengths at $\eta = 0$, it increases up to ~ 1.8 radiation lengths at $|\eta| \sim 1.4$ and then decreases to $\lesssim 1$ radiation length for $|\eta| > 2$. Studies performed on collisions data established that the material budget of the tracker detector is presently known with an accuracy better than 10% [63].
The precision of the detector positions with respect to the particles trajectories was 3–4 (14) μm in the barrel (endcap) section already before the first LHC collisions [64]. The performances of the Si

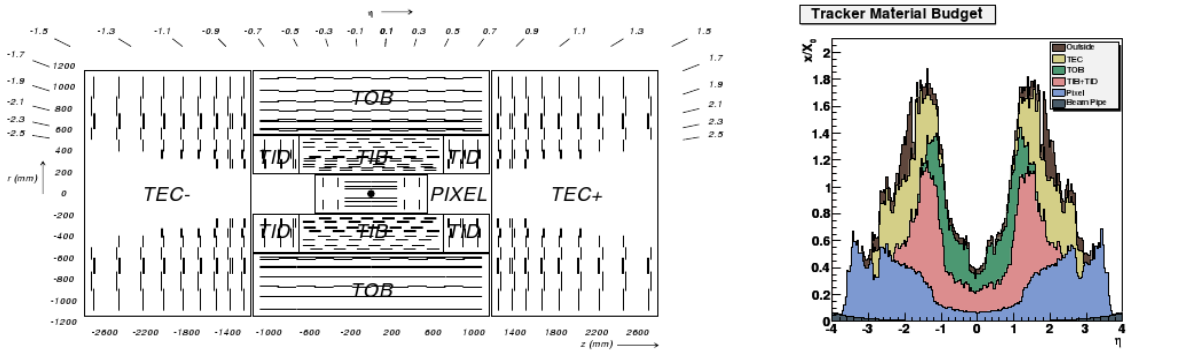


Figure 3.10: (left) Schematic representation of the mechanical layout of the CMS silicon tracker. (right) Material budget due to the presence of the silicon tracker, expressed in radiation lengths. The largest thickness traversed by matter is ~ 1.8 radiation lengths. This happens for $|\eta| \sim 1.4$ and is due to the presence of the detector service cables.

tracker have been evaluated in collision events in terms of track and vertex reconstruction efficiency and spatial resolution (shown in Figure 3.11), and in terms of track momentum resolution and particle identification efficiency [65, 63]. The performances of the detector were found to be in good agreement with expectations.

3.2.4 Trigger and Data Acquisition

The high event rate produced by the LHC collisions cannot be sustained by any storage system presently available. The collision rate at the LHC is in fact 40 MHz, while only a rate of ~ 100 Hz can be stored for offline analysis. An effective online reduction system is thus necessary in order to achieve the required 10^6 reduction factor. This essential task is accomplished by the trigger and data acquisition system.

The CMS trigger system

Figure 3.12 shows the production cross sections of some benchmark Physics process in pp collisions at $\sqrt{s} = 14\,\text{TeV}$. It can be seen that cross sections of processes being searched at the LHC, such as Higgs or SUSY production, are more than 9 orders of magnitude lower than the elastic pp cross section. It is thus clear that the trigger system represents the start of the Physics event selection process.
The CMS trigger system is organised in two levels. The level 1 trigger (L1) consists of custom-designed, largely programmable electronics, whereas the high level trigger (HLT) is a software system implemented in a filter farm of about one thousand commercial processors. The L1 trigger was designed to have an output rate limit of 100 kHz; it uses uses coarsely segmented data from the calorimeters and the muon system, while holding the high-resolution data in pipelined memories in the front-end electronics. The HLT has access to the high-resolution data in pipelined memories in the front-end electronics as well as the information from the silicon tracker. The HLT has access to the complete read-out data and can therefore perform complex calculations similar to those made in the the analysis off-line software if required for specially interesting events. The HLT [67] has been designed in order to provide the maximal flexibility in the deployment of the algorithms. In this way the online selection can be easily adapted to the different running conditions.
For reasons of flexibility the L1 Trigger hardware is implemented in FPGA technology where possible, but ASICs and programmable memory lookup tables (LUT) are also widely used where speed, density and radiation resistance requirements are important. A software system, the Trigger Supervisor, controls the configuration and operation of the trigger components.

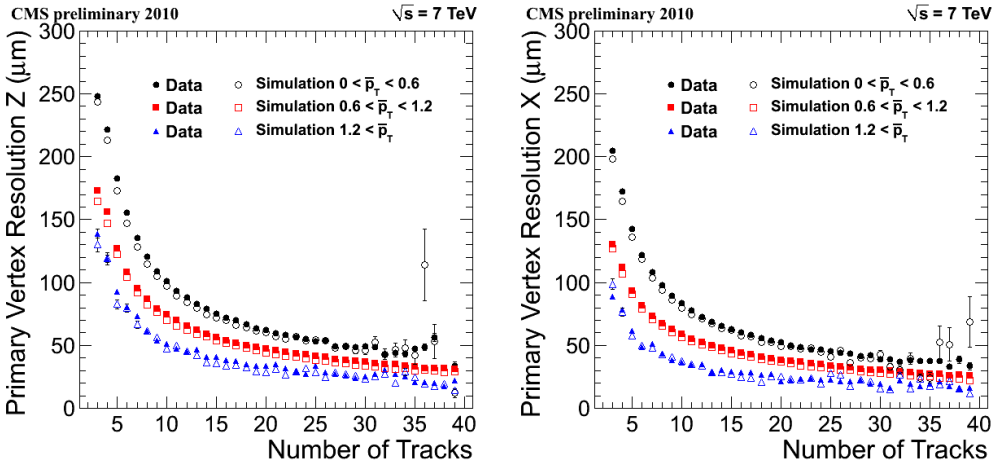


Figure 3.11: Measured resolution in the determination of the primary vertex position in the z (left) and x (right) axes as a function of the number of tracks reconstructed for each vertex in pp collisions at $\sqrt{s} = 7$ TeV. Figure reproduced from [63].

The L1 Trigger schema is shown in Figure 3.13. It has local, regional and global components. At the bottom end, the Local Triggers, also called Trigger Primitive Generators (TPG), are based on energy deposits in calorimeter trigger towers and track segments or hit patterns in muon chambers, respectively. Regional Triggers combine their information and use pattern logic to determine ranked and sorted trigger objects such as electron or muon candidates in limited spatial regions. The rank is determined as a function of transverse energy or momentum and quality, which reflects the level of confidence attributed to the L1 parameter measurements, based on detailed knowledge of the detectors and trigger electronics and on the amount of information available. The Global Calorimeter (GCT) and Global Muon (GMT) Triggers determine the highest-rank calorimeter and muon objects across the entire experiment and transfer them to the Global Trigger (GT), the top entity of the Level-1 hierarchy. The latter takes the decision to reject an event or to accept the event for further evaluation by the HLT. The decision is based on algorithm calculations and on the readiness of the sub-detectors and the DAQ, which is determined by the Trigger Control System (TCS). The Level-1 Accept (L1A) decision is communicated to the sub-detectors through the Timing, Trigger and Control (TTC) system. The architecture of the L1 Trigger is depicted in Figure 3.13. The L1 Trigger has to analyse every bunch crossing. The allowed L1 Trigger latency, between a given bunch crossing and the distribution of the trigger decision to the detector front-end electronics, is 3.2 μs. The processing must therefore be pipelined in order to enable a quasi-deadtime-free operation. The L1 Trigger electronics is housed partly on the detectors, partly in the underground control room located at a distance of approximately 20 m from the experimental cavern. Triggers can be inhibited through a feedback loop received from the front-end devices. While being operated, the front end devices assert synchronously their status in terms of readiness for data taking or buffer occupation near overflow. These states are transmitted through the so called Trigger Throttling System (TTS).

A detailed description of the L1 trigger algorithms can be found in reference [66].

The e/γ candidates are built by the Regional Calorimeter Trigger (RCT) which looks for the calorimeter towers with highest E_{T} deposits. The nearest neighbour with highest E_{T} is added to the e/γ candidate to estimate the candidate total energy. A selection on the lateral shower profile can be required, based on the information produced by the ECAL TPG and, in addition, a maximum threshold on the ratio between the energy deposited in ECAL and HCAL can be required. Other calorimetric candidates like jets, τ and missing energy are also built at the RCT level and finalized at the GCT level.

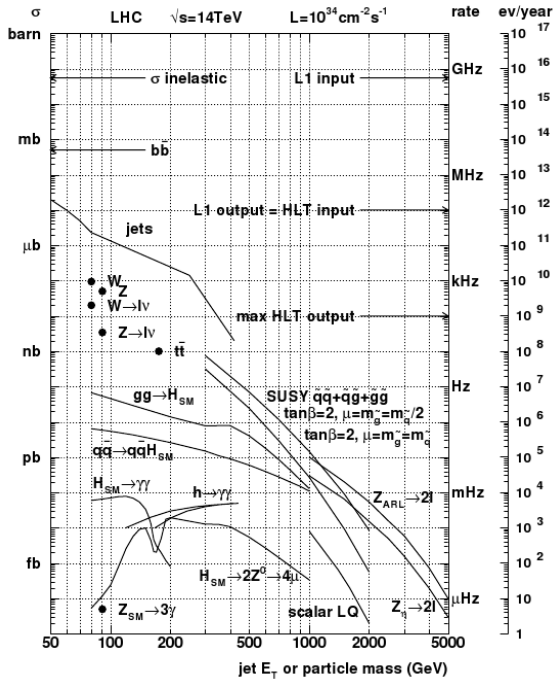


Figure 3.12: Production cross sections of some benchmark Physics process in pp collisions at $\sqrt{s} = 14\,\text{TeV}$. The corresponding event rate for an instantaneous luminosity of $10^{34}\,\text{cm}^{-2}\,\text{sec}^{-1}$ is shown on the right y axis. Figure reproduced from [66].

Muon candidates are built from tracks produced in the Muon System. Tracks are built by the DT and CSC track finders (DTTF and CSCTF), extrapolating track segment sources from one station to the next, according to pre-computed trajectories with their origin set at the vertex. The RPC trigger requires no local processing apart from synchronization and reduction of the cluster of hits. The RPC trigger is segmented in 33 towers along η which have a subdivision of 144 segments in ϕ. The information from the DTTF, CSCTF and RPC is received by the GMT which reduces the redundant information and combines it with information coming from the calorimeter trigger, in order to test the compatibility of the muon candidate with a minimum ionising particle.

Data acquisition system

The Data Acquisition (DAQ) system of CMS is in charge of collecting the data from ~ 650 data sources (see Table 3.1) at the nominal L1 trigger rate, combining the information and forwarding them to the HLT filter farm. Each data source provides event fragments of ~ 2 kB. The average throughput of the system is thus ~ 100 GB/s. A schema of the CMS DAQ system is shown in Figure 3.14.
In the CMS experiment each sub-detector frond-end system stores data in 40-MHz pipelined buffers which are readout synchronously on incoming L1A signals. The data is read out by bi-directional high speed optical links, operated at 800 Mbit/s by the so called Front-End Drivers (FED) . The FEDs are located in the underground counting room $\sim$ 70 m from the detector. The data of each FED is received by the Front-end Readout Link (FRL) boards. Each FED is plugged to a S-LINK 64 Sender Card which performs a Cyclic Redundancy Check (CRC) prior to transmission over the S-LINK 64 [68]. Data is transmitted over an LVDS cable with a maximum throughput of 400 MB/s. A buffer of 1.6 kB is used to sustain high instantaneous rates. If the buffer is filled, back-pressure is generated and automatically fed back to the TTS lines by the FED. Each FED is also responsible by formatting the data received

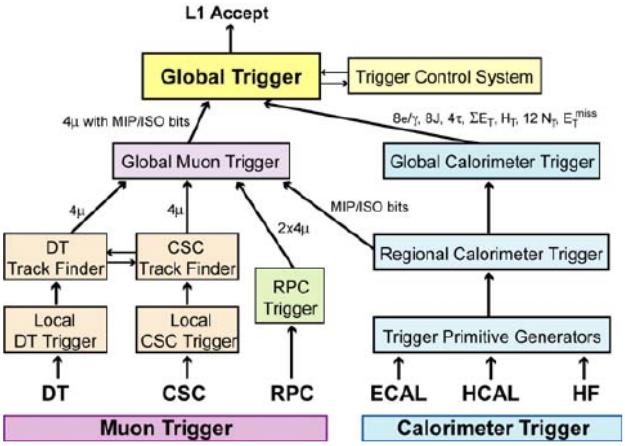


Figure 3.13: Schematic representation of the CMS L1 trigger system. Figure reproduced from [54].

Table 3.1: Sub-detector read-out parameters. Numbers reported from [54].

Sub-detector	N. channels	N. data links	N. data sources (FEDs)
nTracker pixel	$\sim$ 66 M	$\sim$ 1500	40
Tracker strips	$\sim$ 9.3 M	$\sim$ 36 k	440
Preshower	144384	1128	56
ECAL	75848	9k	54
HCAL	9072	3072	32
Muons CSC	$\sim$ 500 k	540	8
Muons RPC	192 k	732	3
Muons DT	$\sim$ 195 k	60	10
Global Trigger	n/a	n/a	3
CSC, DT Track Finder	n/a	n/a	2
Total	$\sim$ 76 M		626

in a common structure which contains a header and a trailer with relevant information to be properly merged into the full event. Although the payload of each data source is only looked into in detail at a higher level, the bandwidth occupied by each data source is limited and thus requires the use of filtering algorithms.

3.2.5 Computing

The total rate of data produced by the online trigger system is $\sim$ 230 MB/s. These data need to be stored for further processing and analysis. Given the large volumes involved, large computing power and storage space are required to efficiently extract Physics results. Also, due to the large size of the CMS collaboration, the users accessing the experimental data are distributed world-wide.
For these reasons, a fully distributed computing model is used for data reconstruction and analysis. The system is based upon Grid middleware, with the common Grid services at centres defined and managed through the Worldwide LHC Computing Grid (WLCG) project [70], a collaboration between LHC experiments, computing centres, and middleware providers.
A hierarchical architecture of computing centres in *Tiers* is employed. The flow of data after the HLT

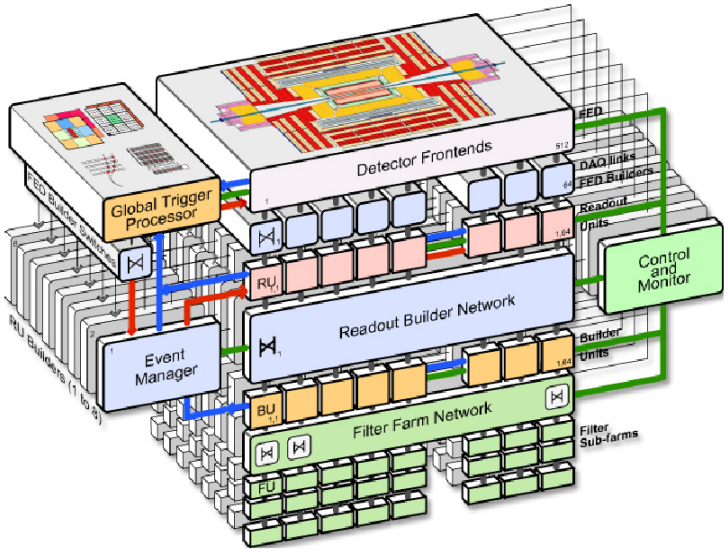


Figure 3.14: Schema of the CMS data acquisition system. Figure reproduced from [69].

is shown schematically in Figure 3.15. Data are classified by the online system in *Primary datasets*, according to the trigger algorithms that were active in each event. Primary datasets are defined in such a way that their overlap is minimised (typically to $\lesssim 10\%$). Data are received by the Tier-0 centre at CERN in the so-called *RAW* format, that contains the full recorded information from the detector, plus a record of the trigger decision.
The Tier-0 centre stores the data and proceeds to a first-pass reconstruction of the data. Reconstructed (*RECO*) data are produced by applying detector-specific filtering and correction of the digitised data, cluster and track-finding and primary and secondary vertex reconstruction, as well as particle identification algorithms. Reconstruction is the most CPU-intensive activity in the CMS data processing chain. The resulting RECO events contain high-level physics objects, plus a full record of the reconstructed hits and clusters used to produce them.
A copy of the RAW and RECO data is exported to the Tier-1 centres ($\sim$ 6 centres), which are used as mass storage and dedicated centres for re-processing of the RAW data.
A fraction ($\sim 10 - 30\%$) of the bandwidth between the online system and the Tier-0 is reserved for calibration and monitoring events. These events are promptly reconstructed at Tier-0 and the alignment, monitoring and calibration results are stored in a dedicated database (*conditions database*) which is used for event reconstruction at the Tier-0.
Tier-1 centres serve the data to the Tier-2 centres where the end-analysis is performed. Tier-2 sites ($\sim$ 50 centres) are also used for the production of Monte-Carlo simulated data and subsequent transfer to the Tier-1 sites. Data can be stored at Tier-2 centres in RECO format or in a lighter Analysis Object Data (AOD) format, designed to allow a wide range of physics analyses whilst occupying sufficiently small storage so that very large event samples may be held at many centres. AOD events contain the parameters of high-level physics objects, plus sufficient additional information to allow kinematic refitting.

3.3 Conclusions

This chapter was dedicated to the description of the LHC apparatus and the CMS detector. The design of both was briefly introduced.
The main operational parameters of the LHC machine were discussed. The first LHC Physics run started

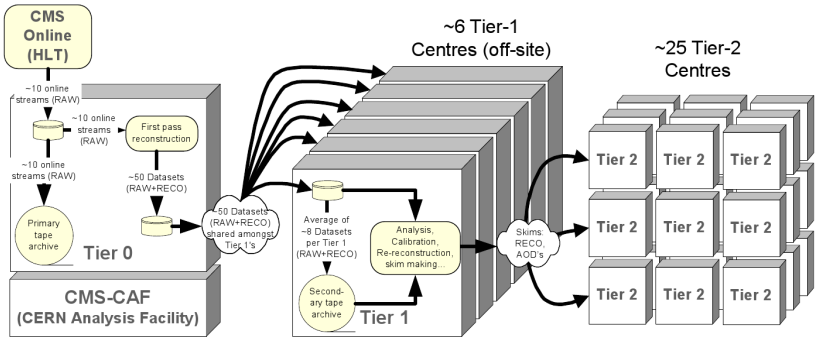


Figure 3.15: Schematic representation of the CMS computing model. Figure reproduced from [54].

at the beginning of 2010. The machine exhibited excellent performances, confirming that the challenging design goals have been successfully met. During the proton-proton run an integrated luminosity of ~ 45 pb^{-1} was delivered to the CMS experiment.
The CMS detector was designed to operate in unprecedented levels of radiation and particles density. It was designed to achieve good energy and momentum resolution in the kinematic regions most relevant for the LHC research program.
The detector was operated successfully during the first LHC run and it showed excellent performances. The alignment and calibration conditions well met the goals set for start-up and in many cases their precision exceeded the expectations. The detector capabilities in terms of reconstruction of Physics processes were observed to meet the expectations and in some case to be already close to design conditions. In the following two chapters, the CMS ECAL, which is the most important component of the CMS detector for the topic of this thesis, will be discussed in more details.

Chapter 4

Trigger and Data Acquisition of the CMS Electromagnetic Calorimeter

Contents

The CMS Electromagnetic calorimeter has been described in Chapter 3. As mentioned there, the detector is made by more than 76000 $PbWO_4$ crystals and their information is used by the CMS first level trigger processor.
The high granularity of the ECAL design implies an equally large amount of readout channels. The large data reduction factors required by the LHC event rates call for fast and efficient trigger algorithms and effective online reduction of the data volume. These tasks are accomplished by the ECAL trigger and data acquisition system (TriDAS).

Part of the work described in this thesis has been devoted to the development of the online software responsible for the operation and monitoring of the ECAL TriDAS system. The latter comprises roughly 3000 ASICs front-end boards, 230 VME boards and 1000 mezzanine boards. Roughly 8000 optical and 600 electrical links are used to transport the data.
The software system is in charge of configuring the $\sim 2 \times 10^6$ parameters of the system and of monitoring the status of the system during data taking. It also allows to acquire data for the purpose of detector calibration.

In this chapter the ECAL trigger and data acquisition system is briefly described and then the architecture and functioning of the online software is discussed in detail. The discussion focuses on the system used for the crystal readout. The software system is used also to drive the preshower data acquisition, which uses a configuration database schema derived from the one described below.

4.1 The hardware system

The ECAL TriDAS system [71] is physically divided in two sections: the on-detector and the off-detector electronics. The first is responsible for the signal digitisation and the trigger primitives (TP) production and consists of radiation tolerant ASICs located on the detector. The second is in charge of the finalisation of the trigger primitive calculation and of the readout and reduction of the full granularity data.
The on-detector electronics, named Front End (FE) in the following, comprise roughly 3000 boards with radiation-tolerant chips, developed in 25 μm CMOS technology.
The off-detector electronics are organised in 54 Readout Units, each comprising three types of VME boards: the Clock and Control System (CCS), the Trigger Concentrator Card (TCC) and the Data Concentrator Card (DCC). In order to limit the system output rate to a level acceptable by the CMS DAQ system, the data volume is reduced by means of a dedicated algorithm (Selective Readout algorithm), based on the classification of the detector in high and low interest regions performed by the Selective Readout Processor (SRP) [72].

4.1.1 Trigger and data-reduction algorithms

Before describing the details of the different electronic components, an overview of the algorithms that they implement is presented. The basic detector units used by the algorithms are the trigger towers, i.e. groups of 25 adjacent crystals. In the barrel, these crystals have a squared 5×5 topology in $\eta - \phi$ space, while for the endcaps they have more complicated topologies.

The ECAL data, in the form of trigger primitives, are sent to the Level-1 calorimeter trigger processor, at every bunch-crossing. Roughly 3000 words, representing the transverse energy deposited in each ECAL trigger tower, are sent to the Regional Calorimeter Trigger (RCT) at the 40MHz bunch-crossing frequency. The level-1 accept (L1A) signal is emitted by the global trigger processor within 3 μs.
The selected events are read out through the data acquisition system to the Filter Farm where further rate reduction is performed using the full detector data. The read-out system is structured in sets of 5×5 crystal matrices in both the barrel and endcap regions. The FE cards store the data, in 256 clock cycles deep memory banks, awaiting a Level-1 trigger decision during at most 160 bunch crossings after the collision occurred.

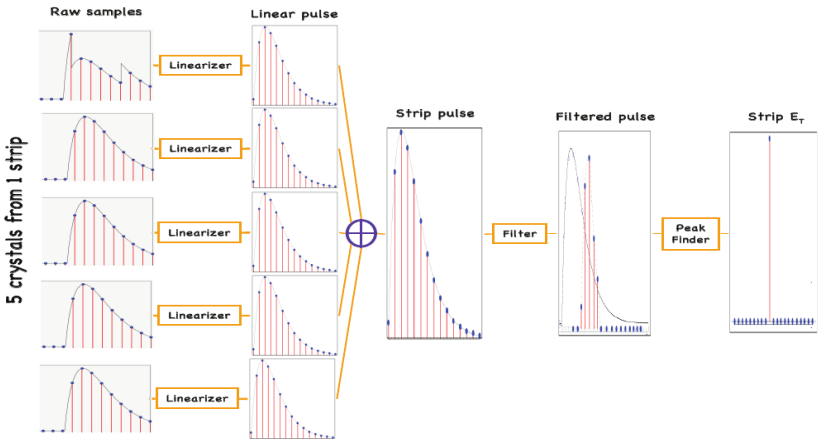


Figure 4.1: Illustration of the trigger primitive digital filter used for to the calculation of the strip transverse energy.

The algorithm implemented by the ECAL Trigger Primitive Generator (TPG) logic is illustrated in Figure 4.1. Five digital signals from the crystals (strip), subtracted of the mean amplifier pedestal, are multiplied by linearisation coefficients that take into account the electronics gain, the crystal calibration and the geometric factor to obtain the transverse energy deposit. The linearised signals of the five crystals are then summed up to produce the strip signal. A Finite Impulse Response (FIR) filter is applied to measure the amplitude of the strip pulse. The filter is based on linear weighted sums where the weights take into account the expected shape of the signal and subtract dynamically a possible residual pedestal. The output of the amplitude filter is then processed by a peak finder stage that keeps only the maximum as a measure of the transverse energy contained in the strip.
The signals of five strips are then added up to provide a measurement of the total transverse energy in a trigger tower. This step is performed by the FE chip in the barrel region and by the TCC board in the endcap one.
The dynamic range of the transverse energy is then reduced from 10 to 8 bits, using dedicated look-up tables, chosen to roughly reproduce the TP energy resolution. In parallel, the signals of the strips are combined in a Fine Grain filter producing 1 bit, the fine grain bit, indicating the transverse extent of the electromagnetic energy deposit. Thus, the ECAL trigger primitives are made, comprising 8 bits encoding the total transverse energy of the trigger tower and the fine grain bit. One such 9-bit word is transmitted for every trigger tower and every bunch crossing.
The trigger primitives are then delivered to the RCT after a constant latency of 52 clock cycles, of which 22 are used for transmission over optical fibres and cables.

About 100 kB per event are allocated for ECAL data. The ECAL event size, if all channels are read out, exceeds this target by a factor of nearly 20. Reduction of the data volume is performed by the Selective Readout algorithm, in such a way the that the suppression applied to a channel takes account of energy deposits in its vicinity, with the goal of achieving the best energy resolution for large energy deposits.
The calorimeter regions are classified according to the transverse energy deposited in the different trigger towers. This task is achieved by the TCC board which flags the trigger towers as low-, medium- or high-interest by comparing the TP with programmable transverse energy thresholds. For the 2010 run, the medium and high interest thresholds have been set to 1 and 2 GeV respectively.
Such flags are transmitted to the SRP system upon receipt of the L1A signal. The SRP then determines the readout mode to be applied to the different detector regions. Towers close to a high or medium interest tower are identified and different suppression levels, corresponding to high, low or no zero suppression can be applied to the different regions according to the configuration. The calculated flags are then sent

to the DCC boards that implement the zero suppression algorithm.
In the schema used during the 2010 run, only one zero suppression threshold, corresponding to $\sim 2.5\sigma$ of the noise is applied. Low interest towers are read-out in zero-suppression mode as well as towers next to medium-interest ones. Full readout is performed for high-interest towers and their neighbours, and for medium-interest towers.
For debugging purposes, the selective readout can be deactivated and either a global zero suppression (same threshold for every channel) or no zero suppression (full readout) applied. Even when the selective read-out is not applied, the selective read-out flags are appended to the data stream and can be used offline for validation purposes.

4.1.2 The Front End

The on-detector electronics has been designed to read a complete trigger tower in EB (5×5 crystals in $\eta \times \phi$) or a super-crystal in EE. It accomplishes the following functions:

- shape, amplify and digitise the signals generated in the photo-detectors;
- calculate the trigger primitives used by the CMS first level trigger processor;
- store the digitised signals during the first level trigger decision latency ($\sim 3\,\mu$s).

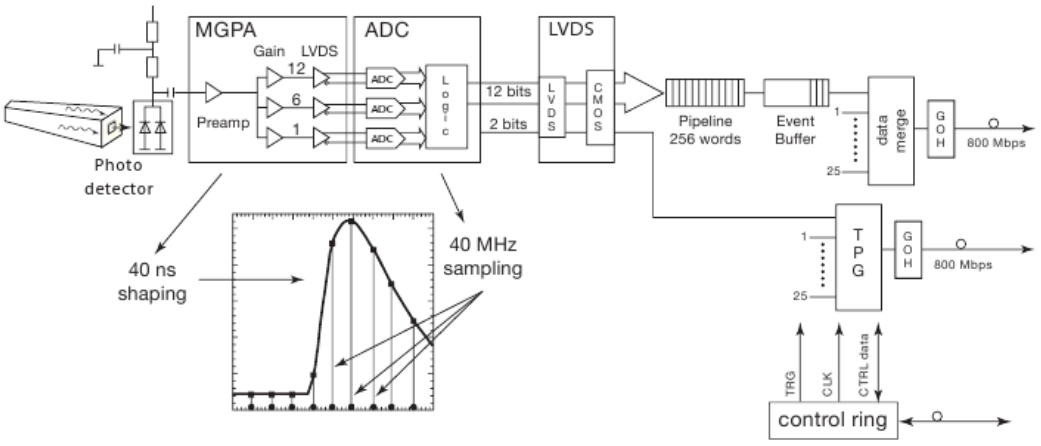


Figure 4.2: Schematic view of the on-detector electronics: scintillation light is collected by photodetectors (in the figure, a APD is presented), the signal is shaped by a Multi-Gain Pre-Amplifier and digitised by 40-MHz ADC; a radiation-hard buffer (LVDS) adapts the ADC output to the FE card, where data is pipelined and Trigger Primitives Generation (TPG) is performed; trigger words are sent every 25 ns, while data are transmitted on reception of a Level-1 trigger; GOHs provide in both cases the data serialiser and the laser diode, driving the signals through about 100 m of optical fibre to the off-detector electronics. A control token ring connects groups of FE cards, providing Level-1 trigger (TRG) and clock (CLK) signals, together with control data in and out (CTRL data).

It consists of five Very Front End (VFE) boards, one Front End (FE) board, two (EB) or six (EE) Gigabit Optical Hybrids (GOH), one Low Voltage Regulator card (LVR) and a motherboard.

The motherboard connects to 25 photo-detectors using Kapton flexible printed circuit boards (EB) or coaxial cables (EE). In the case of the EB, the motherboard distributes and filters the APD bias voltage. In the case of the EE, the operating voltages for the VPTs are distributed and filtered by a separate HV

filter card. One LVR and five VFE cards plug into the motherboard.

The signals are pre-amplified, shaped and then amplified in parallel by three amplifiers with nominal gains of 1, 6 and 12. This functionality is built into the Multi Gain Pre-Amplifier (MGPA) [73] ASIC. The full scale signals of the APDs and VPTs are 60 pC and 12.8 pC corresponding to $\sim$ 1.5 TeV and 1.6–3.1 TeV for EB and EE respectively. The shaping is done by a CR-RC network with a shaping time of $\sim$ 40 ns. The output pulse non-linearity is less than 1%. The MGPA contains three programmable 8-bit DACs to adjust the baseline of the ADC inputs. An integrated test-pulse generator with an amplitude adjustable by means of an 8-bit DAC allows the testing of the read-out electronics over the full dynamic range.

A schematic view of the signal read-out is given in Figure 4.2. The 3 analog output signals of the MGPA are digitised in parallel by a multi-channel, 40-MHz, 12-bit ADC [74]. An integrated logic selects the highest non-saturated signal as output and reports the 12 bits of the corresponding ADC together with two bits coding the ADC channel number. If the read-out switches to a lower gain as the pulse grows, the read-out is forced to continue using the lower gain for the next five samples.

The noise obtained with the VFE cards installed into supermodules is typically 1.1, 0.75 and 0.6 ADC counts for gains 12, 6 and 1 respectively. This corresponds to $\sim$ 40 MeV for gain 12. The FE card [75] stores the digitised data during the Level-1 trigger latency in pipelines that are 256 clock cycles deep. Five such pipelines and the logic to calculate the energy sum of the 5 channels for every bunch crossing are integrated into an ASIC developed in 0.25 μm technology called FENIX. Each VFE card is serviced by a FENIX chip, which sums the energy of 5 crystals along a strip in ϕ. In the case of the EE the five strip sums are transmitted by five GOHs to the TCC board, while in the case of the EB a sixth FENIX sums the five strip sums and calculates the fine-grain electromagnetic bit. The trigger tower energy sum together with the fine-grain bit are transmitted using one GOH to the TCC. Upon reception of a Level-1 trigger the corresponding data, ten 25 ns samples per channel, are transmitted in $\sim$ 7.5 μs to the DCC board using an identical GOH.

The VFE and FE electronics are controlled using a 40-MHz digital optical link system, controlled by the CCS boards. The FE boards are connected through a token-ring-type network [76]. A 12-fibre ribbon is connected to the token ring link board, which translates signals into an electrical control ring. The Clock and Control Unit (CCU) ASIC provides the FE with an interface to the token rings. Each supermodule has 8 token rings which connect to groups of eight to ten FE cards including the two FE cards of the laser monitoring electronics module (MEM). In the endcap, each 40° sector has 3 or 4 token-rings. The system has redundancy and is able to operate as long as there are no two consecutive FE cards malfunctioning. The token-ring connection provides fast and slow control functions. While the fast control transmits the level one trigger information and the 40-MHz clock, the slow control comprises the configuration of the FE and VFE electronics as well as the read-out of status information, temperatures, voltages and APD leakage currents.

4.1.3 The Off-Detector Electronics

The ECAL off-detector read-out and trigger architecture [71] is illustrated schematically in Figure 4.3. The system is located in the CMS service cavern and is composed of different electronic boards sitting in eighteen 9U VME crates (the read-out units comprising CCS, TCC and DCC modules) and in one 6U VME crate (the SRP system). The system serves both the DAQ and the trigger paths. In the DAQ path, the DCC performs data read-out and data reduction based on the selective read-out flags computed by the SRP system. In the trigger path, at each bunch crossing, trigger primitive generation started in the FE boards is finalised and synchronised in the TCC before transmission to the regional calorimeter trigger.

The off-detector electronics receives signals from the TTC system, namely the clock, the L1A signals plus the so-called B-Go commands, used for synchronisation and acquisition of special triggers. The

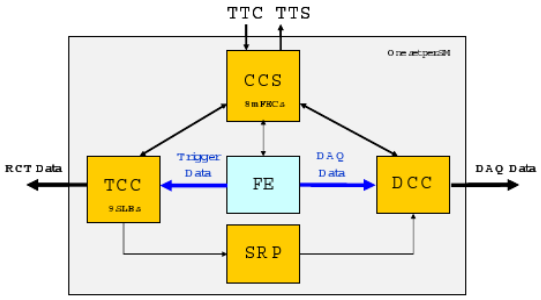


Figure 4.3: Schematic view of ECAL off-detector electronics.

TTC system is organised in several *partitions*. These receive a single TTC fibre from the CMS central trigger system. The signal is received by the TTCci boards [77], which distributes them to the whole partition. The TTCci boards can also receive inputs from the Local Trigger Controller [77] (LTC) board for debugging and calibration runs. The ECAL off-detector electronics are organised in four partitions, geometrically corresponding the negative an positive sides of the barrel and of the endcap detectors.

A feedback system, the Trigger Throttling System (TTS), is used in order to limit the trigger rate in case buffers may overflow. The TTS receives inputs from CMS Front End Drivers (FED). These inputs are processed by the Fast Merging Module (FMM) boards which combine them into a single signal that is transmitted to the Global Trigger Processor (GTP). The granularity of the TTS system matches the one of the TTC system. Thus one TTS signal per TTC partition is received by the GTP.

In the following, the general characteristics of the ECAL off-detector electronics are enumerated.

The Clock and Control System

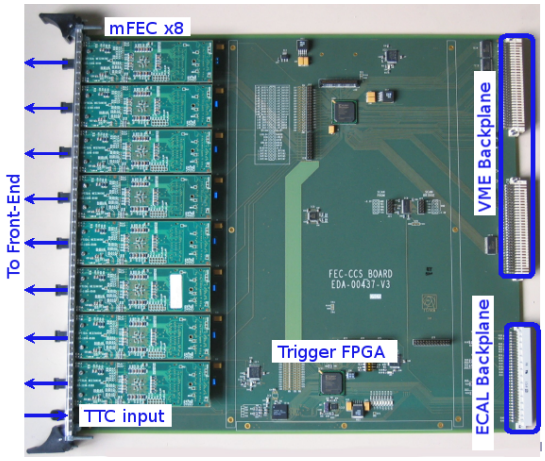


Figure 4.4: The Clock and Control System Board.

A picture of the CCS board is shown in Figure 4.4. It is a 9U VME board hosting up to eight *mezzanine* Front End Controller cards (mFEC). The board has one optical input to receive TTC signals and is electrically connected to the DCC and TCC boards through a dedicated back-plane. The board has three

main functions:

- distribution of the system clock and TTC commands to the on-detector electronics and to the other components of the readout units;
- configuration and control of the on-detector electronics;
- merging of the trigger throttling signals coming from the TCC and the DCC and the forwarding of the merged signal to the Trigger Throttling System (TTS).

The TTC signals transmitted to the FE are translated and encoded by suppression of clock edges and sent to the mFEC cards, which are optically interfaced with a FE token-ring. At the same time, the TTC signals are transmitted through the electrical back-plane to the TCC and DCC.
The board is equipped with a Field Programmable Gate Array (FPGA), which implements the TCC signal translation and TTS signals merging. A simple programmable trigger logic is also implemented and can be used to generate TTC signals autonomously from the TTC system.

The Trigger Concentrator Card and the Synchronisation Link Board

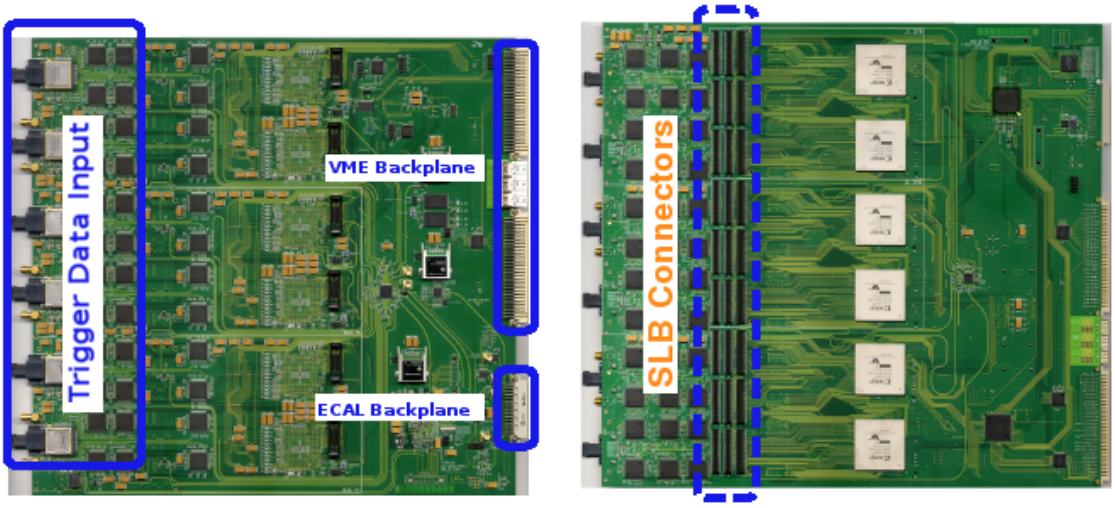


Figure 4.5: The EB Trigger Concentrator Card. The two sides of the board are shown.

Figure 4.6: The Synchronisation Link Board.

The TCC board (Fig. 4.5) is responsible for the reception and synchronisation of the TPs at the LHC bunch-crossing frequency of 40 MHz and their transmission to the Regional Calorimeter Trigger (RCT). It has three other important functions: the completion of the TP calculations, the classification of trigger

towers by categories of deposited transverse energy, required by the Selective Read Processor, and the storage of the trigger primitives during the Level-1 latency for subsequent reading by the DCC.
The transmission of the TPs to the RCT is delegated to the Synchronisation Link Board [78], represented in Figure 4.6. These mezzanine boards are connected to the TCC boards and are responsible for the alignment of the TPs with respect to the bunch-crossing time.
Each TCC collects trigger data from 68 FE boards in the barrel, corresponding to a supermodule, and from up to 48 FE boards in the endcaps corresponding to the inner or outer part of a 20°ϕ sector. In the endcaps, trigger primitive computation is completed in the TCCs, which must perform a mapping between the trigger data collected from different pseudo-strips from different supercrystals and the corresponding trigger towers.

The Selective Readout Processor

The selective read-out processor (SRP) is responsible for the implementation of the selective read-out algorithm. The system is composed by a single 6U VME crate with twelve identical algorithm boards (AB). The ABs compute the selective read-out flags in different calorimeter partitions. The flags are composed of 3 bits, indicating the suppression level to be applied to the associated read-out units.

The Data Concentrator Card

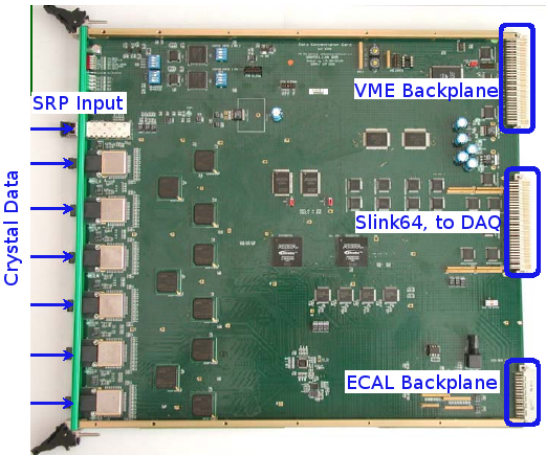


Figure 4.7: The Data Concentrator Card.

The Data Concentrator Card [79], shown in Figure 4.7, is the third component of the ECAL readout unit. Its specific task is the readout of the full-granularity crystal data upon a L1A decision which happens, in CMS, at the average frequency of 100 kHz. Before transmitting the data to the DAQ system, the board applies data reduction algorithms based on the regional classification performed by the SRP.
The signal amplitude for each crystal is calculated by means of a FIR filter and the zero-suppression thresholds determined by the SRP are applied.
In addition to the full-granularity crystal data, the TP generated for each tower are transmitted from the TCC to the DCC at L1A-time. At each L1A signal, the DCC merges the crystal-data together with the TP block and the SR flags to build an event fragment and transmits it to the DAQ system of CMS. In the barrel part each DCC reads the crystal data from one supermodule and the TPs from one TCC board. In the endcap part a DCC reads the data from a 40° sector and the 4 corresponding TCC boards. The DCC performs an extensive integrity check of the data received from the FE. Synchronisation errors, formatting or parity errors can be detected online and error conditions propagated through the Trigger

Throttling System (TTS). All data-integrity errors are flagged in the event header. A buffer protection system is also implemented and can activate the *Warning Overflow* and *Busy* signals over the TTS lines. DCC events are transmitted to the central CMS DAQ using a S-LINK64 [68] data link interface at a maximum data rate of 528 MB/s, with an average transmission data flow of 200 MB/s. Calibration triggers occur with a programmable frequency and synchronously with the LHC gap. No data reduction is applied for these events, which are read-out following a special TTC command. A VME memory is used for local DAQ, but also allows VME access to physics events and calibration events in spy mode.

4.2 The configuration database

The operation of the ECAL trigger and data acquisition system requires the configuration of $\sim 2 \times 10^6$ parameters. Part of the configuration parameters used in each run are needed in the data-reconstruction process and thus need to be stored for the long term. The full system configuration has to be reproducible a-posteriori for verification and debugging purposes.
The need to have fast retrieval and reliable mid- and long-term storage of these configuration data, naturally calls for the use of a Data-Base Management System (DBMS). The CMS collaboration has chosen to use an Oracle-based DBMS for the storage of configuration and condition data. An Oracle relational database schema has been implemented to store the ECAL TriDAS configuration data.
The configuration parameters have been sorted into three groups according to their function and update interval. The three groups are identified by *configuration keys*:

- The operation mode of the system is defined by the RUN_KEY. This defines the source of L1A triggers for the system, the type of data reduction algorithm to be applied, the latency of the data and trigger paths, the FE gain, as well as all registers defining the operation mode of the electronics. In the case of a calibration run, it also describes the list of parameters that are changed during the run.
- All the parameters pertaining to the DAQ path are defined by the FE_KEY. This specifies the values of the MGPA DAC, the FE time alignment registers, the zero-suppression FIR weights and the list of channels to be excluded from the read-out.
- All the parameters defining the TPG configuration are defined by the TPG_KEY. These include the crystal linearisation coefficients, the pedestal-subtraction coefficients, the TP FIR weights, the energy compression look-up tables (LUT), the fine grain bits transverse energy thresholds and the list of crystals and towers to be excluded from TP generation.

The total number of parameters configured is $\sim 2 \times 10^6$. The details are reported in Table 4.1. Each configuration key is associated with a table in the database, the rows of which are uniquely identified by the combination of a string (*tag*) and a version number.
A dedicate graphical user interface (GUI) has been developed for the modification of the database contents.

4.2.1 RUN_KEY

The RUN_KEY is associated to the ECAL_RUN_CONFIGURATION table (see Figure 4.8). This defines the type of run and the system operation mode.
Associated to the ECAL_RUN_CONFIGURATION table, the ECAL_SEQUENCE table allows to define several set of settings to be applied to the system during each run. This is particularly important for calibration runs where a manipulation of the system parameters (e.g. the electronics gain) is often required during the run.
The RUN_KEY parameters are fully specified by the ECAL_CYCLE table. The latter is linked to dedicated tables specifying the configuration of the different system resources. The following tables are defined:

Table 4.1: Number of configuration parameters associated to the different configuration keys of the ECAL TriDAS system.

		EB	EE	Total
RUN_KEY		37228	20758	57,986
FE_KEY	MGPA DAC	183,600	43,404	227,004
	Time aling.	2,448	579	3,027
	ZS FIR weights	367,200	86,808	454,008
	Total	555,696	131,369	687,065
TPG_KEY	TP FIR weights	379,440	73,620	453,060
	Lin. coeffs.	367,200	73,440	440,640
	Ped. subt.	183,600	36,720	220,320
	LUT	7,344	108	7,452
	FG	24,480	12,420	36,900
	Total	962,064	196,308	1,158,372
Total		1,554,988	348,435	1,903,423

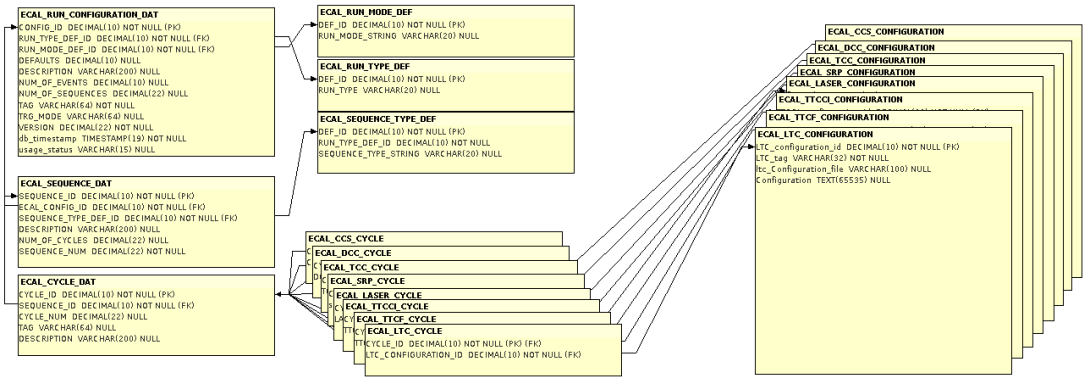


Figure 4.8: Diagram showing the contents of the ECAL_RUN_CONFIGURATION that contains information on the type of run and the system operation mode.

- ECAL_CCS_CONFIGURATION, defining the configuration of the CCS board registers and of global FE parameters.
- ECAL_DCC_CONFIGURATION, defining the configuration of the DCC board registers.
- ECAL_TCC_CONFIGURATION, defining the configuration of the TCC board registers.
- ECAL_SRP_CONFIGURATION, defining the configuration of the SRP system.
- ECAL_TTCci_CONFIGURATION, defining the configuration of the TTCci board.
- ECAL_LTC_CONFIGURATION, defining the configuration of the LTC board.

The resource configuration tables are linked to the ECAL_CYCLE through dedicated tables, which allows the schema to be easily extended to accommodate the configuration of more resources.
Several types of runs are implemented by the system and can be described by the RUN_KEY. The information about the type of run performed is stored in the event header and the reference to the configuration keys used is logged in the condition database. Two run-types are defined for data acquired in non-local (global) mode: *PHYSICS* for collision runs and *COSMIC* for cosmic ray runs. Besides these, a number of local calibration runs are defined:

PEDESTAL runs are used to read-out the amplifier pedestal levels in all gains, something that is not possible during Physics runs. Three cycles are implemented for this run type. In each cycle the electronics gain is forced to be 1, 6 or 12 and random L1A signals are generated to sample the electronics noise level.

TEST_PULSE runs are taken to exercise the full electronics chain and allow measuring the ratio of gains. A test charge, generated by the MGPA test circuit is injected in the amplifiers and the resulting signal read-out. The procedure is repeated for the three electronics gains.

LOCAL_CALIBRATION runs acquire the same sequence of events used for the online calibration monitoring. In particular laser, LED, charge-injection and pedestal events can be acquired. Calibration events are sent, in turn, to different parts of the detectors, in the same fashion as during collision runs. The TTC commands specifying the sequence of calibration events are generated by the TTCci or LTC boards.

PEDESTAL_OFFSET_SCAN runs are meant to scan the dynamic range of the MGPA baseline DAC in order to find the optimal settings on a crystal-by-crystal basis. The DAC settings are probed for each electronics gain using random L1A signals.

TEST runs are used for test and debugging purposes. The RUN_KEY schema allows full flexibility in the definition of these type of runs.

The list of run types known to the system is stored in a dedicated table (ECAL_RUN_TYPE_DEF) that can be easily extended.

4.2.2 FE_KEY

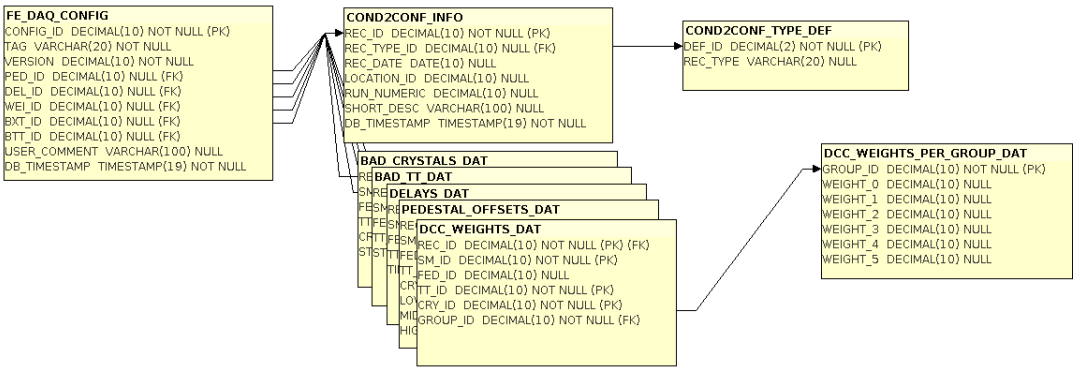


Figure 4.9: Diagram showing the contents of the FE_DAQ_CONFIG that combines information regarding values of the MGPA DAC, the FE time alignment, the zero-suppression FIR weights and the list of channels to be excluded from the read-out.

The FE_KEY is associated to the FE_DAQ_CONFIG table (see Figure 4.9). This table aggregates the references to the tables containing the values of the MGPA DAC, the FE time alignment registers, the zero-suppression FIR weights and the list of channels to be excluded from the read-out.
In order to optimise the volume of the stored data, each of the records above can be updated independently, following the evolution of the corresponding constants.
A dedicated table (COND_TO_CONF_INFO) stores meta-data associated with the update of each record.

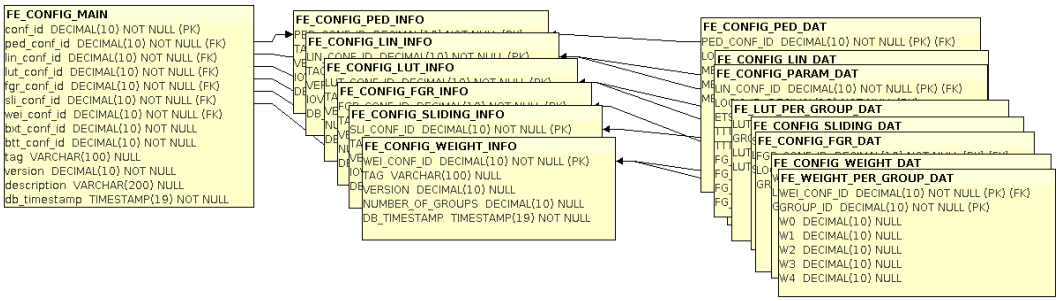


Figure 4.10: Diagram showing the contents of the FE_MAIN_CONFIG that combines information regarding the parameters needed in the trigger primitive generation process.

4.2.3 TPG_KEY

The data structure associated to the TPG_KEY is very similar to the one used for the FE_KEY. The table FE_MAIN_CONFIG collects references to the tables with the crystals linearisation coefficients, the pedestal-subtraction coefficients, the TP FIR weights, the energy compression LUT, the fine grain bits transverse energy thresholds and the list of crystals and towers to exclude from TP generation.
Also in this case, each of the aggregated records can be updated independently and dedicated tables store the meta-data associated to the update of each record.

4.3 The Online Software

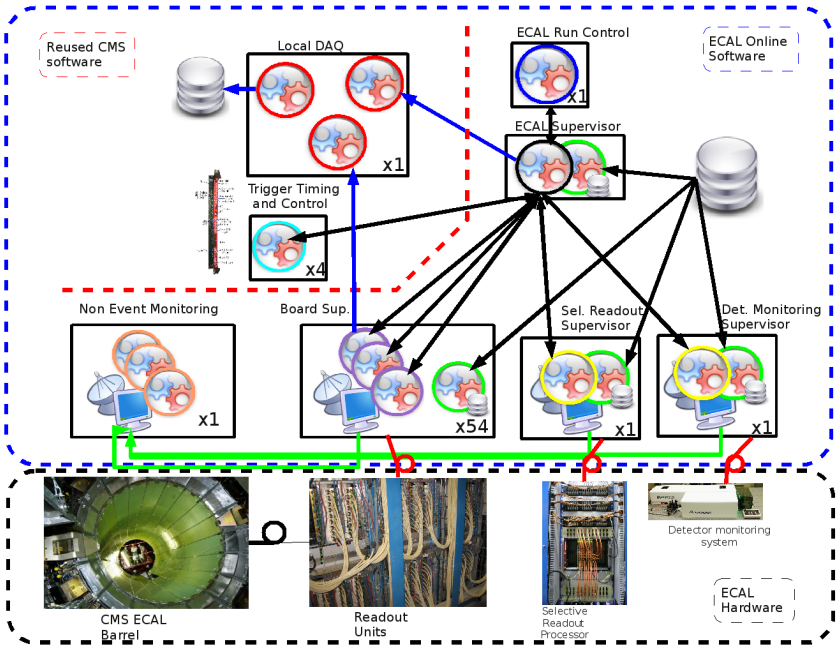


Figure 4.11: Overview of the CMS ECAL trigger and data acquisition system.

The ECAL online software is responsible for the configuration, operation and monitoring of the ECAL

trigger and data acquisition system. The software has been developed as a modular distributed system based on web technologies. The implementation is based on common frameworks developed in the context of the CMS experiment. Most of the components are based on the C++ XDAQ framework [80], while the top layer of the system is based on the JAVA RCMS [81] framework.

The system accomplishes several tasks: it configures the system hardware according to the parameters in the configuration database; it configures the ECAL calibration monitoring system; it monitors the system hardware during data taking and reports error conditions to the user; it integrates the ECAL sub-system in the CMS TriDAS system; and it can perform calibration runs in "local DAQ mode", i.e. independently of the CMS DAQ infrastructure.

The system is largely based on the C++ XDAQ framework, which provides the basic services needed by the ECAL online software, such as:

- inter-process communication, by means of SOAP-based communication channels;
- a data transport service, based on the I2O [82] binary protocol;
- an hardware access library;
- a vendor-independent database access service;
- a distributed logging, monitoring and error-reporting infrastructure.

The system, sketched in Figure 4.11 has a hierarchical organisation. Several applications participate to the system:

ECAL Function Manager. It is the top layer of the system. It is a web application based on the RCMS framework, responsible for managing the life-cycle of the lower layer applications and for configuring the local DAQ event builder. It also provides the user interface to the system.

ECAL Supervisor. It is the application responsible for the configuration of the Electromagnetic calorimeter. It receives commands from the ECAL run control and dispatches them to the applications responsible for the configuration of the different parts of the system, in sequence or in parallel, as required.

Resource Supervisors. For each of the components of the system, dedicated applications, generically called *resource supervisors* have been implemented. These applications receive the configuration parameters from the database, according to the directives coming through the ECAL supervisor, and are responsible for the configuration and monitoring of the system resources.

Middleware applications. In order to exploit the services provided by the XDAQ framework, dedicated modules, used by all applications in the system have been implemented. These modules are responsible for the hardware discovery (*Crate Scanner*), mutually exclusive resource locking (*Resource Locker*) and to retrieve the parameters from the database (*DB Manager*).

Monitoring system. During data taking, all system components are permanently monitored in order to detect and diagnose possible arising problems. The monitoring system has been built using the monitoring capabilities of the XDAQ framework and is integrated with the rest of the system.

The software runs on a cluster of rack-mounted PCs located in the CMS service cavern. Eighteen machines are connected to the VME crate controllers and host the resource supervisors in charge of operating the Readout Units. One machine is connected to the 6U VME crate hosting SRP system. One PC is dedicated to the client-side of the monitoring system. Three PCs host instances of the ECAL supervisor and the local DAQ system and one PC is dedicated to the storage.

4.3.1 Software architecture

All operations performed on the ECAL TriDAS system originate in the ECAL function manager application, in response to user input or to a request of the CMS run control. A SOAP-based communication protocol allows to then propagate the information to all system components. In particular, operations performed on the hardware system are delegated to the ECAL supervisor, which propagates the commands to the resource supervisors, as illustrated by the diagram in Figure 4.12. The set of operations performed on the hardware is fully specified by the list of detector components to be operated (FED_VECTOR) and the list of configuration keys (RUN_KEY, FE_KEY, TPG_KEY).

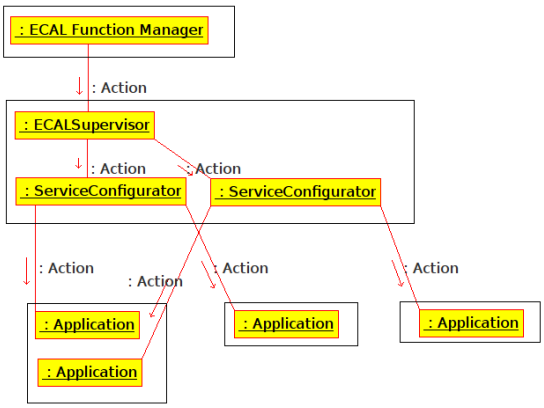


Figure 4.12: Collaboration diagram showing a set of actions executed upon a state transition.

The system has two modes of operation: *global* and *local* DAQ mode. In the first case, L1A signals are generated by the CMS first level global trigger processor and the data are collect by the CMS DAQ system. In the second case, L1A signals are generated by the TTCci or LTC boards and local instances of the CMS RUBuilder [69] and Filter Farm software, controlled by the ECAL function manager, are used to collected the data, acquired through the VME interface of the DCC boards from their internal memories.

The building blocks of the system are the ECAL supervisor and the resource supervisors. Both are web applications built on top of the XDAQ framework. A common model is followed in the implementation of all the applications. A standard Finite State Machine (FSM), which receives inputs from a SOAP-based interface, constitutes the core of each application. Specific hardware drivers are instantiated by each application in order to operate system resources. A standard monitoring facility allows applications to produce data and distribute it through the XDAQ Monitoring and Alarming System (XMAS).
The FSM implemented by all applications is illustrated in Figure 4.14. The goal of the FSM is to standardise the different steps necessary to prepare the system for data taking.

- Upon instantiation, all applications' FSM are in the *Initial* state. When an *Initialise* state transition is triggered, the software tests the connection to all hardware components and initialises all the required hardware drivers. The hardware is uniquely identified by the host to which the VME crate is connected and the slot occupied in the crate. The information about the available hardware resources is propagated to the ECAL supervisor and ECAL function manager level.

- The *Initialise* transition brings the system to the *Halted* state. In this state, the information about the available hardware components has been collected, all hardware drivers initialised, and the system is ready to be configured. When a *Configure* state transition is triggered, all the hardware

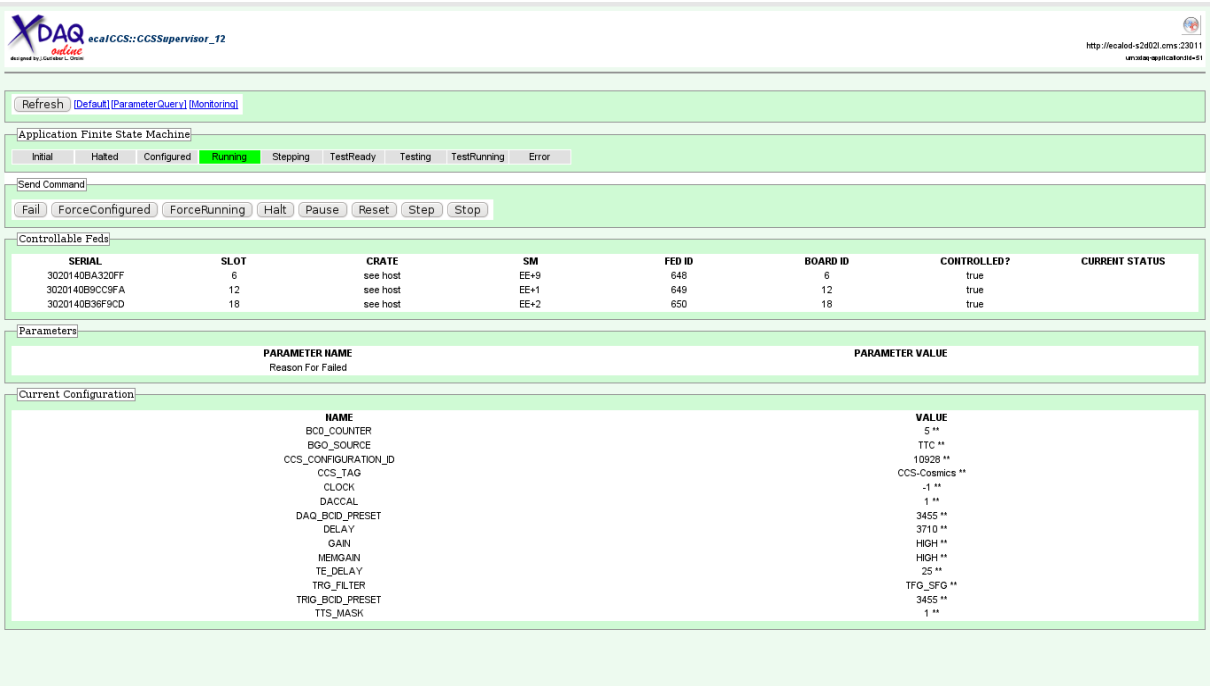


Figure 4.13: Web user interface of the ECAL online application.

configuration parameters are read from the database and all hardware registers set accordingly. Each application verifies that the configuration of the corresponding hardware has been successful. The system monitoring is enabled.

- Upon success of the *Configure* transition, the system is brought to the *Configured* state. In this condition the system is ready to start generating and accepting triggers. When the *Start* state transition is triggered, the status of the hardware is re-assesed and all counters cleared.
- As a result of the *Start* transition, the system goes to the *Running* state. In this state, trigger primitives are generated and sent to the CMS first level trigger and data delivered to the CMS DAQ upon reception of L1A signals. During this phase the hardware receives commands from TTC system and reports possible error conditions through the TTS system. The software performs a periodic monitoring of the hardware status. From the *Running* state, a *Stop* transition can be triggered. In this case, triggers are no longer accepted by the hardware and the system returns to the *Configured* state.
- From the *Running* and *Configured* states, the *Halt* transition can be triggered, which causes the system to stop accepting triggers and to return to the *Halted* state.
- If a fatal error condition is raised during any state transition, the system is brought into the *Failed* state.

In "local DAQ" mode, the system is able to change only some of the configuration parameters during the data-taking. Additional states are used in order to allow such parameters to be modified. From the *Running* state, a *Step* transition can be generated. No L1A signals are delivered to the hardware and the system is brought to the *Stepping* state. The hardware configuration is modified and then the conditions to continue the data taking restored. The system is brought back to the *Running* state and L1A signals re-enabled.

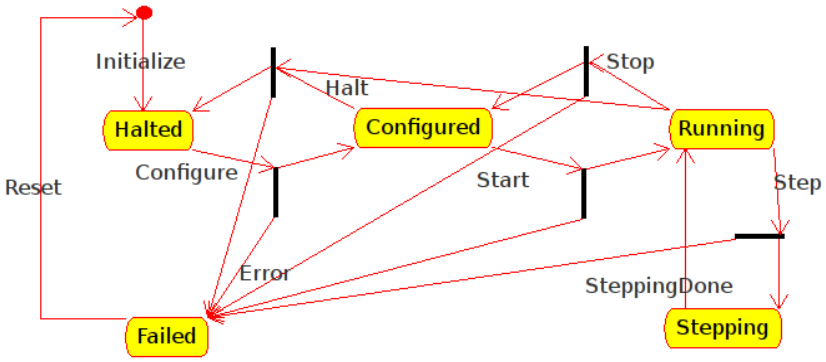


Figure 4.14: Finite State Machine implemented by the ECAL online applications.

A mutually exclusive resource locking mechanism, based on the linux™ Inter Process Communication (IPC) facility, has been implemented. This allows to run several instances of the system in parallel without risks of interference between the different sessions. A dedicated web application, called *Resource Locker*, has been developed with this purpose. An instance of such application is run on each machine by each software session, allowing the resource supervisors to request exclusive control of the resources.

A basic set of classes have been developed in order to facilitate the implementation of the application model. Figure 4.15 shows the most important ones.

- The *ecal::Application* class is the basic building block of the architecture. It defines the common FSM and exports the SOAP interface for its operation.
- The *ecal::I2OApplication* class implements the interface to the CMS RUBuilder, needed for the local DAQ mode.
- The *ecal::Monitorable* class allows to export monitoring data to the XMAS system;
- The *ecal::ApplicationUI* class implements a web HTML interface for operation and monitoring (Figure 4.13);
- The *tools::ConfigUtilities* class is used by the applications to access the system configuration data.

4.3.2 The Resource Supervisors

Many applications have been implemented to manage the different resources operated by the ECAL online software. These applications are generically called resource supervisor and are ultimately responsible for the operation and monitoring of the different system components.

Each type of VME board in the system has an associated resource supervisor. The system comprises:

- 54 *DCCSupervisor* applications, each controlling one DCC boards;
- 18 *CCSSupervisor* applications, each controlling 3 CCS boards;
- 36 *TCCBarrelSupervisor* applications, each controlling 1 TCC board in the barrel region;
- 18 *TCCEndcapSupervisor*, each controlling the 4 TCC boards (corresponding to a 40° sector) in the endcap regions;
- one *SRPSupervisor* application, controlling the SRP system;

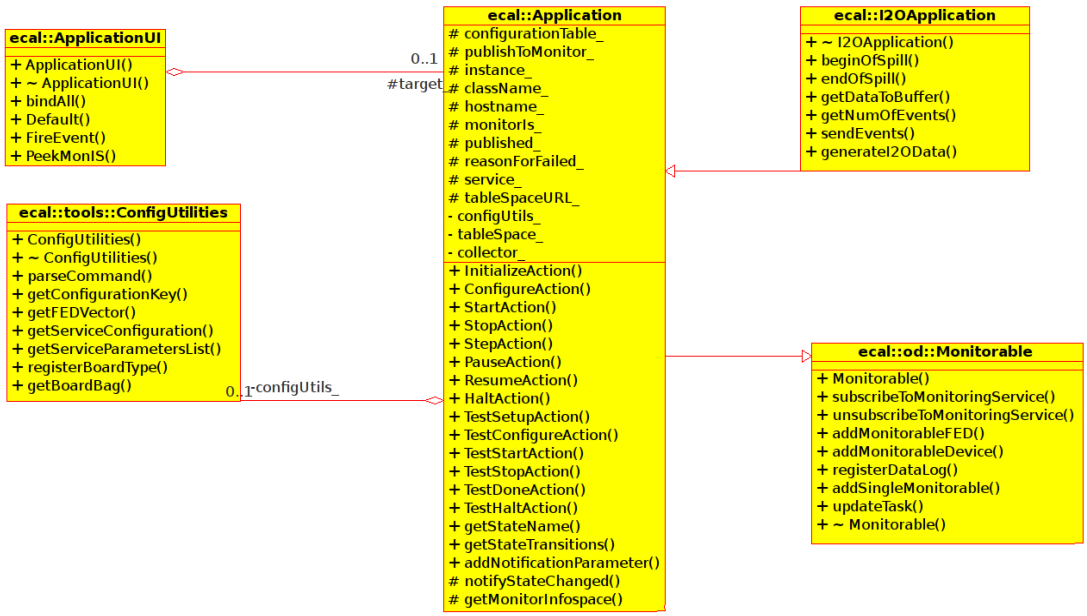


Figure 4.15: Class diagram showing the basic components of the ECAL online applications.

- one *LaserSupervisor* application, controlling the calibration monitoring system.

All these applications implement the common FSM described in the previous section and provide monitoring data for the controlled resources.

4.3.3 The ECAL Supervisor

The ECAL supervisor application is responsible for the configuration of all resources controlled by the ECAL online software. These include the VME readout-units, the SRP, the calibration monitoring system and the local TTC interfaces. All of these resources are controlled by XDAQ-based web applications, most of which are part of the ECAL online software.

The ECAL supervisor receives the configuration keys and the FED_VECTOR from the ECAL Function Manager and dispatches commands to other applications in the system. During the *Initialise* state transition, a map of the system is built. Applications controlling the same type of resources are organised in groups called *services* and detector components are mapped to each application. A global list of resources is exported to the ECAL function manager.

During the *Configure* state transition, the application determines which services need to be operated, based on the contents of the RUN_KEY and also determines which applications to target, based on the requested FED_VECTOR. SOAP messages are then dispatched to the concerned applications to trigger the FSM state transitions. This happens asynchronously and in parallel for all resource supervisors of a given service.

The communication with the controlled applications is performed through a set of dedicated classes, called *ServiceConfigurators*. These are in charge of compiling all the information needed by each type of resource for each operation. A common interface, providing methods for each FSM transition is implemented by these classes. This allows to easily cope with the different types of resources in the system and provides a simple mechanism to extend the application's functionalities.

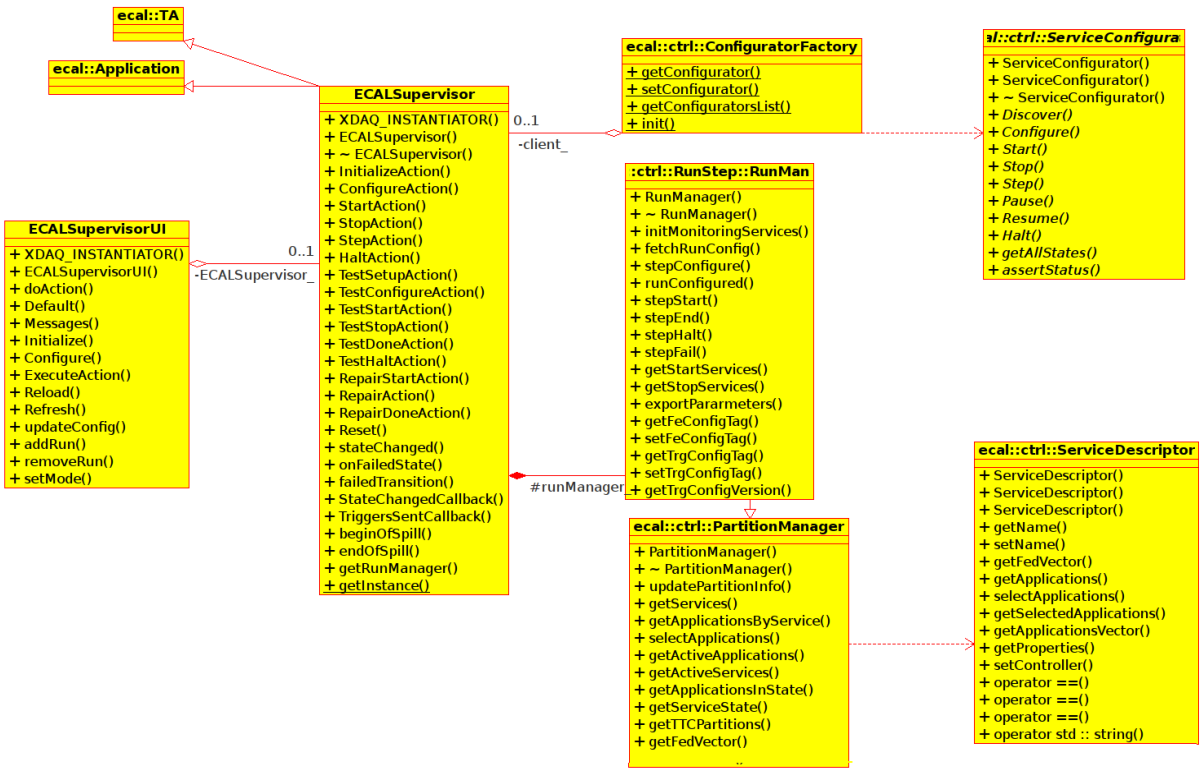


Figure 4.16: Diagram showing the ECALSupervisor implementation.

At each *Start* transition, the ECAL supervisor receives the run number from the ECAL function manager, dispatches it to all resource supervisors and creates a record in the CMS Condition Database to log the set of configuration keys and FED_VECTOR used. This ensures that the system configuration used in any run can be reproduced at any time.

A class diagram showing the details of the ECAL supervisor application implementation is shown in Figure 4.16:

- The *ECALSupervisor* class implements the application FSM and inherits from the *ecal::Application* class.
- The *ECALSupervisorUI* class provides an HTML web interface for control and monitoring.
- The *RunManager* and *PartitionManager* classes are in charge of managing information concerning the system configuration and the resources respectively.
- The *ConfiguratorFactory* class implements a registry of the available *ServiceConfigurators*.

Figure 4.17 shows a picture of the combined user web interface of the ECAL supervisor and ECAL function manager applications. Such interface reports the status of the system applications and of the resources being configured.

4.3.4 The Local DAQ system

A dedicated facility has been developed to allow the read-out of the ECAL detector without the need of the CMS DAQ system. Such facility is called *local DAQ* and it is used to acquire calibration data.

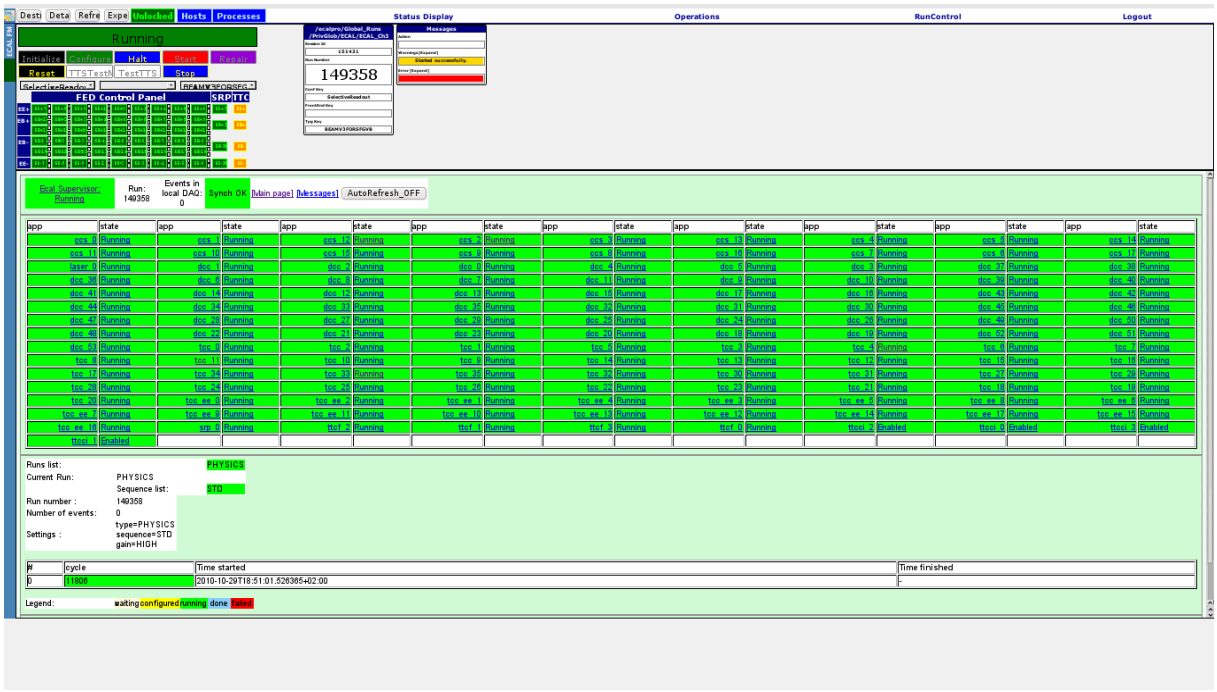


Figure 4.17: Combined web user interface of the ECAL supervisor (bottom part) and ECAL function manager (top part) applications.

In local DAQ mode, the L1A triggers are generated using the LTC or the TTCci boards and the data are stored in a dedicated memory of the DCC board, read-out through the VME interface.

The local DAQ system reuses the software developed for the CMS *RUBuilder* and *Filter-Farm*. A small instance of these components is run on the ECAL cluster machines and is controlled by the ECAL function manager.

A set of classes have been developed in order to implement the sender side of the RUBuilder data protocol. The data readout through the VME from the DCC boards is formatted and sent to the CMS RUBuilder using the I2O binary protocol. The CMS RUBuilder collects the data fragments coming from the DCC boards in order to build complete events. The latter are then forwarded to the Filter-Farm software which proceeds with the storage. The produced data format is 100% compliant with the standard CMS data format and the standard CMS reconstruction software can be used to process the acquired runs.

The local DAQ system is driven by the ECAL supervisor and its operation can be described as follows:

1. The system is configured with an appropriate RUN_KEY. The latter describes the sequence of system parameter sets to be applied to the system and the number of events to be acquired for each set.

2. At the end of the *Start* action, a programmable number of triggers is generated using the LTC or TTCci board.

3. When all triggers have been sent, the data are read-out from the VME to the PC memory.

4. A synchronisation check is performed in order to ensure that all DCC boards have received the same number of triggers. In case of a desynchronisation, the data read are discarded;

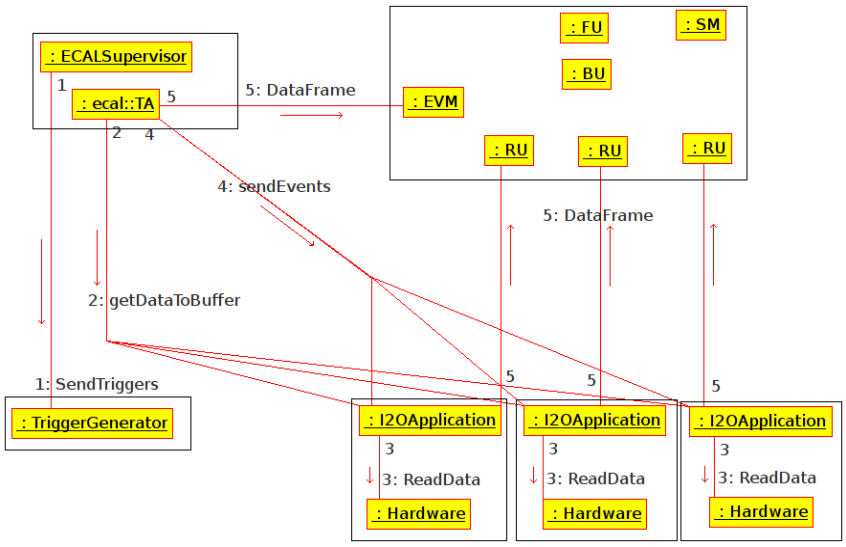


Figure 4.18: Collaboration diagram showing the local DAQ system operation.

5. All data are sent to the RUBuilder;
6. The system configuration is switched to the next set of parameters (by means of the *Step* FSM transition) and more triggers are sent.
7. The process continues until all requested data have been acquired. When this condition is satisfied, the system is automatically brought to the *Halted* state.

Steps 3–5 are described by the collaboration diagram in Figure 4.18.

4.3.5 The online Database Interface

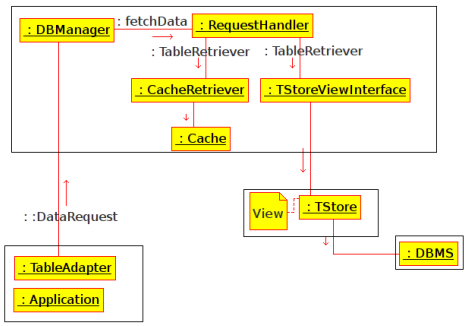


Figure 4.19: Collaboration diagram showing the query mechanism of the DBManager.

The data required for the system configuration are stored in the CMS online database service. At runtime, it needs to be retrieved by the ECAL online software.

A dedicated database access layer has been developed and integrated in the system. It is based on the XDAQ database access facility. The latter consists of a SOAP-based protocol, the server-side of which is implemented by the *TStore* web application. The latter provides a vendor-independent database access. A set of XML configuration files (*views*) contains the SQL statements needed to retrieve the data from the DBMS and specific classes implement the vendor-specific layer. The data transport is achieved through the use of binary attachments to the SOAP messages. A class called *xdata::Table* is used for this purpose.

In the ECAL online software, a dedicated application, the *DBManager*, has been implemented with the purpose of interfacing the *TStore* application. The additional layer allows to implement a cacheing mechanism for the configuration data and to easily switch to a different database access service.

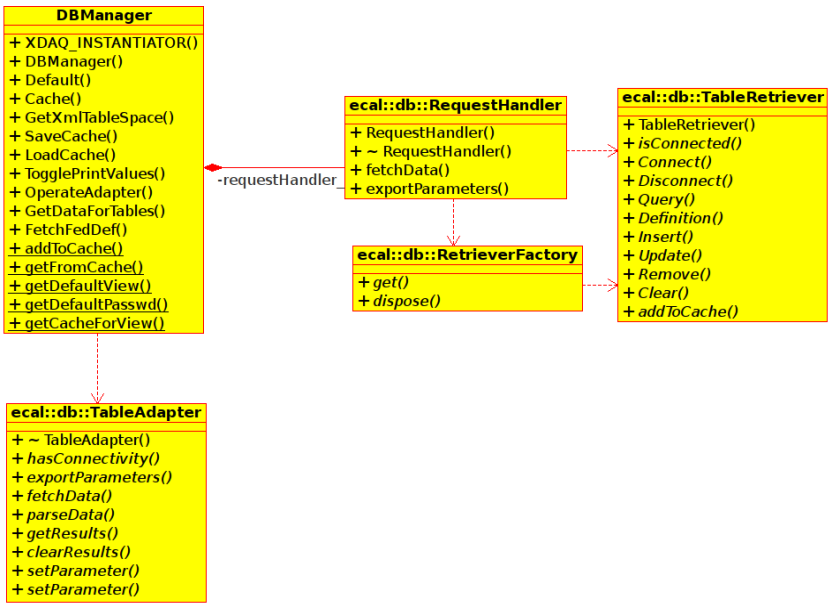


Figure 4.20: Class diagram showing the DBManager application implementation.

All applications in the system, namely the ECAL supervisor and the resource supervisors, communicate with the DBManager in order to retrieve the configuration data. This communication is performed through a SOAP-based protocol. The client side of this protocol is implemented by a set of classes called *table adapters*, which convert the configuration data from the *xdata::Table* to high-level C++ data structures used by the applications.

A diagram showing the implemented model is shown in Figure 4.19. A *DataRequest* SOAP message is sent to the *DBManager* application: the message payload specifies the type of query requested. A flag is set to specify whether the data can be extracted from the cache or should be retrieved directly from the database. The processing of the requests is delegated to the *RequestHandler* class which instantiates classes that allow to retrieve the data from cache or trough other services, in particular the *TStore* one. If the cache search is unsuccessful other services are tried, with a priority that can be specified by the client.

Figure 4.20 reports a class diagram showing the details of the *DBManager* class implementation. The application exports the SOAP interface to implement the server side of the *DataRequest* protocol. Data requests are delegated to the *RequestHandler* class, which satisfies them through the use of classes implementing a common *DataRetriever* interface. The *DBManager* application exports a set of HTML web pages for monitoring and implements an interface for the test of the *table adapter* classes.

4.4 System performance

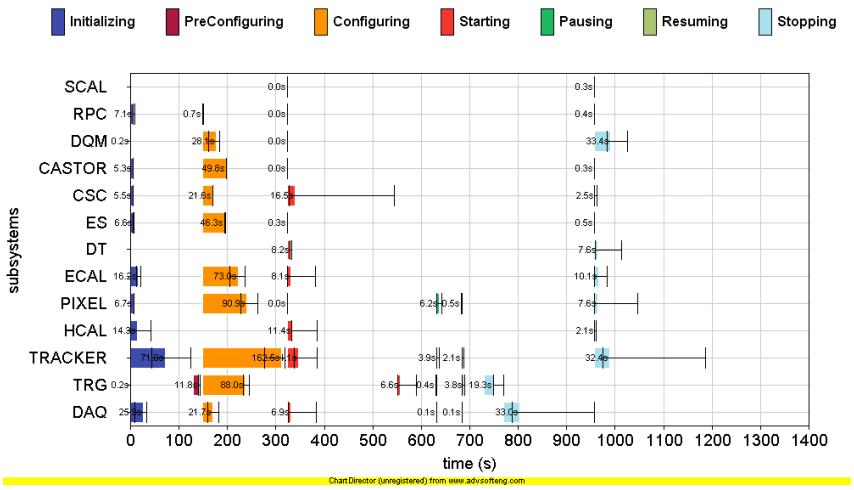


Figure 4.21: Measurement of the time taken by different CMS DAQ subsystems to complete the FSM transitions during beam-on periods in the 2010 LHC run. Figure reproduced from [83].

The data acquisition system of the CMS ECAL has been an essential tool for detector commissioning. The development of the ECAL online software started with the Barrel electronics integration. The system constantly evolved to follow the detector construction and commissioning. All the on-detector and off-detector electronics have been tested and validated during the integration campaigns. After construction, a large fraction of the ECAL was pre-calibrated with electron beams and all the Barrel using cosmic rays, using the DAQ system. During the in-situ installation, detector functionalities were carefully verified employing the DAQ system. Over the time, the system has been optimised, made more robust and has always been operated with very good efficiency.
The time required for the first system configuration in each session is $\sim 86\ s$. Subsequent configurations take $\sim 63\ s$. The 30% reduction in the configuration time budget is achieved partly thanks to the

cacheing of the configuration information performed by the DBManager application and partly because the initialisation of the FE token rings is performed only on the first configuration.
During the 2010 proton run, the average configuration time of the ECAL during beam-on time has been of ~ 74 s (see Figure 4.21). Such time budget is similar to the one needed for the configuration of other CMS sub-detectors with similar complexity. The availability of the system during beam-on periods has been $\sim 99.5\%$[1].

4.5 Conclusions

The design and architecture of the CMS ECAL Trigger and Data Acquisition System was presented in this chapter. The system hardware and the implemented algorithms were described. The database schemata used to store the system configuration data were discussed as well as the architecture and implementation of the online software.
The system software has been implemented in order to achieve large flexibility, efficiently manage the system configuration and reliably monitor its status.
The data acquisition system of the CMS ECAL has been an essential tool for detector commissioning. It followed the detector development through the construction process, the in-situ commissioning and collision running at the LHC.
The level of performance achieved during the 2010 LHC run confirms that the design goals of the system have been met. The time required for the configuration of the ECAL varies between ~ 86 s and ~ 63 s and the system availability in beam-on periods during the 2010 LHC run was 99.5%.

[1] computed using informations from `https://cmswbm.web.cern.ch/cmswbm/cmsdb/servlet/FillSummary`

Chapter 5

Performance of the Electromagnetic Calorimeter

Contents

During his PhD work the author of this manuscript had the privilege to contribute to the commissioning process of the CMS detector. He had the fortune of participating to the last phase of the construction of the ECAL and to all commissioning steps between the in-situ installation of the calorimeter and the start-up of the LHC.

It is hard to render in written words the formative value of such an experience or to summarise the great number of studies performed in order to insure the highest possible quality of the detector. In this chapter, two examples of such studies will be discussed. The author of this manuscript gave a personal contribution to both studies, which are related to the Physics subject of this thesis.

The first study was performed using the data taken during the 2008 CMS Cosmic ray run and assesed the performances of the first level CMS electromagnetic trigger. The second was performed on the data from the 2009 LHC commissioning run at $\sqrt{s} = 900$ GeV and pertains to the study of the reconstruction algorithms of electromagnetic objects.

5.1 Performance of the Level-1 electromagnetic trigger in Cosmic Ray events

During October-November 2008, the CMS collaboration conducted a month-long data taking exercise, known as the Cosmic Run At Four Tesla (CRAFT), with the goal of commissioning the experiment for extended operation [84]. With all installed detector systems participating, CMS recorded 270 million events triggered by cosmic rays with the solenoid at its nominal axial field strength of 3.8 T.
During the CRAFT, the calorimeter triggers were configured to trigger on instrumental noise and energy deposited by cosmic rays. Only the ECAL barrel was used to provide e/γ triggers, since the ECAL endcap trigger electronics were not yet installed at the time. The ECAL trigger primitive transverse energy was sent to the RCT on a linear scale, with a least significant bit (LSB) corresponding to 250 MeV, the maximum possible value being 63.75 GeV. In order to minimise the contribution from noise, trigger primitives below 750 MeV were suppressed. This value corresponds to between three and four times the noise. For most of the CRAFT, the electromagnetic trigger was active.
The full Regional Calorimeter Trigger was used during CRAFT. The default RCT configuration used during cosmic ray runs produced e/γ candidates from ECAL barrel trigger primitives, and region sums from the sum of ECAL (barrel) and HCAL (barrel, endcap and forward) trigger primitives. Noisy or absent ECAL and HCAL channels were masked in the RCT LUTs. The RCT input LUTs were generated to give linear E_{T} with a 250 MeV LSB. The isolation, "fine-grain", and H/E criteria were ignored in the production of e/γ candidates. The E_{T} of the e/γ candidate, which is the sum of E_{T} in a pair of contiguous ECAL trigger towers, was transmitted to the global calorimeter trigger (GCT) and thence to the GT using a linear scale with a LSB of 500 MeV. This algorithm is referred to as the e/γ trigger (EG), although the requirement is simply of an ECAL energy deposit above a configurable threshold.
The performance of the barrel e/γ trigger was evaluated in terms of rate, resolution and efficiency. The distribution of the rate of the L1_SingleEG1 e/γ trigger, which nominally fires whenever a single electromagnetic energy deposit above 1 GeV is detected, is shown in Figure 5.1 for a typical CRAFT run. The L1 decision is based on the sum of 2 ECAL trigger towers. The average single crystal noise as measured during the entire running period was 40 MeV [85], so the L1 candidate noise is expected to be around 280 MeV. On the other hand, a rate of 22.65 Hz implies that the threshold of the L1_SingleEG1 trigger is roughly 5σ away from the detector noise.

5.1.1 Data Selection

Of the 270 million cosmic-ray events recorded by CMS at 3.8 T during the CRAFT, a total of 246 million were used for ECAL reconstruction and analysis. Of these, 158 million events were taken with

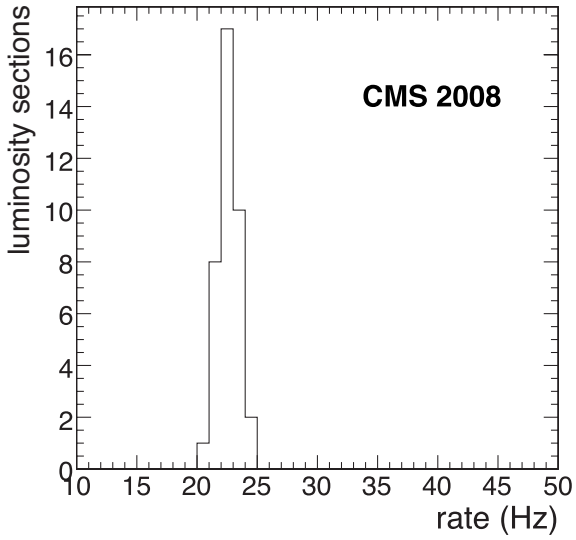


Figure 5.1: Rate distribution of the L1_SingleEG1 trigger for a typical run. The average rate is compatible with that expected from the noise level measured in the ECAL.

the nominal APD gain of 50, in order to study trigger performance, noise and the signatures of minimum ionising particles (MIP) in the same configuration used for collision data.
For the measurement of the full trigger chain efficiency runs with a consistent configuration of the full L1 e/γ trigger (namely the ECAL, the RCT and the GCT) were selected. For the trigger primitive generator efficiency measurement, only the ECAL trigger primitive generator was required to be properly configured. Finally, the study was restricted to the regions of the detector that had no known hardware problems.
To study the trigger resolution and efficiency, ECAL superclusters (SC) originated by muon radiation in the lead tungstate crystals and reconstructed offline as described in [85] were used as tags to probe for the production of L1 e/γ trigger candidates.
The e/γ trigger resolution and efficiency were measured using events that were taken with a muon trigger. Online data reduction for the ECAL was obtained through the selective readout algorithm described in Chapter 4. The low-interest energy threshold was set to 250 MeV and neighbours of low-interest towers were read out in full.
A dedicated algorithm was developed to reconstruct the energy deposited by cosmic ray muons in ECAL. The main differences with respect to the standard reconstruction algorithm used in collision events were driven by the requirement of good reconstruction and association efficiency down to low energy and for energy deposits associated with muons not pointing to the nominal interaction vertex. The energy reconstruction algorithm identifies clusters starting from a central crystal (seed) with an energy deposit of at least 15 ADC counts (139.5 MeV), or from the most energetic of a pair of adjacent crystals with at least 5 ADC counts (46.5 MeV) each, the cluster energy was determined as the sum of all channels above 2 ADC counts (18.5 MeV) belonging to a 5×5 matrix of crystals. Contiguous clusters are merged to form superclusters (SC) which size depends on the impact angle of the muon at the calorimeter surface. Pulse amplitudes were reconstructed using a weighted sum of the signal samples, as done for the reconstruction of collision events.
Only superclusters containing at least one crystal with a reconstructed energy above 400 MeV, corresponding to roughly ten times the noise RMS, were retained for analysis. This ensured accurate timing reconstruction [86] thus allowing the rejection of asynchronous cosmic ray deposits. In order to emulate collision events for which the trigger primitive generator was designed to be efficient, the phase between

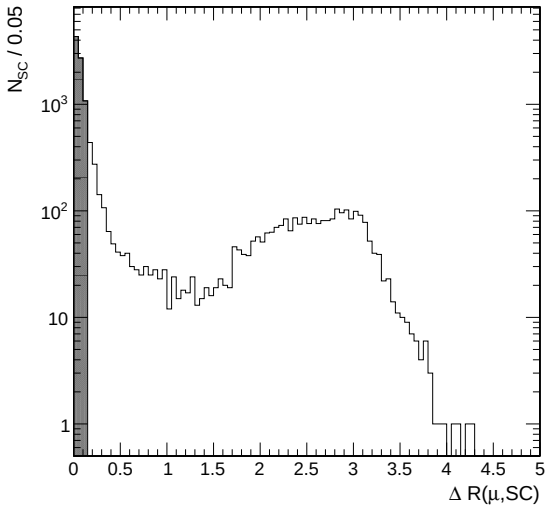


Figure 5.2: Distance in (η, ϕ) space between the reconstructed muon position extrapolated to the ECAL surface and the reconstructed supercluster. The shaded region indicates the selected candidates.

the reconstructed timing of the seed crystal and the clock was required to be smaller than 3.75ns.
Superclusters in ECAL were then associated with muon tracks according to a geometrical match between the track extrapolation to the calorimeter surface and the centre of gravity of the energy deposit. Figure 5.2 shows the distance in $\eta - \phi$ space between the extrapolated impact point of the muon track and ECAL supercluster. Only SC with $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.1$ were considered for analysis.

5.1.2 Resolution and Efficiency

After obtaining a pool of ECAL superclusters validated by reconstructed muons, the L1 e/γ trigger resolution and efficiency were probed. The distributions for $\Delta\eta = \eta(\text{L1}) - \eta(\text{SC})$ and $\Delta\phi = \phi(\text{L1}) - \phi(\text{SC})$ are shown in Figure 5.3 (left) for L1 candidates satisfying the L1_SingleEG1 requirement of $E_\text{T}(\text{L1}) > 1$ GeV. The binning used corresponds to the dimensions of one ECAL trigger tower, 0.087 in both η and ϕ. The highly populated region of 4×4 ECAL trigger towers corresponds to the resolution with which the L1 e/γ candidate position is reported.
Figure 5.3 (right) shows the relative difference between $E_\text{T}(\text{L1})$ and $E_\text{T}(\text{SC})$ for $E_\text{T}(\text{L1}) > 10$ GeV. The resolution from fitting a Gaussian to the distribution is 5.2% and the mean -4.7%. The non-zero mean is caused by differences in the way E_T is obtained in L1 and off-line reconstructions. These are: the clustering algorithm, the signal amplitude determination, and the effect of integer truncation in the L1 E_T determination, which had a LSB of 500 MeV. Each of these effects contributes to the L1 E_T being lower than the E_T which is recovered offline, the effect above 5 GeV being of the order of the LSB. The resolution is also of the order of the LSB; due to the steeply falling energy spectrum of these events, the sample is dominated by those just above the 10 GeV threshold.
To further characterise the e/γ trigger, two efficiencies were measured: the trigger primitive (TP) generator efficiency and the Level-1 e/γ candidate efficiency, which correspond to the first and the last steps in the calorimeter trigger chain. Due to the requirement of an energy deposit in the ECAL, this measurement evaluates the trigger efficiency only in the active part of the detector and is relative to the detector efficiency to detect muons and electromagnetic energy.
The trigger primitive generator was considered efficient if a muon-tagged ECAL supercluster had an associated TP in the same ECAL trigger tower. The TPG efficiency is shown in Figure 5.4 (left) as a

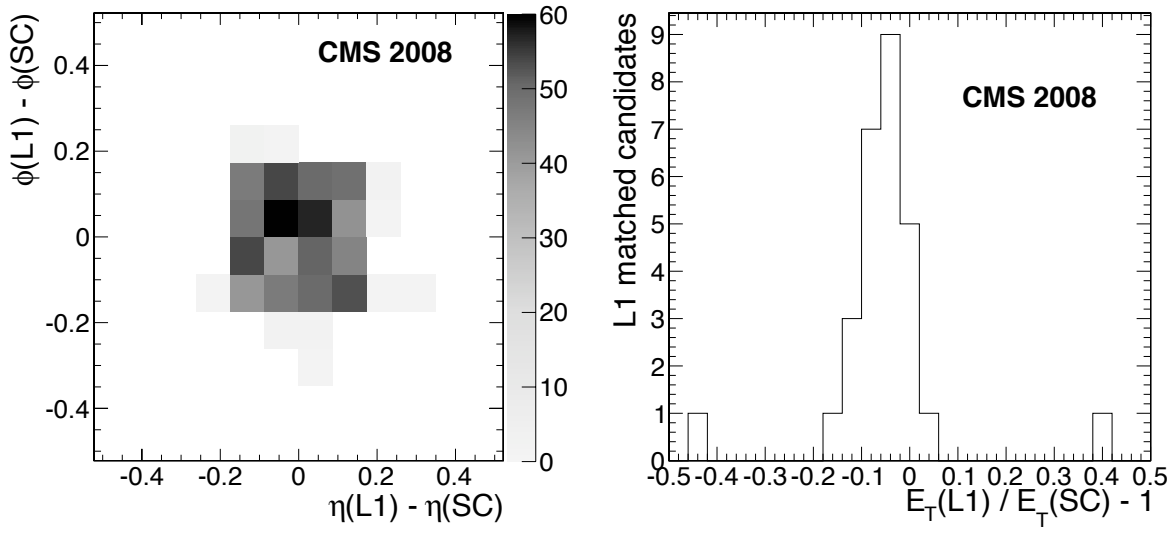


Figure 5.3: (left) $\Delta\eta$ and $\Delta\phi$ of Level-1 e/γ trigger objects with respect to reconstructed ECAL superclusters. (right) Level-1 e/γ E_{T} resolution for $E_{\mathrm{T}}(\mathrm{L1}) > 10$ GeV. The non-zero mean results from a combination of factors and is compatible with the LSB of $E_{\mathrm{T}}(\mathrm{L1})$.

function of $E_{\mathrm{T}}(\mathrm{SC})$. It rises with increasing $E_{\mathrm{T}}(\mathrm{SC})$ before reaching a plateau of 100%, the 50% efficiency turn-on point being at 0.70 ± 0.03 (stat.) ± 0.02 (syst.) GeV, compatible with the threshold set at 750 MeV. The systematic error is determined by varying the cuts applied to the selection.
L1 e/γ candidates are recorded along with the event data and are the result of the full calorimeter trigger chain: Trigger Primitive Generator, Regional Calorimeter Trigger and Global Calorimeter Trigger.
The full e/γ trigger chain was considered efficient if a muon-tagged ECAL supercluster had an associated L1 e/γ candidate with energy above the L1 e/γ threshold under study. The measure of association used is the distance between the L1 candidates and muon-tagged ECAL superclusters, $\Delta R\,(\mathrm{L1}, \mathrm{SC}^{\mu})$.
If $\Delta R\,(\mathrm{L1}, \mathrm{SC}^{\mu}) < 0.5$ and the L1 e/γ candidate rank was above the threshold under study, the event is considered efficient. With this selection the efficiency for the L1 e/γ trigger chain is shown in Figure 5.4 (right) for three different trigger algorithms: L1_SingleEG1, L1_SingleEG5 and L1_SingleEG10 with nominal thresholds at 1, 5 and 10 GeV, respectively. An unbinned maximum likelihood fit of an error-function was performed for each of the datasets.
The turn-on point of the 1 GeV threshold algorithm, L1_SingleEG1, was found to be 1.19 ± 0.02 (stat.) $\pm$ 0.02 (syst.) GeV. For the corresponding 5 and 10 GeV algorithms the turn-on points were measured to be 5.23 ± 0.09 (stat.) ± 0.14 (syst.) GeV and 10.2 ± 0.2 (stat.) ± 0.3 (syst.) GeV, respectively. Systematic errors were estimated by varying the timing selection window from ± 1.25 ns to ± 5 ns and the tagging distance $\Delta R\,(\mathrm{SC}, \mu)$ from 0.1 to 0.5. The discrepancies between the measured and expected turn-on points are a reflection of the effects already mentioned above, and mainly affect the L1_SingleEG1, since the L1 e/γ E_{T} LSB is 500 MeV. Above 10 GeV the effect is of no appreciable consequence.
During the 2010 LHC run the electromagnetic trigger efficiency was measured in pp collisions at $\sqrt{s} =$ 7 TeV [87]. Events were selected requiring a minimum bias trigger and energy deposits in the electromagnetic calorimeter reconstructed with the standard algorithms. The study was extended endcap regions; in the barrel region, the behaviour of the trigger efficiency algorithms as a function of the deposited energy was found to be in agreement with what was measured during the cosmic ray run (Figure 5.5, left). The absolute trigger efficiency, which includes malfunctioning trigger channels was also measured. For the L1_SingleEG5 algorithm a 100% efficiency plateau was observed to be reached for $E_{\mathrm{T}} \sim 20 GeV$ (Figure 5.5, right).

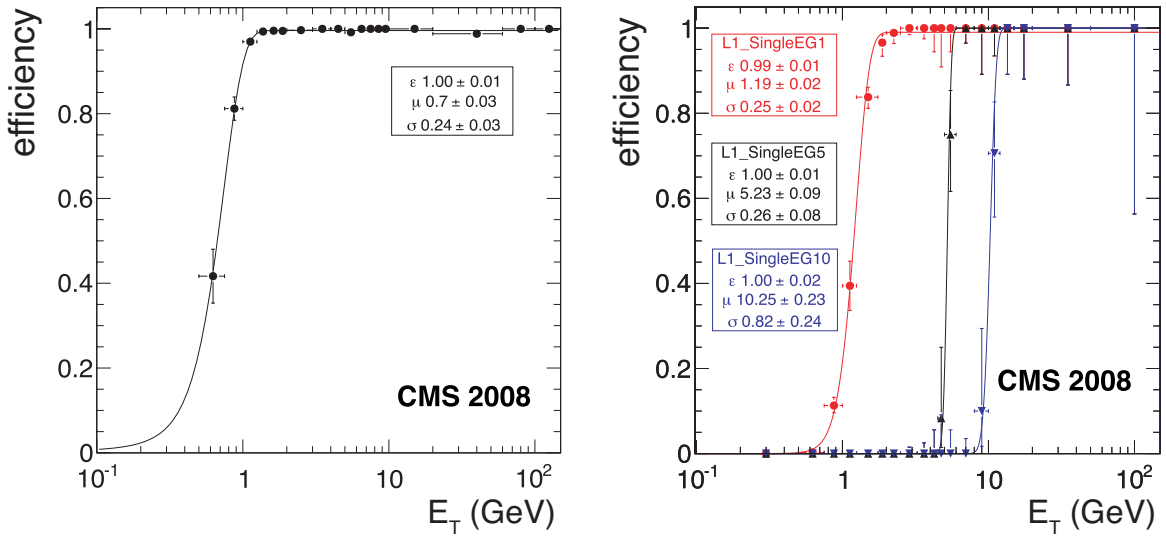


Figure 5.4: The ECAL trigger primitive production efficiency (left) and the full Level-1 e/γ trigger efficiency (right) as a function of the E_{T} reconstructed offline in ECAL. Parameters are obtained from fits of error functions to the data. In the case of the right figure, an unbinned fit was used.

A relatively sharper turn-on can be observed in the measurement performed with cosmic ray events with respect to what observed in collision events. This is a consequence of the different look-up tables used in the two cases and of the smaller spread of the energy deposits in comic ray events with respect to collision ones.

5.2 Electromagnetic objects reconstruction in $\sqrt{s} = 900\,\mathrm{GeV}$ pp collisions

LHC started operation at the end of 2009 and delivered $\sim 10\ \mu\mathrm{b}^{-1}$ of integrated luminosity to each of the LHC experiments.
Approximately 200 thousand minimum bias events recorded by the CMS detector at a centre of mass energy of 900 GeV were used to commission the reconstruction of the electron and photon physics objects. All the basic ingredients that contribute to the reconstruction and identification of electromagnetic physics objects were checked: electromagnetic calorimeter (ECAL) and preshower (ES) clusters, in the ECAL and in the hadron calorimeter (HCAL) as measured around electromagnetic deposits.
Given the low integrated luminosity used for the study and the absence of identification requirements, the comparison was mainly carried out on a sample dominated by energy deposits from hadronic decays. Nevertheless this was still sufficient to assess the general quality and the proper functioning of the algorithms and the modeling of the detector response in the simulation.

5.2.1 Datasets and event selection

Minimum bias events were triggered using scintillator planes that are located in front of the Hadron Calorimeter Forward detectors (Beam Scintillator Counters or BSC). The Level-1 trigger required at least one hit in one of the BSC.

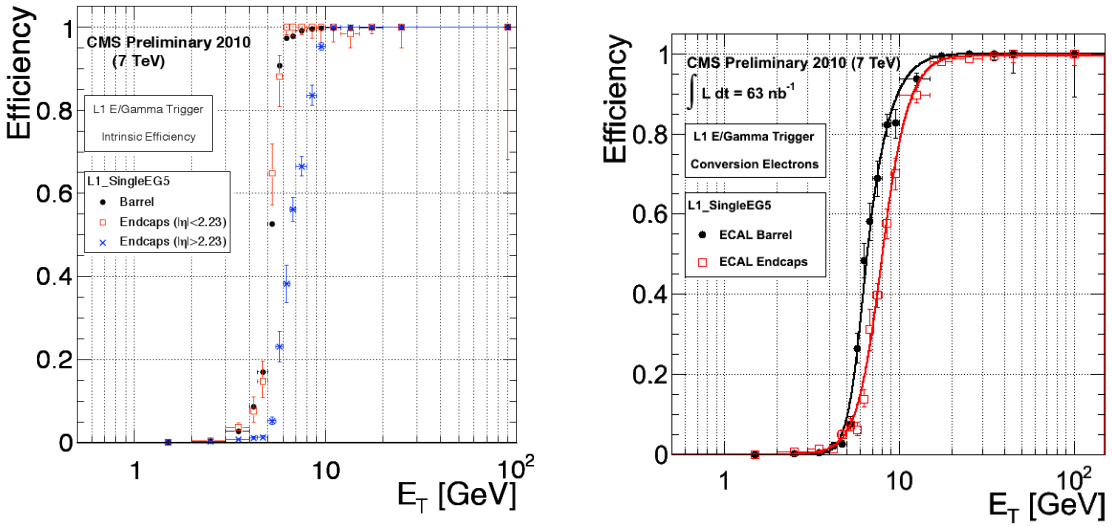


Figure 5.5: (left) Intrinsic efficiency of the L1 electromagnetic trigger algorithm measured in $\sqrt{s} = 7\,\text{TeV}$ pp collisions. (right) Absolute efficiency of the L1 electromagnetic trigger. Figures reproduced from [87].

Offline selection was then applied. The requirements were the following:

- The event had to be in time with a valid beam crossing measured by the coincidence of the two beam pickup monitors (BPTX);
- The presence of at least one primary vertex reconstructed with more than four tracks consistent with the beam spot was required;
- The fraction of *high purity* tracks [65] was requested to be at least 25% for events with at least 10 tracks;
- At least one hit with energy greater than 2 GeV in each of the Forward Hadron calorimeter detectors (HF) was required.

Runs where all relevant subdetectors were properly functioning were selected. The selected sample consists of 185330 minimum bias events.
These data were compared to a full MC simulation based on GEANT4 [88] of 10 million minimum bias events generated with PYTHIA 6.4 [17]. The simulation was carried out using mis-alignments, mis-calibrations and dead channel lists corresponding to the startup conditions of the CMS detector.

5.2.2 Supercluster reconstruction

Electron and photon showers deposit their energy in several ECAL crystals. The presence of material in front of the calorimeter results in bremsstrahlung and photon conversions. Because of the strong magnetic field the energy reaching the calorimeter is spread in ϕ. The energy is therefore clustered at the electromagnetic calorimeter level by building a cluster of clusters (supercluster or SC), which is extended in ϕ, thus minimising the cluster containment variations due to radiative processes. The clustering threshold is approximately 1 GeV in transverse energy. More details on the algorithms employed for electron and photon reconstruction can be found in Chapter 6.

The study was focused on the raw energy measurement of the superclusters, namely the sum of the energies deposited in the ECAL crystals. The response of the preshower detector was also analysed.
Isolated seemingly high energy deposits were observed during collision data taking at 900 GeV, 2.36 TeV and 7 TeV centre-of-mass energies. These energy deposits are believed to be caused by direct ionization of the avalanche photodiode (APD) sensitive volumes by highly ionising particles - mainly protons and heavy ions - produced during proton-proton collisions. Due to the fact that these deposits are observed in single crystals, they are often termed "spikes". The rate of these energy deposits is estimated to be approximately 1 in 10^3 minimum bias events. It is also observed to scale with $\sqrt{s}$ at a rate consistent with the measured increase in charged particle multiplicity [89].
Specific algorithms were developed to flag and reject the anomalous energy deposits. These are flagged based on topological and timing characteristics. A topological variable comparing the energy in a single crystal, $E1$, to the summed energy in the four adjacent crystals in η and ϕ, $E4$, is computed for each channel. A cut of $(1 - E4/E1) < 0.95$ has a rejection power that depends on the signal transverse energy[59]. More than 99% of the anomalous signals are removed for energy deposits above 10 GeV.
Analysis of the signal pulse shape provides further discrimination between energy deposits from electromagnetic showers and those from direct ionization of the APD. The former are a convolution of the time profile of light emission from the lead tungstate crystals (80% of light emitted in 25 ns) and the response of the front-end electronics (shaping time $\tau \sim 40$ ns), whereas, for the latter, only the electronics shaping time contributes. This difference in shape produces a bias in the measured time of the signal pulse when the anomalous energy deposits are reconstructed with the expected pulse shape for scintillation light. Signals produced by scintillation light are peaked at zero, whereas the apparent time of the anomalous signals is peaked at -10 ns due to the faster rise-time of their pulses.
For each reconstructed pulse, a flag is set based on the reconstructed time and the expected time resolution. A pulse is declared 'out-of-time' if the difference between the measured and expected time is greater than 5 standard deviations.
For the analyses described here, anomalous energy deposits were removed by rejecting signals with $(1 - E4/E1) > 0.95$.
The selective readout mid- and high-interest thresholds were set to 1 and 2 GeV respectively for the 2009 LHC run. During the 2009 data taking, some of the ECAL trigger towers, particularly in the endcaps, were not perfectly timed in and this caused the SR to sometimes not trigger the full read-out even when a region was above threshold.
No correction was applied to MC to account for this effect and thus some small discrepancies for variables that are sensitive to very little amount of noise or to negative fluctuations below the pedestal were observed.

5.2.3 Supercluster distributions

The comparison of the distribution of kinematic and shower shape variables between data and MC simulation was carried out. SCs were selected with the following requirements:

- the SC η was required to satisfy $|\eta| < 1.4442$ and $1.566 < |\eta| < 2.5$, avoiding the transition region between barrel and endcaps.
- the raw SC E_{T} was required to be larger than 2 GeV.

In total 3226 SCs were selected in data. Of these, 2120 were reconstructed in the barrel and 1106 in the endcaps. In the following plots MC distributions are normalised to the total number of SCs observed in data, in case of barrel/endcaps plots two different normalisations were used.
Figure 5.6 shows the distributions for data and Monte Carlo of the number of superclusters per event and of the supercluster raw transverse energy. Figure 5.7 shows the distributions for data and Monte Carlo as a function of the pseudorapidity of the SC. Figures 5.8 and 5.9 show two shower shape variables: the ratio between the maximum crystal energy and the total energy of the 5x5 crystal array centred around

the maximum energy crystal and the ratio between the sum of the crystals in the two neighbouring 1x5 eta strips, labelled as 2x5, containing the shower and the same 5x5 crystal array sum. This ratio may be larger than the unity due to fluctuation of the electronic noise to the lower side of the pedestal for crystals in regions that are fully read out.
Figure 5.10 shows the distribution of R_9: the ratio of the energy contained in the 3x3 region around the seed crystal and the total supercluster energy. R_9 is used to discriminate between converted and unconverted photons [66]. Figure 5.11 shows the distribution of H over E: the ratio of the energy in the HCAL behind the SC and the energy of the SC itself. For HCAL, only towers with energy greater than 0.7 and 0.8 GeV were considered in the sum respectively for barrel and endcaps.
All shape variables in the endcaps show some discrepancy between data and MC for values larger than one, due to the differences in the SR behaviour in data and MC. Apart from this, good agreement can be observed for all the variables considered.

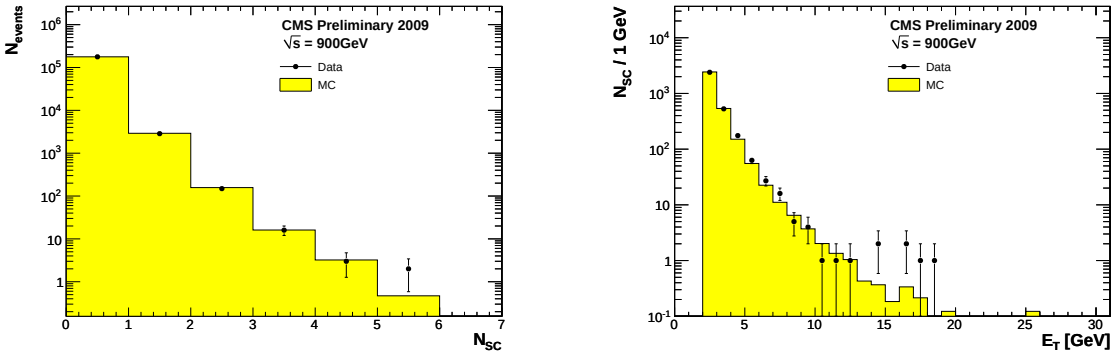


Figure 5.6: Number of selected superclusters per event (left) and raw Supercluster transverse energy distribution (right). The black points correspond to data and the histogram to simulated minimum bias events.

5.2.4 Preshower clusters

The preshower detector (ES) is located in front of the ECAL endcaps and covers the $1.65 < |\eta| < 2.6$ region. Figures 5.12 and 5.13 show the energy ratio of the 2^{nd} to the 1^{st} ES plane and the energy of preshower clusters in the two planes expressed in "equivalent minimum ionising particles". The preshower clusters were associated to ECAL superclusters with raw transverse energy larger than 2 GeV and $|\eta| < 2.5$.

5.2.5 Isolation studies

Isolation variables for superclusters were compared between data and simulation. These are analogous to the ones described in details in Chapter 6. Since the applied thresholds and cones definition slightly differ from the one discuss there, the precise definition is reported here:

- *Track isolation*: the sum of the transverse momenta of tracks that were reconstructed in a hollow cone around the reconstructed object. The dimensions of the cone are $0.04 < \Delta R < 0.3$. Only tracks with transverse momentum greater than 0.7 GeV were considered in the sum.
- *ECAL isolation*: the sum of the transverse energy reconstructed in individual channels of the ECAL in a cone around the reconstructed object with outer cone size $\Delta R = 0.3$ and inner cone radius corresponding to the size of 3 ECAL crystals ($\Delta R \sim 0.05$ in barrel region). The transverse energy

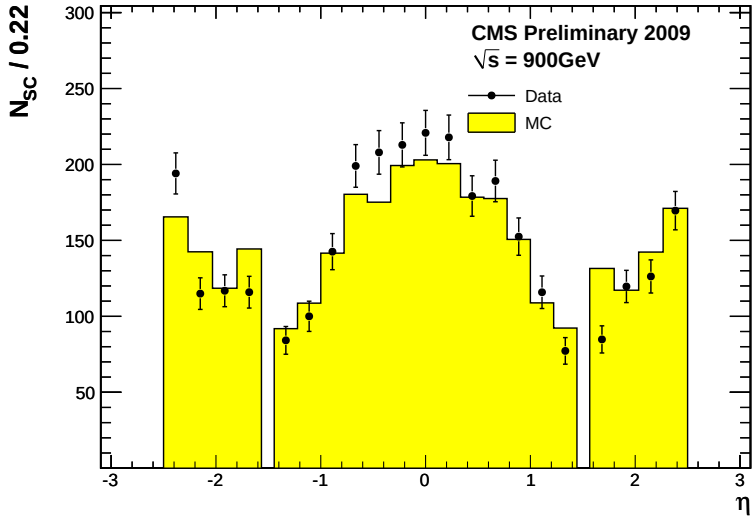


Figure 5.7: Pseudorapidity distribution of the superclusters. The black points correspond to data and the histogram to simulated minimum bias events.

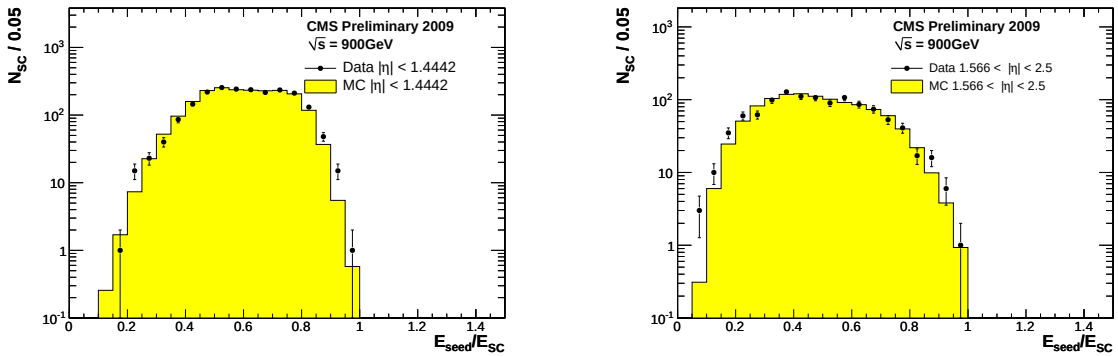


Figure 5.8: Ratio between the most energetic crystal energy and total supercluster energy for the barrel (left) and endcaps (right) SCs. The black points correspond to data and the histogram to simulated minimum bias events.

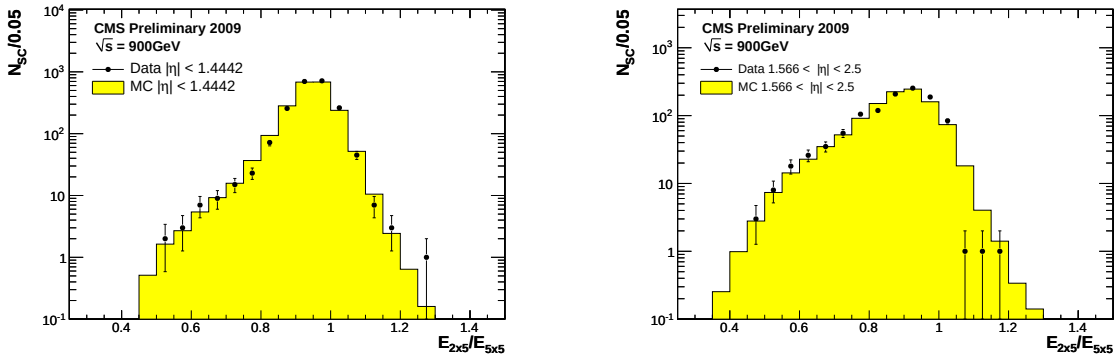


Figure 5.9: Ratio between the energy contained in the 2x5 crystal eta-strip and the energy in the 5x5 crystal array for the barrel (left) and endcaps (right) SCs. The black points correspond to data and the histogram to simulated minimum bias events.

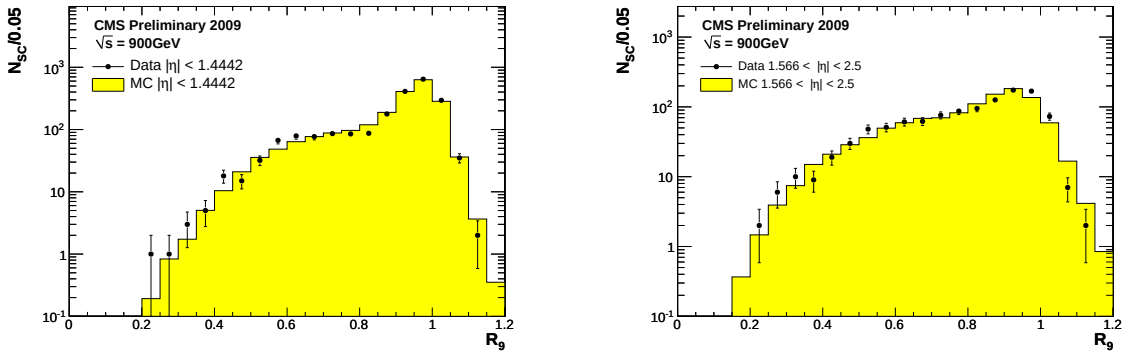


Figure 5.10: R_9: ratio between the energy contained in the 3x3 region around the seed crystal and the total supercluster energy for the barrel (left) and endcaps (right) SCs. The black points correspond to data and the histogram to simulated minimum bias events.

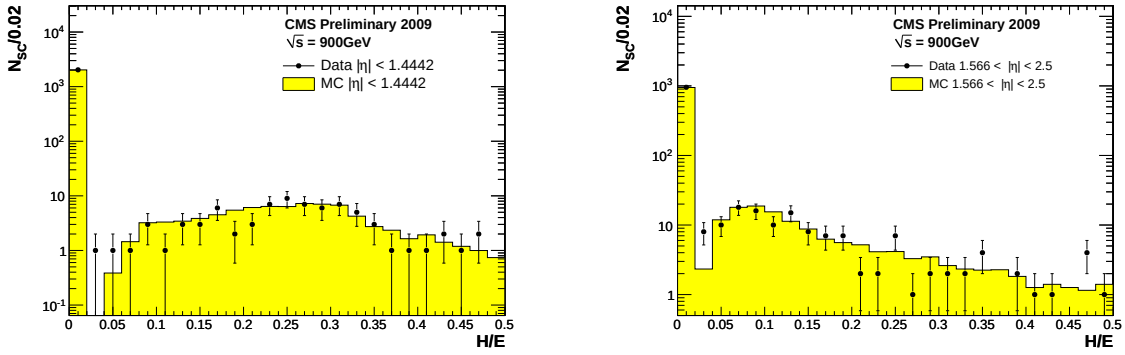


Figure 5.11: H over E distribution for SCs in the barrel (left) and endcaps (right). The black points correspond to data and the histogram to simulated minimum bias events.

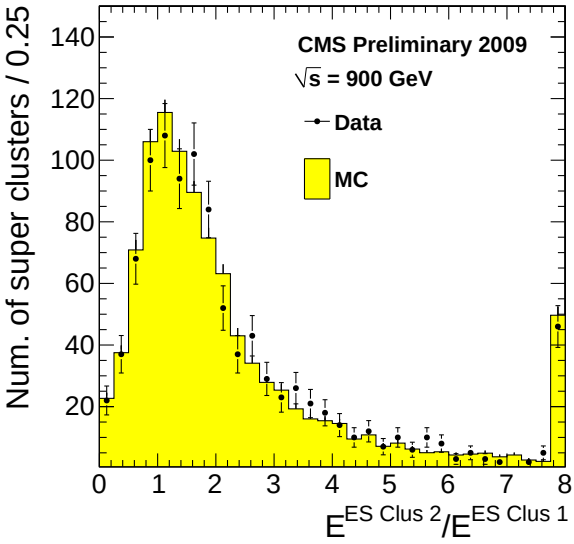


Figure 5.12: The ratio of the energy deposit associated with a supercluster on the second ES plane to the first one for supercluster raw transverse energy larger than 2 GeV. Overflows are added to the last bin.

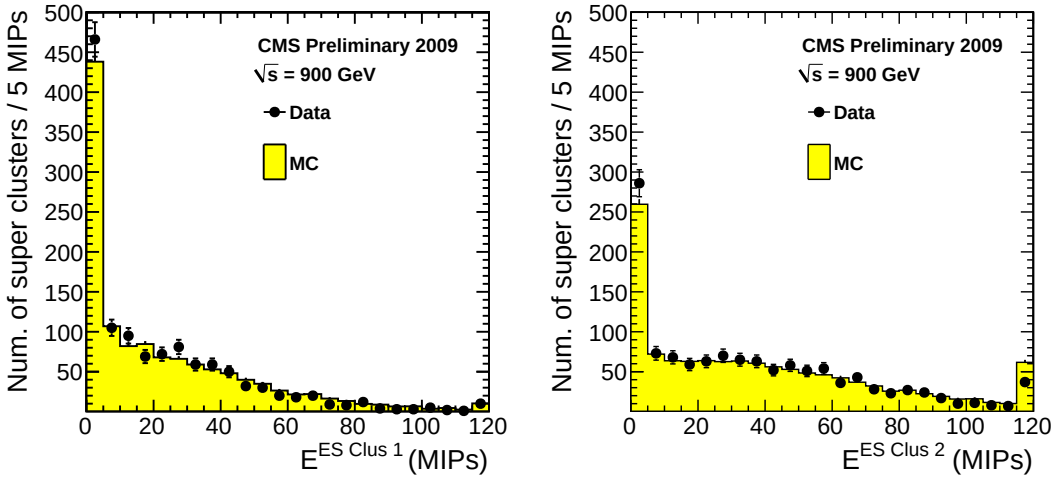


Figure 5.13: Energy deposited in each of the two ES planes for supercluster raw transverse energy larger than 2 GeV. Overflows are added to the last bin.

in channels found in a strip along ϕ centred at the ECAL position of the reconstructed object with an η-width of 3 crystals were also not considered in the sum. Only those reconstructed hits with the absolute value of the energy greater than 0.08 GeV in the ECAL barrel (EB) and with the absolute value of the transverse energy of 0.1 GeV in the ECAL endcaps (EE) were considered. The cut on the absolute value of the energy was aimed at averaging out the effect of noise and it may give rise to negative values of the ECAL isolation variable.

- *HCAL isolation*: the sum of the transverse energy of HCAL towers in the region behind the ECAL cluster of the reconstructed object. The sum was performed in a cone with dimensions $0.15 < \Delta R < 0.4$. The energy is summed of towers which have an energy greater than 0.7 GeV in the barrel and 0.8 GeV in the endcap.

In all plots the Monte Carlo sample were normalised to the number of superclusters in data, in case of barrel/endcaps plots two different normalisations were used. Superclusters were selected with the same requirements described above.
Figure 5.14 shows the track isolation, Figure 5.15 shows the ECAL isolation and Figure 5.16 shows HCAL isolation computed for the selected superclusters.
All isolation variables can be observed to be well reproduced by the MC simulation.

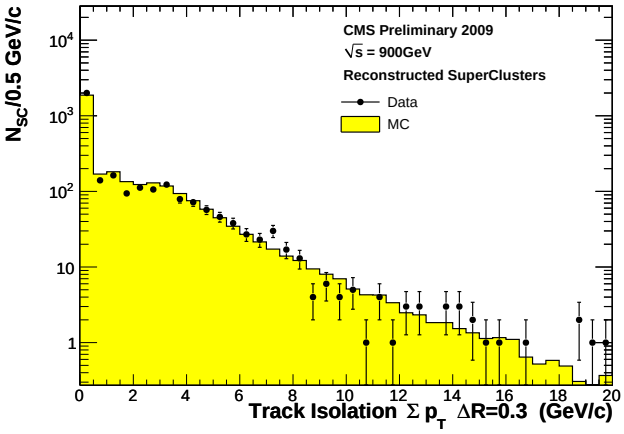


Figure 5.14: Track isolation distribution for superclusters. The black points correspond to data and the histogram to simulated minimum bias events.

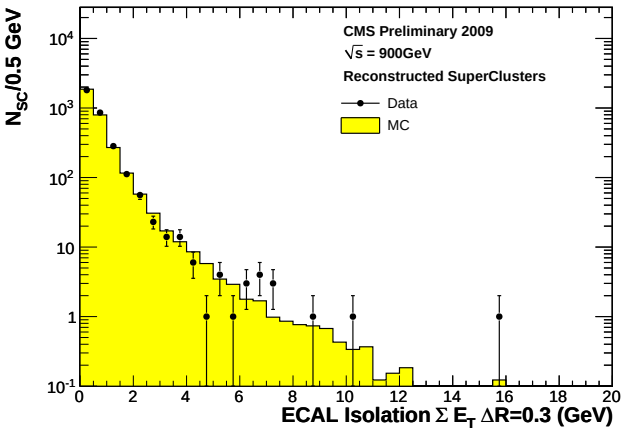


Figure 5.15: ECAL isolation distribution for superclusters. The black points correspond to data and the histogram to simulated minimum bias events.

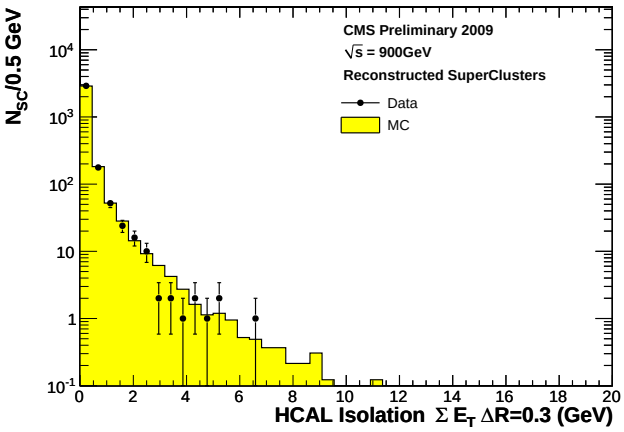


Figure 5.16: HCAL isolation distribution for superclusters. The black points correspond to data and the histogram to simulated minimum bias events.

5.3 Conclusions

This chapter was devoted to the discussion of the performance of the CMS ECAL in-situ. Given the vastness of the subject, the author tried to provide a glimpse of effort that the CMS collaboration devoted to the tuning and understanding of the detector performances during the commissioning phase.
Two of the studies that contributed to the commissioning process were selected as representative of a larger class. The author of this manuscript gave his direct contribution to both studies.
The subject of first study was the characterisation of the electromagnetic trigger. Using data acquired during the CMS CRAFT exercises, the electromagnetic trigger was probed using radiative losses from cosmic ray particles. The energy and position resolution of the L1 e/γ trigger algorithms were found to be consistent with expectations. The algorithm was shown to be fully efficient and the E_{T} turn-on points for different trigger thresholds were measured.
The subject of the second study was the performances of the electromagnetic objects reconstruction algorithms. Using the first proton-proton collision data collected with the CMS detector in 2009 at a center of mass energy of 900 GeV, comparisons between data and simulation were carried out for ECAL clusters, tracker, ECAL and HCAL isolation variables for electromagnetic candidates. Due to the limited statistics of the available sample, this first evaluation of the reconstruction algorithms was carried out on a sample dominated by energy deposits from hadronic decays.
All kinematic and identification variables which were considered show a good agreement between data and Monte Carlo, leading to the conclusion that the response of the subdetectors is well meddled in the simulation and that the algorithms designed and optimised in the simulation show a behaviour consistent with what is expected.
In the next chapter, the application of the trigger and reconstruction algorithms to the measurement of Physics processes will be discussed.

Chapter 6

Photon reconstruction and identification in CMS

Contents

This chapter is devoted to the discussion of the photon reconstruction and identification techniques employed in the CMS experiment. In particular a study aimed at the definition of photon identification criteria for Physics analyses performed on early LHC data at $\sqrt{s} = 7$ TeV is presented. The criteria for photon identification were tuned to obtain a robust selection. The performance of the selection was verified using collision data of the first LHC run.
In the following, the algorithms for photon reconstruction and identification in the CMS detector are described. Then, the tuned selection criteria and their performance with first LHC data are discussed. The selection criteria developed in the study discussed here have then been employed for the measurement of the isolated photon cross section, which is discussed in the following chapter.
Here, the term photon reconstruction refers to the techniques used to determine the photon energy and position. The goal of this step is to obtain a reconstruction efficiency for prompt photons close to 100% and to have the best energy and position resolutions. A second step, referred to as photon identification, comprises all those techniques used to discriminate prompt photons from their background, mainly due light neutral mesons in jets decaying into photon pairs.

6.1 Photon reconstruction

Photons are primarily reconstructed through the energy deposited in the ECAL. A photon incident perpendicularly on the detector release $> 95\%$ of its energy in a matrix of 25 crystals. However, the presence of material in front of the detector causes the photons to convert into electron-positron pairs. The solenoidal field makes the electron and positron trajectories to separate in the ϕ direction and this leads the energy deposits to be spread over larger regions. The effect is well illustrated by figure 6.1, which shows the photon conversion probability and the fraction of the true photon energy reconstructed in a 5×5 crystal matrix for converted and unconverted photons as a function of the photon pseudorapidity.

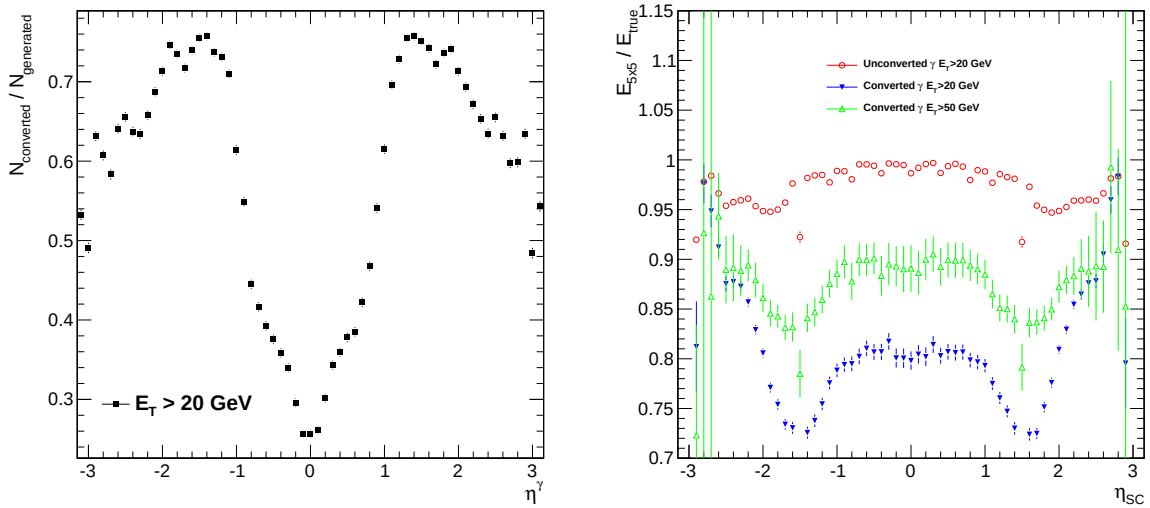


Figure 6.1: (left) Fraction of photons converting into electron-positron pairs before reaching the ECAL as a function of the photon pseudorapidity. The plot is obtained from a simulated sample of direct photons with $E_T > 20\ GeV$. (right) Fraction of photon energy contained in a 5×5 crystal matrix for the same sample of photons as function of the reconstructed super-cluster pseudorapidity.

In order to efficiently reconstruct the full photon energy, specific algorithms, called clustering algorithms, have been designed. They group energy deposits in individual crystals to obtain objects called *super-*

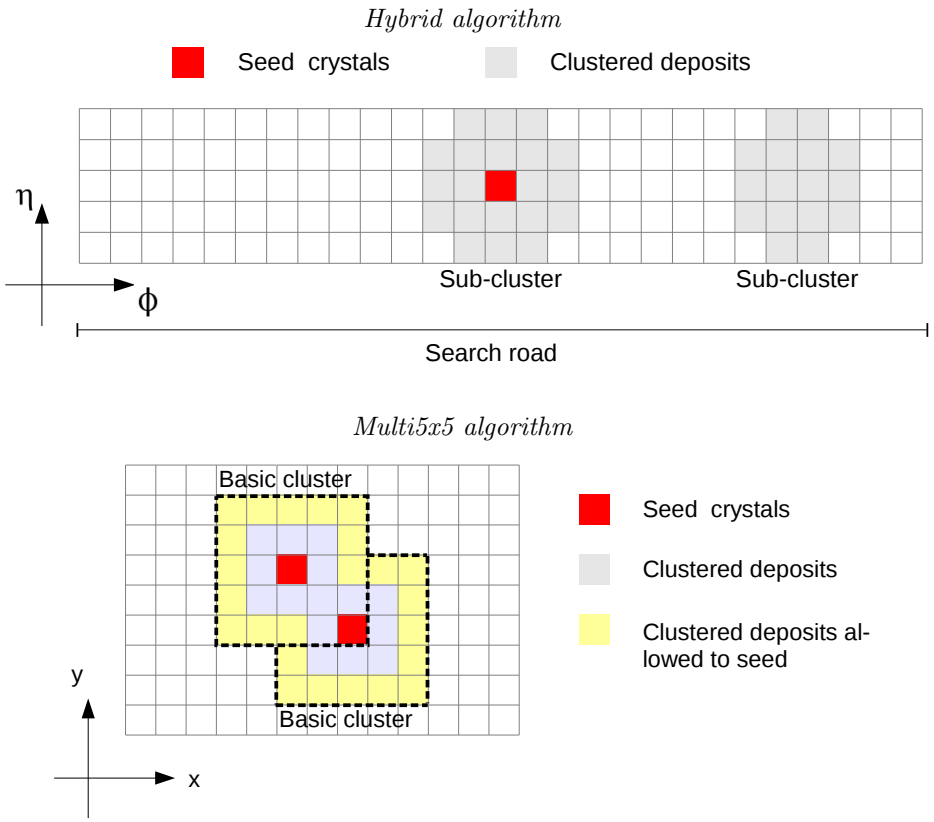


Figure 6.2: Illustration of the clustering algorithms used in the barrel (top) and endcap (bottom) regions.

clusters (SC) aimed at recovering the entire energy of the original photon. The case of prompt electrons is very similar to that of prompt photons; bremsstrahlung emission has effects analogous to pair-production and in fact the same clustering algorithms are employed for electron and photon reconstruction.

6.1.1 Clustering algorithms

The solenoidal geometry of the CMS magnetic field causes the energy from bremsstrahlung and pair-production to be spread along the ϕ direction only, since the trajectory in the $r - z$ plane is unaffected by the magnetic field. The goal of the clustering algorithm is then to group neighbouring crystals to reconstruct individual showers and to associate showers reconstructed at the same η within a given window in the ϕ direction.
To perform this task, two different clustering algorithms have been designed for the barrel and endcap region, due to the different mechanical layout of the two sections. The algorithm used in the barrel region is called *Hybrid algorithm* and exploits the $\eta - \phi$ geometry of this part of the detector. For the endcap region, a different algorithm, called *Multi5x5 algorithm* is used, which employs 5×5 crystal matrices to gather energy deposits.

The Hybrid algorithm can be summarised as follows [90]:

1. At each step, all crystals not already belonging to a cluster are tested in decreasing energy order. To avoid noise contamination and low energy backgrounds, the crystal transverse energy E_T is

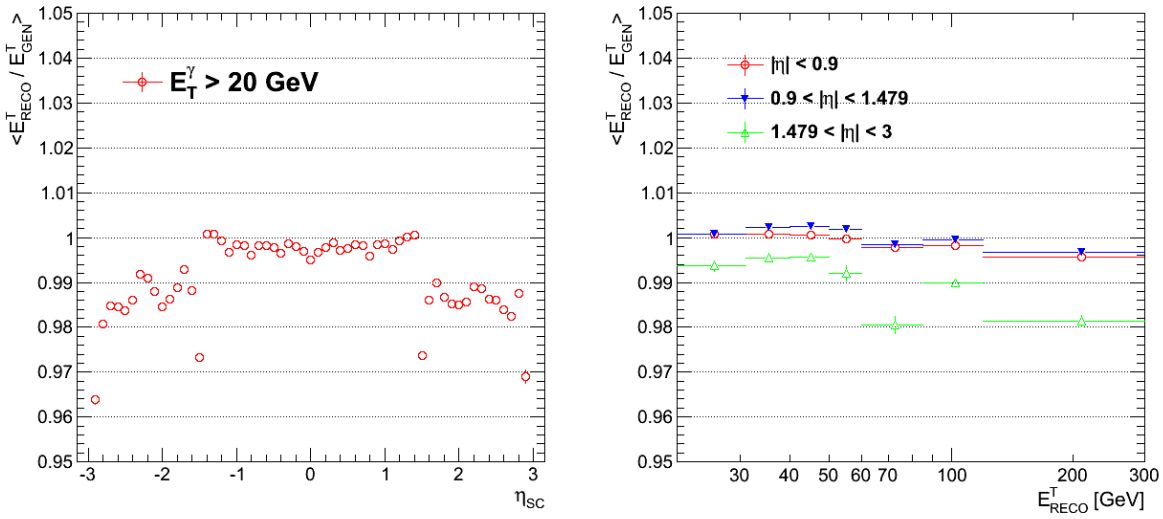


Figure 6.3: Ratio of reconstructed to true photon energy as a function of the reconstructed super-cluster pseudorapidity (left) and corrected transverse energy (right). The plot refers to a simulated sample of direct photons with energy corrections obtained using isolated electrons. Discontinuities in the ratio as a function of pseudorapidity are due to inactive regions of the detector. Residual effects as a function of transverse energy can be ascribed to the differences between electrons and photons.

required to be above a minimum threshold $E_T^{seedthr}$. If $E_T > E_T^{seedthr}$ the crystal can seed the clustering process. Otherwise, the next crystal is examined.

2. A 5×1 domino of crystals in $\eta - \phi$ direction around the seed crystal is built.

3. The second step is repeated for all crystals with the same η as the seed one that satisfy $|\phi_{crystal} - \phi_{seed}| < \Delta\phi_{road}$ (search road). The domino is included in the cluster if $E_{domino} > E_{domthr}$.

4. The dominoes with $E_{domino} > E_{domthr}$ that were not aggregated to the main clusters are then searched for local energy maxima and secondary clusters are formed around maxima where the highest crystal energy is above a second threshold E_{locthr}.

5. The algorithm continues until all crystals have been examined.

The result of the procedure are super-clusters made up by several showers at constant η but spread in the ϕ-direction, like in the example reported figure 6.2.

The Multi5x5 algorithm proceeds as follows:

1. At each step unclustered crystals are examined in decreasing transverse energy order. If the transverse energy $E_T > E_T^{seedthr}$, the crystal can seed the clustering process. Otherwise, the next crystal is examined.

2. The crystal is tested for being a local maximum by comparing its energy to its four neighbours by side in a *Swiss Cross* pattern. If the crystal is not a local maximum, the algorithms goes back to the first step.

3. A *basic cluster* is created including crystals in the 5×5 window around the seed that not already assigned to other basic clusters.

4. The algorithms continues until all crystals have been examined.

To recover the energy of secondary showers, a rectangular window along η and ϕ is opened around basic clusters with transverse energy above a threshold E_T^{bcthr}. Other basic clusters falling within the window are added to form the super-cluster. All basic clusters are examined in descending transverse energy order, with the constraint that each basic cluster can be assigned to only one super-cluster. In the region covered by the preshower detector, the energy detected in the latter is added to the super-cluster energy. The association is performed extrapolating the super-cluster position towards the interaction point.
Figure 6.2 shows, schematically, examples of super-clusters resulting from the two clustering algorithms. The parameters of the algorithms that were used in the reconstruction software for this thesis are reported in Table 6.1.

Regardless of the algorithm, the SC position is estimated through a weighted average of the position of all the crystals, where each crystal enters with a weight, $w_i = \max(0, 4.7 + \log(E_i/E_{SC}))$. Super-clusters are subsequently promoted to photon candidates and the direction of the photon momentum is estimated connecting the SC position to that of the primary vertex in the event. If several primary vertexes are found, the one with the highest scalar sum of track transverse momenta is chosen.

Table 6.1: Parameters of the clustering algorithms as used in this thesis.

Hybrid algorithm *(barrel)*		Multi5x5 algorithm *(endcap)*	
$E_T^{seedthr}$	1 GeV	$E^{seedthr}$	180 MeV
E_{domthr}	350 MeV	E^{bcthr}	1 GeV
E_{locthr}	100 MeV		
$\Delta\phi_{road}$	17 crystals	$\Delta\phi_{road}$	0.14
		$\Delta\eta_{road}$	0.6

6.1.2 Energy corrections

A high resolution on the photon energy is essential for all the measurements and searches done with final states containing photons. Several detector effects concur to the degradation of the energy resolution and need to be properly corrected in order to obtain the best possible performances. For this purpose, after the clustering step, correction factors are applied to the total super-cluster energy in order to achieve an homogeneous response in the full calorimeter volume. The reconstructed photon energy is given by:

$$E_\gamma = F_\gamma \cdot G \cdot \sum_{i=1..N} c_i \cdot A_i \tag{6.1}$$

here A_i are the digital amplitudes measured by the ECAL readout electronics and c_i are a set of coefficients that equalise the response of the different crystals; G is a global energy scale, defined such that sum of the amplitudes of a 5×5 crystal matrix multiplied by G amounts to the total energy of an incident unconverted photon. The factor F_γ represents a correction to the super-cluster energy. Three types of effects are taken into account by this term:

Variations of the shower containment as a function of the position in the detector are parametrised by a function labelled $C_{EB}(\eta)$. Such an effect is important only in the barrel region, where the non-uniformities in the lateral shower leakage due to the off-pointing geometry of the crystals need to be corrected. This correction is obtained from MC simulations and has been found to be in good agreement with test-beam data. Overall, the C_{EB} correction is $\lesssim 1\%$ [91].

Variations in the algorithm response to different super-cluster topologies are corrected through a function called $f(brem)$. The $brem$ variable is defined as

$$brem = \frac{\sqrt{\sum \left(\eta_i - \bar{\eta}\right)^2 c_i A_i}}{\sqrt{\sum \left(\phi_i - \bar{\phi}\right)^2 c_i A_i}} \tag{6.2}$$

where the sum runs over all crystals in the super-cluster and $(\bar{\eta}, \bar{\phi})$ refers to the super-cluster position. This function is insensitive, within certain limits, to the amount of material in front of the calorimeter and can be obtained from MC simulations. The size of this term is $< 7\%$ in the barrel region and $\lesssim 20\%$ in the endcap region.

Residual variations due to the non-uniform distribution of material in the detector and the energy-dependence of the energy-collection efficiency are corrected through a function $f(E_T, \eta)$. Since this function depends on the details of the material distribution in the detector its determination has to be performed *in-situ*. In the case of electrons, $f(E_T, \eta)$ can be measured using $Z \to ee$ events with a luminosity of $\mathcal{O}(10\ \mathrm{pb}^{-1})$. Differences between electrons and photons are expected to be small and can be modelled using MC simulation, until a sufficient sample of prompt high energy photons is accumulated using events with associated $Z\gamma$ production.

Many of the effects parametrised by F_γ are not relevant in the case of unconverted photons and the corresponding corrections should not be applied. For this purpose, unconverted photons are selected by the ratio of the energy deposited in the 3×3 crystal matrix surrounding the seed crystal to total super-cluster energy. If this ratio, denoted R_9, exceeds some preset thresholds, the photon energy is reconstructed from the energy of the 5×5 matrix containing the seed crystal and only the C_{EB} correction is applied. The R_9 thresholds used to produce the results of this thesis are 94% and 95% for the barrel and endcap regions respectively.

The ratio between the true and reconstructed photon energy for a simulated sample of direct photons is shown in figure 6.3. To produce this figure, energy corrections have been obtained using simulated electrons from $Z \to ee$ decays. A residual variation of $\mathcal{O}(0.2-1\%)$, due to the difference between electrons and photons can be observed.

6.2 Photon Identification

The most important background to prompt photons is due to jets fragmenting mainly into light neutral mesons, such as π_0 and η, which subsequently decay to a pair of photons. The angular separation between the photons in the laboratory frame is $\sim 2/\Gamma$ where Γ is the ratio between the meson energy and its mass and it becomes of the order of the ECAL granularity ($\Delta\eta \times \Delta\phi = 0.0174^2 \div 0.05^2$) for π_0's of transverse energy above 20 GeV. Thus the separation of the prompt photon component from this background becomes very challenging at high energy.

Even though an event-by-event discrimination is very difficult, a separation of signal and background is possible on a statistical basis. A powerful instrument to separate signal and background consists in the analysis of the topology of the energy deposits. Several variables sensitive to differences between signal and background can be constructed and are classified under the name of *shower shapes*.

The neutral mesons contributing to the background are produced in jets, thus in association with other particles, while this is not true in the case of prompt photons. The jet background can therefore be reduced requiring the reconstructed super-clusters to be *isolated*, i.e. limiting the amount of energy carried by other particles surrounding them.

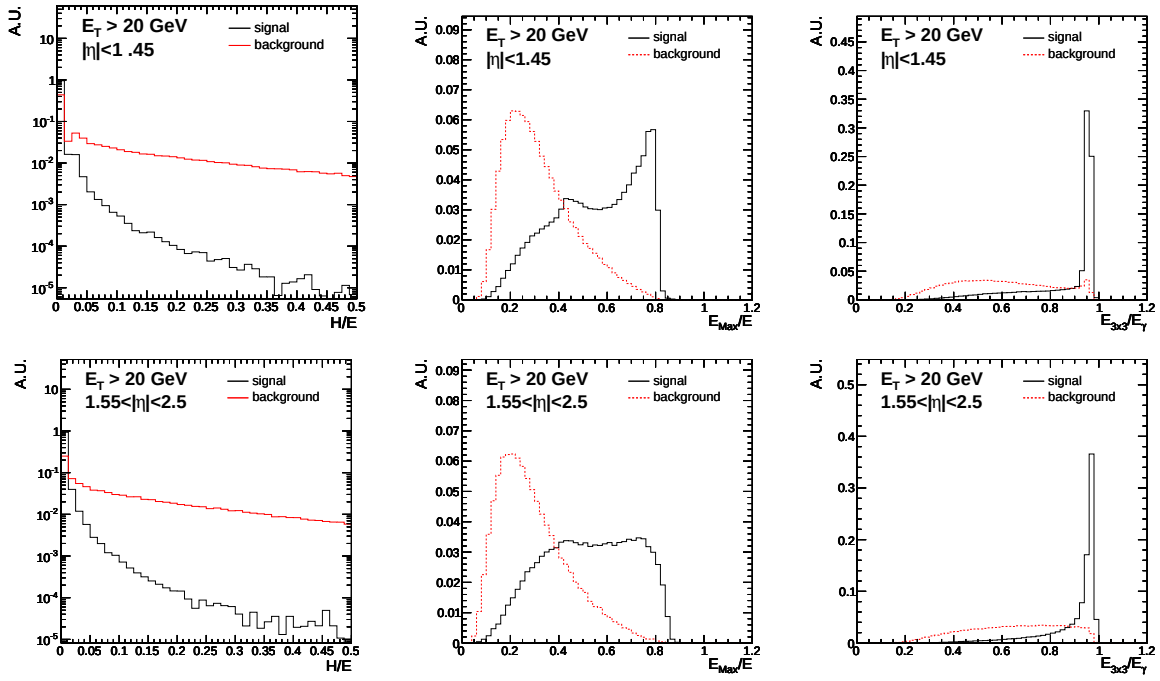


Figure 6.4: Comparison of shower shape variables for prompt photons and jet background: H/E, R_1 and R_9 variables. (top row) Barrel region. (bottom row) Endcap regions.

6.2.1 Shower shapes

The shower topology is an important discriminator to disentangle prompt photons from jet background. Several variables can be constructed to parametrise the differences between signals due to prompt photons from those caused by the interaction of hadrons in jets. The distribution of some of these variables for signal and background is reported in figures 6.4 and 6.5:

Hadronic over electromagnetic ratio (H/E) : the compatibility of the shower with a purely electromagnetic interaction with the detector can be tested looking at the ratio between the energy deposited in the HCAL and ECAL. Due to the large thickness of the ECAL, the probability of *punch through* for an electromagnetic shower is, in fact, very small. The H/E ratio is defined as the ratio between the energy deposited in the HCAL in a cone of radius $\mathcal{R} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.15$ around the SC direction and the energy of the photon candidate.

Rectangular ratios : the hypothesis of a single photon shower versus multiple overlapping showers can be tested looking for a double peak structure in the topology of the SC. The most simple type of variables sensitive to this aspect that can be constructed are ratios of the energy deposited in $N \times M$ crystals window around the SC seed crystal to the energy deposited in an $N' \times M'$ window or the total SC energy. Examples of this kind of ratios are the R_1 ratio, defined as the ratio of the seed crystal energy to the SC energy and the R_9 ratio, defined before. Also, the ratio of the energy deposited in a 1×5 and 2×5 window to the 5×5 one can be used to discriminate single photons from multiple photons showers.

Shower moments : a more sophisticated way of looking for double peak structures is to look for elongated deposits. It is in fact possible to define a covariance matrix describing the spatial extension of the

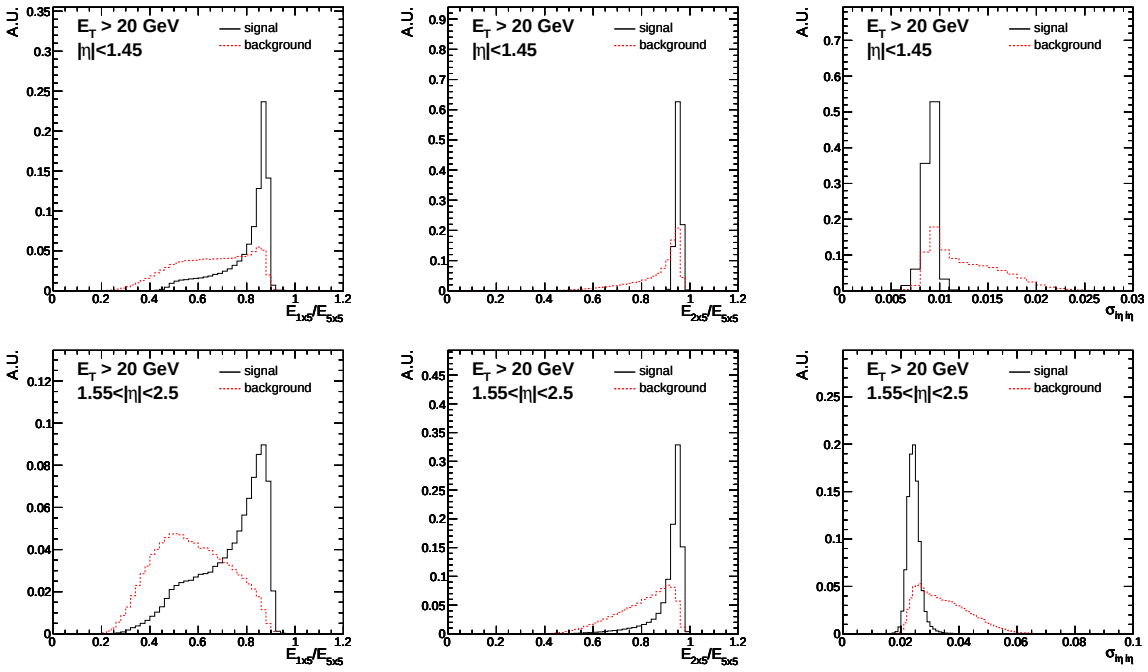


Figure 6.5: Comparison of shower shape variables for prompt photons and jet background: $E_{1\times5}/E_{5\times5}$, $E_{2\times5}/E_{5\times5}$ and $\sigma_{i\eta i\eta}$ variables. (top row) Barrel region. (bottom row) Endcap regions.

shower in $\eta-\phi$ space, $\begin{pmatrix} \sigma^2_{\eta\eta} & \sigma^2_{\eta\phi} \\ \sigma^2_{\eta\phi} & \sigma^2_{\phi\phi} \end{pmatrix}$,

$$\sigma^2_{\alpha\beta} = \frac{\sum_{i=1}^{N} (\alpha_i - \bar{\alpha})(\beta_i - \bar{\beta}) w_i}{\sum_{i=1}^{N} w_i} \tag{6.3}$$

where $(\bar{\eta}, \bar{\phi})$ refer to the SC position, (η_i, ϕ_i) to the position of the crystals entering the sum and the weights w_i are a function of the energy deposited in each crystal E_i (defined as $E_i = G \times c_i \times A_i$). Slightly different definitions of the shower covariances can be obtained depending on the choice of the weights and the set of crystals used in the computation. For instance, in expression (6.2) $w_i = E_i$ and the sum runs over the full SC. For identification purposes, the 5×5 matrix around the seed crystal is used and $w_i = \max(0, 4.7 + \log(E_i/E_{5\times5}))$. The resulting set of variables is labelled as $\sigma_{i\eta i\eta}, \sigma_{i\eta i\phi}, \sigma_{i\phi i\phi}$.

6.2.2 Isolation

The requirement that the photon candidate is not produced in association with other particles, can be specified quantitatively through the definition of isolation sums. These express the total amount of energy/momentum carried by particles surrounding the candidate. Three type of sums have been employed in this thesis: the total energy detected in the electromagnetic and hadronic calorimeters and the total tracks momentum measured in the tracker. All sums are defined in a cone of radius $\mathcal{R} = 0.4$ around the photon candidate. Specific veto regions have been defined for each isolation sum with the purpose of removing the energy deposits due to the photon candidate.

The Iso_{TRK} is defined as the scalar sum of the p_T of the tracks consistent with the primary vertex of the interaction. An exclusion region consisting of an inner cone of radius $\mathcal{R} = 0.04$ and a rectangular

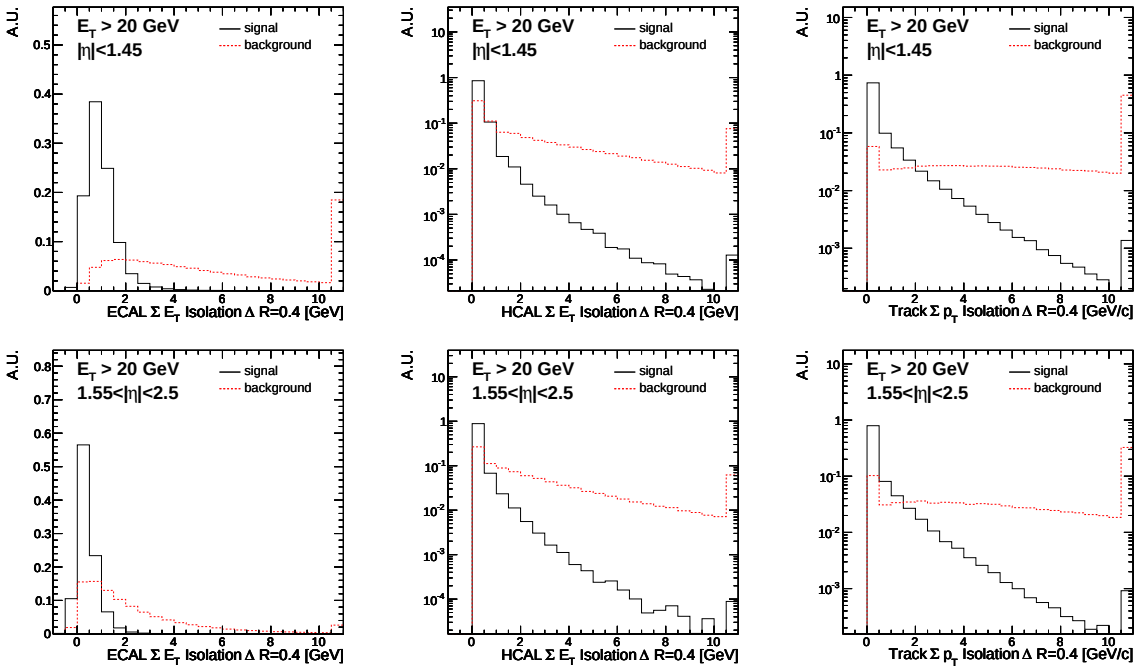


Figure 6.6: Comparison of isolation sums for prompt photons and jet background: $\mathrm{Iso_{ECAL}}$, $\mathrm{Iso_{HCAL}}$ and $\mathrm{Iso_{TRK}}$. (top row) Barrel region. (bottom row) Endcap regions.

strip of $\Delta\eta \times \Delta\phi = 0.015 \times 0.4$ is defined in order to remove from the sum tracks from pair conversions.

The $\mathrm{Iso_{ECAL}}$ is defined as the sum of transverse energy deposited in the electromagnetic calorimeter crystals. In order to remove the contribution due to the photon candidate, an exclusion region consisting of an inner cone of radius $\mathcal{R} = 0.06$ and a strip of dimensions $\Delta\eta \times \Delta\phi = 0.04 \times 0.4$ is defined. In the barrel (endcap) region, energy (transverse energy) deposited in a single crystal is required to be greater then 80(100) MeV in absolute value.

The $\mathrm{Iso_{HCAL}}$ is defined as the sum of the energy deposited in the hadronic calorimeter towers, excluding an inner cone of radius $\mathcal{R} = 0.015$.

The distribution of these variables for signal and background is shown in figure 6.6.

6.2.3 Photon identification criteria for start-up analyses

The variables described in the previous sections can be combined in order to discriminate prompt photons from their backgrounds. The most important aspect for the development of a set of start-up identification criteria is the minimisation of the impact of the imperfect understanding of the detector performance.
In the absence of a data-driven control sample to tune the identification criteria, the set of variables entering the selection was restricted to a small number of more easily understandable ones and the criteria optimised using MC simulations. The biggest impact on the description of the photon identification variables is expected to come from the simulation of the detector noise, the description of the material in front of the calorimeter and the simulation of the underlying event.
The tuning of the selection criteria was driven by the requirements of minimising the inputs from MC simulation in the determination of the selection efficiency and maximising the background rejection. The set of variables chosen is the following:

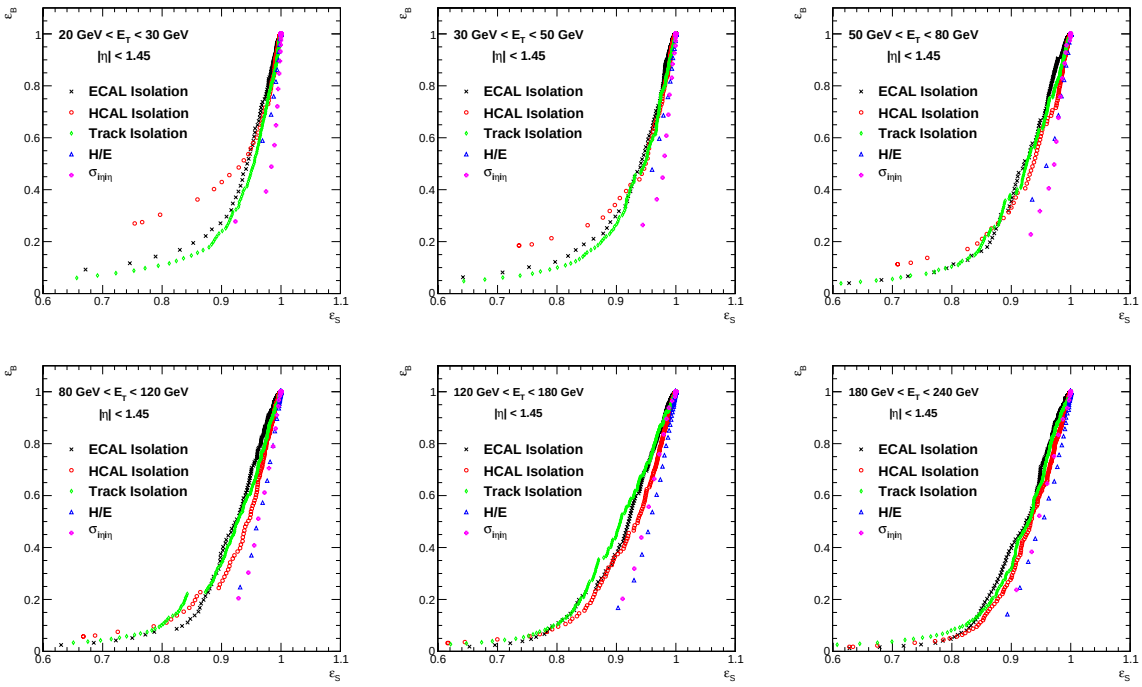


Figure 6.7: Efficiency-rejection curves for the Iso_{ECAL}, Iso_{HCAL}, Iso_{TRK}, H/E and $\sigma_{i\eta i\eta}$ variables for photon candidates in the barrel. Six bins in transverse energy between 20 and 240 GeV are shown.

- Iso_{TRK}
- Iso_{ECAL}
- Iso_{HCAL}
- H/E
- $\sigma_{i\eta i\eta}$

The designed isolation variables allow to extract the selection efficiency from data. Due to the choice of the veto regions in the isolation cones, the selection efficiencies for prompt electrons and photons are expected to be very similar. This implies that electron control samples can be used to study the photon selection efficiency.
Similar considerations can be drawn concerning the use of the $\sigma_{i\eta i\eta}$ variable. Looking only at the η direction allows to minimise the effects due to the imperfect description of the detector material, which are expected to play a role mostly in the direction of the bending caused by the magnetic field. For the same reason, prompt electrons and photons are expected to have similar shower profiles in the η direction, which is less affected by bremsstrahlung and pair-production, thus allowing to use electron control samples to study photon identification variables.

In order to obtain the most simple selection criteria, an approach based on simple unidiemensional selections (*rectangular cuts*) was chosen. Figures 6.7 and 6.8 show the background cut efficiency as a function of the signal cut efficiency for the identification variables. The curves are plotted for different transverse energies and pseudorapidities of the photon candidates. It can be seen that the discriminating power of the different variables evolves with the energy of the candidate and that depending on the desired signal efficiency different choices can be made. The correlation between the identification variables is reported in Table 6.2, where it can be seen that non-negligible correlations are present.

Table 6.2: Correlation matrix for photon identification variables in the barrel and endcap regions. The reported numbers correspond to correlation coefficients in signal (background).

Variable	Track Isol.	ECAL Isol.	HCAL Isol.	H/E	$\sigma_{i\eta i\eta}$
			Barrel		
Tracker Isol.	1	+0.21 (+0.32)	+0.08 (+0.32)	+0.09 (+0.16)	-0.03 (+0.11)
ECAL Isol.		1	+0.14 (+0.29)	+0.00 (+0.21)	+0.10 (+0.18)
HCAL Isol.			1	-0.02 (+0.26)	+0.02 (+0.11)
H/E				1	-0.05 (+0.13)
$\sigma_{i\eta i\eta}$					1
			Endcaps		
Tracker Isol.	1	+0.17 (+0.32)	+0.16 (+0.31)	+0.19 (+0.14)	-0.11 (-0.07)
ECAL Isol.		1	+0.06 (+0.36)	+0.02 (+0.12)	-0.06 (-0.04)
HCAL Isol.			1	+0.12 (+0.40)	-0.02 (-0.03)
H/E				1	-0.09 (+0.07)
$\sigma_{i\eta i\eta}$					1

A requirement on the selection criteria, stemming from the robustness and simplicity arguments, is that of yielding a signal efficiency as uniform as possible as a function of the photon E_T and η. Even though not essential, this requirement allows to simplify the assumptions on the signal distribution when unfolding the effects. In order to obtain a constant efficiency as a function the photon transverse energy, the residual contribution to the isolation sums due to the photon foot-print have been subtracted. Furthermore, the range of the allowed rectangular cuts has been restricted such that the efficiency variation as a function of the photon pseudorapidity is less than 2%.
To account for the different constraints and correlations between variables, an iterative algorithm was developed to tune the selection criteria. The goal of the algorithm was to maximise the background rejection for a given signal efficiency. In order to avoid local maxima and to handle the correlation between variables, the cut optimisation was performed iteratively. Given a target signal efficiency ε_S and a maximum number of iterations n_{iter}, the algorithm works as follows:

1. For each step, an intermediate target signal efficiency is defined, in such a way that $\varepsilon_i > \varepsilon_{i+1} > .. > \varepsilon_{n_{iter}} = \varepsilon_S$ and a set of rectangular cuts $\mathcal{C}_i$ is obtained in such a way that $\mathcal{C}_i \supseteq \mathcal{C}_{i+1} \supseteq .. \supseteq \mathcal{C}_{n_{iter}}$

2. The background vs signal efficiency curves are obtained for all variables, given the set of rectangular cuts from the previous step $\mathcal{C}_{i-1}$.

3. All cuts for witch the background efficiency is smaller than the signal efficiency are examined. The cut that maximises the background rejection, while satisfying the constraint on the signal efficiency $\varepsilon_S(c) > \varepsilon_i$, is added to the set of rectangular cuts $\mathcal{C}_{i-1}$ to build the set $\mathcal{C}_i$.

4. The algorithm stops when $i = n_{iter}$ or when no improvement is found, i.e. when no variables are selected on the third step.

The selection has been tuned using an inclusive sample of direct photons generated with PYTHIA, while the background was taken from an inclusive PYTHIA jet sample. Full detector simulation was performed and photon candidates with transverse energy above 20 GeV have been considered. To account for an observed (small) E_{T} dependence of the isolation variables, a term proportional to E_{T} was subtracted from each isolation sum. The slope of the scaling terms was obtained from a linear fit to the mean of each isolation sum as a function of E_{T}. The obtained values can be read from Table 6.3 and are between 1×10^{-3} and 3×10^{-3}. In order to avoid numerical problems, the number of iterations was varied finding that the algorithm is stable (i.e. it converges to the same set of rectangular cuts) for $n_{iter} \gtrsim 3n_{variables}$.

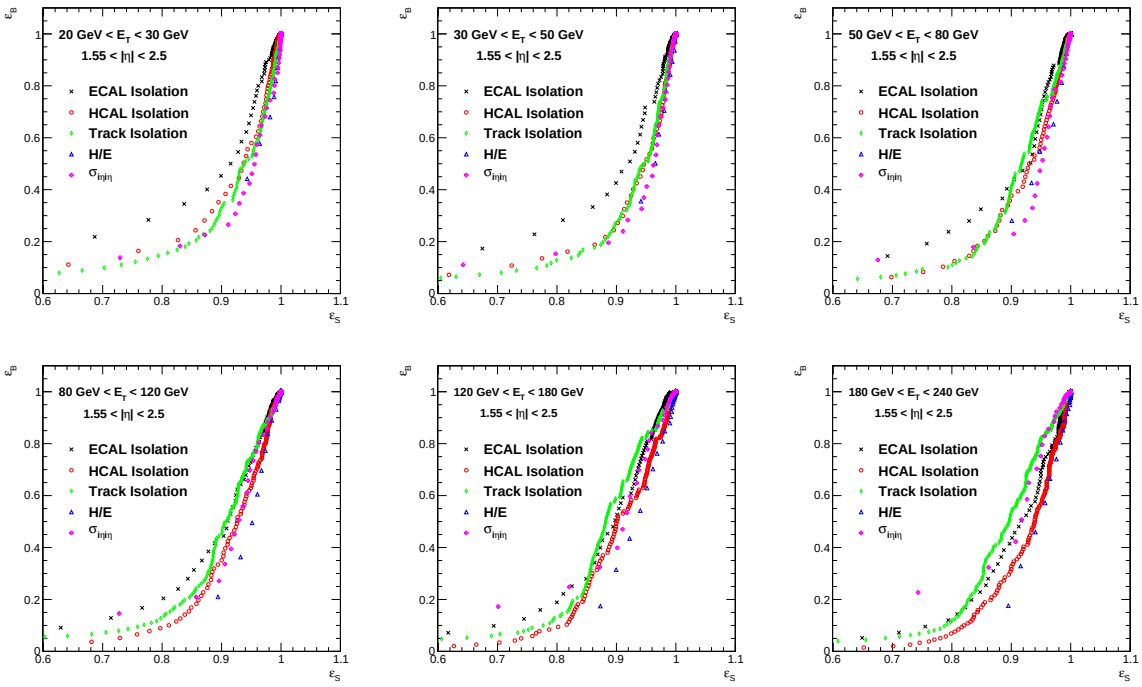


Figure 6.8: Efficiency-rejection curves for the $\mathrm{Iso_{ECAL}}$, $\mathrm{Iso_{HCAL}}$, $\mathrm{Iso_{TRK}}$, H/E and $\sigma_{i\eta i\eta}$ variables for photon candidates in the endcaps. Six bins in transverse energy between 20 and 240 GeV are shown.

The procedure has been repeated for different target efficiencies and restricting the allowed range of the rectangular cuts until a uniform signal efficiency as a function of η has been obtained. Finally, the set maximising the ratio of signal-to-background efficiency $\varepsilon_S(\mathcal{C})/\varepsilon_B(\mathcal{C})$ has been chosen. The set of selection criteria obtained with this procedure is shown in Table 6.3.

Table 6.3: Cut thresholds for the barrel and the endcap regions.

Variable	Barrel	Endcap
$\mathrm{Iso_{TRK}}$	< 2.0 GeV $+ 0.001 \cdot E_{\mathrm{T}}^{\gamma}$	
$\mathrm{Iso_{ECAL}}$	< 4.2 GeV $+ 0.003 \cdot E_{\mathrm{T}}^{\gamma}$	
$\mathrm{Iso_{HCAL}}$	< 2.2 GeV $+ 0.001 \cdot E_{\mathrm{T}}^{\gamma}$	
H/E	< 0.05	
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03

Given the geometrical definition of the veto cones, the set of cuts in Table 6.3 has very high efficiency for a sample of prompt electrons. In order to remove this possible source of background, the SC position is extrapolated to the interaction vertex and no match with hits in the pixel detector compatible with tracks from the hard interaction region is required.

The total efficiency of the selection criteria in table 6.3 for prompt photons of transverse energy above 20 GeV in the barrel is $\sim 91\%$ and the total variation as a function of the photon pseudorapidity is $\pm 1\%$, the veto on pixel seed hits brings the efficiency down to $\sim 88\%$ while keeping the variation as a function of the pseudorapidity unchanged. In the case of the endcaps, the selection criteria in table 6.3 have an efficiency of $\sim 94\%$, with a variation as a function of the pseudorapidity of $\pm 1\%$, while the pixel hits

veto brings the efficiency down to $\sim 80\%$ with a variation of $\pm 7\%$. Applying all selections, the signal-to-background ratio of an inclusive photon sample with transverse energy above 30 GeV is predicted to be $\gtrsim 0.6(0.4)$ in the barrel (endcaps) regions.
The signal-to-background ratio obtained with the described selection is relatively low. Higher background rejection factors could be achieved requiring stricter selection criteria. In this case, however, the details of the detector simulation would start to play an important role. In particular, the contribution of the calorimeter noise to the isolation sums would introduce a dependence of the signal efficiency on the supercluster position which would require a precise tuning of the simulation to be well modelled. Furthermore, imposing stricter selection criteria may introduce significance differences in the behaviour of the isolation algorithms in the case of isolated electron and photon, making more difficult the study of the selection efficiency in early data. For these reasons, a relatively loose set of selection criteria was chosen for startup analyses.

6.3 Commissioning of photon reconstruction and identification with collision data

The performance of the algorithms described in the previous sections was tested on pp collision data acquired during the 2010 LHC run. The study extended the one discussed in Chapter 5 and was divided in two parts. In the first part, the clustering algorithm performances were tested on a background-dominated sample, obtained applying loose selection criteria to the SC population. In the second part, the performances of the identification criteria described in section 6.2.3 were studied.
A sample of data recorded during 2010 proton-proton collisions, amounting to a luminosity of 77 ± 8 nb^{-1}, was compared to a sample of simulated events, generated with PYTHIA 6 [17], consisting of inclusive photon, inclusive jet and inclusive $W \to e\nu$ production. The D6T [92] tune was used in the simulation.

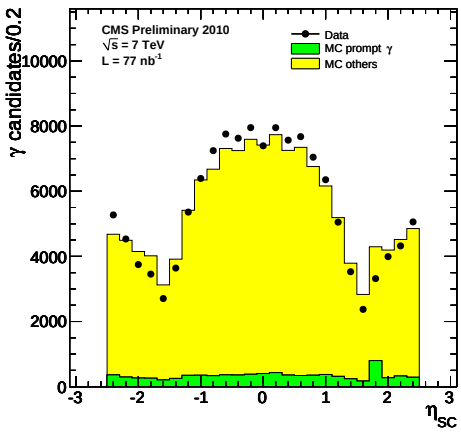


Figure 6.9: Detector pseudorapidity of super-clusters in 7 TeV minimum bias collisions.

6.3.1 Super-cluster selection and observables

In order to compare observables important to SC reconstruction between the data collected in 2010 and Monte Carlo simulations, the following selection was imposed:

- The events were required to pass the high-level trigger path HLT_Photon15_L1R, which requires the presence of a supercluster with uncorrected transverse energy above 15 GeV and is seeded by

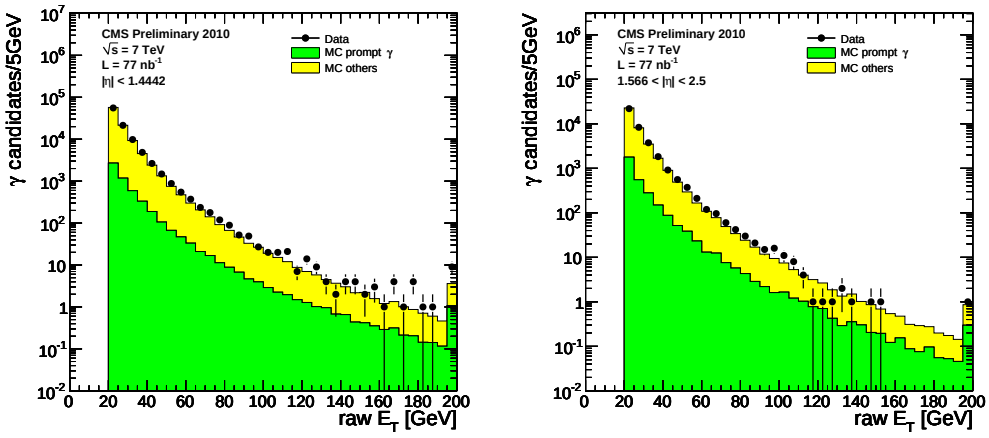


Figure 6.10: Uncorrected transverse energy spectra for EB (left) and EE (right) superclusters. The Monte Carlo results are normalised separately for each plot to the number of entries in the data histogram.

the Level-1 EG5 trigger [6].

- The SC *raw transverse energy*, corresponding to the quantity $G \cdot (\sum c_i \cdot A_i) \cdot \sin\theta_{SC}$ in the notation of Equation (6.1), was required to be greater than 20 GeV.

- Only super-clusters in the region covered by the silicon tracker detector ($|\eta_{SC}| \leq 2.5$) were considered. The region between the barrel and the endcaps ($1.4442 < |\eta_{SC}| < 1.566$), which is occupied to a significant extent by services, was excluded.

- Super-clusters containing anomalous ECAL signals were removed by imposing the criteria on topology and signal time.

- The hadronic to electromagnetic energy ratio was required to be $H/E < 0.05$ to reject jets with a substantial hadronic component, and to avoid differences with the simulation of the Level-1 trigger, which did not take into account the relaxed H/E cut used online at startup.

Figure 6.9 shows the pseudorapidity distribution of reconstructed super-clusters. Here the Monte Carlo and data statistics are normalised using the integrated luminosity in data. The measurements are fairly well reproduced by the Monte Carlo simulation. The dips at the transition between the ECAL Barrel and Endcap is due to the fiducial cut introduced above.
Figure 6.10 shows the distribution of the uncorrected transverse energy of reconstructed super-clusters. The supercluster E_T spectrum is well reproduced by the simulation over a wide range. Figure 6.11 shows the distribution of the number of basic clusters and the number of crystals per SC. The different patterns due to the behaviour of the Hybrid and Multi5x5 algorithms observed in data appear to agree well with the MC prediction.

Super-cluster energy corrections have been described in section 6.1.2. The most important variables needed to go from the super-cluster energy to the corrected photon energy are the *brem* and the R_9 variables. It is thus very important to verify the correct modeling of the two. Figure 6.12 shows the distributions of *brem* and R_9 for the barrel and endcap regions. In both cases, the shape of the distribution is similar in data and Monte Carlo. The fraction of the sample of photon candidates lying above or below the 0.94 (0.95) R_9 cut agrees with Monte-Carlo within 10%.

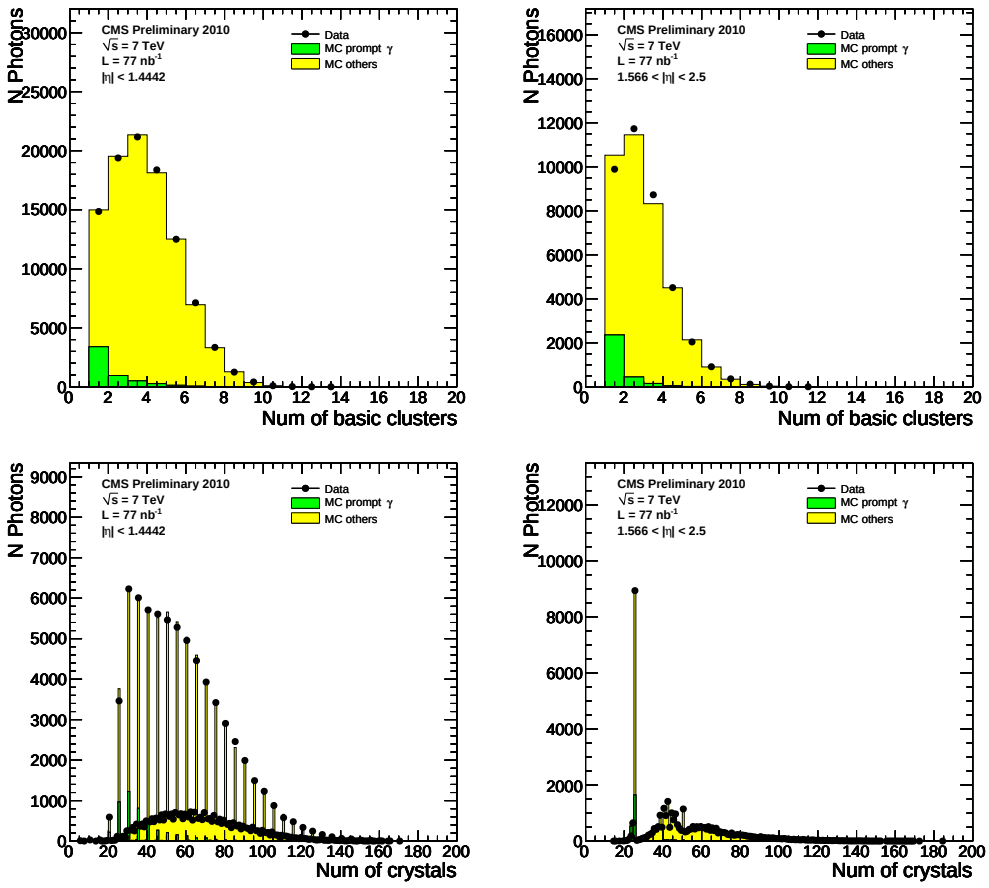


Figure 6.11: (top) Number of basic clusters per super-clusters for EB (left) and EE (right) super-clusters. (bottom) Number of crystals per super-clusters for EB (left) and EE (right) super-clusters. The Monte Carlo results are normalised separately for barrel and endcap regions to the number of entries in the data histogram.

6.3.2 Photon selection criteria

To increase the purity of the photon sample, the identification criteria described in section 6.2.3 were applied: the cuts reported in table 6.3 were imposed as well as the requirement of no matching pixel hits. nAs mentioned, with this selection, the MC predicts a purity of about 60(40)% signal photons with $E_T > 30$ GeV in EB (EE). The total number of photon candidates which fulfilled all the selection criteria in the analysed data sample is about 2.9×10^3.

The signal-enriched sample was used to compare photon variables in data to the Monte Carlo. Photon candidates in 2010 data were compared to the simulated sample mentioned above. For each variable, the MC distributions were normalised to the total number of candidates in data, fixing the relative normalisation of each MC component using the cross-section calculated by PYTHIA. The difference between this normalisation and the one predicted using the total integrated luminosity and the total PYTHIA cross-section varies case-by-case and is $\lesssim 15\%$ for all cases.
In each figure, the contribution due to isolated photons (including direct and fragmentation components) was separated from the one due to other sources. To compare the distribution of the variables entering the selection, all cuts except the one involving the variable under comparison were applied, obtaining

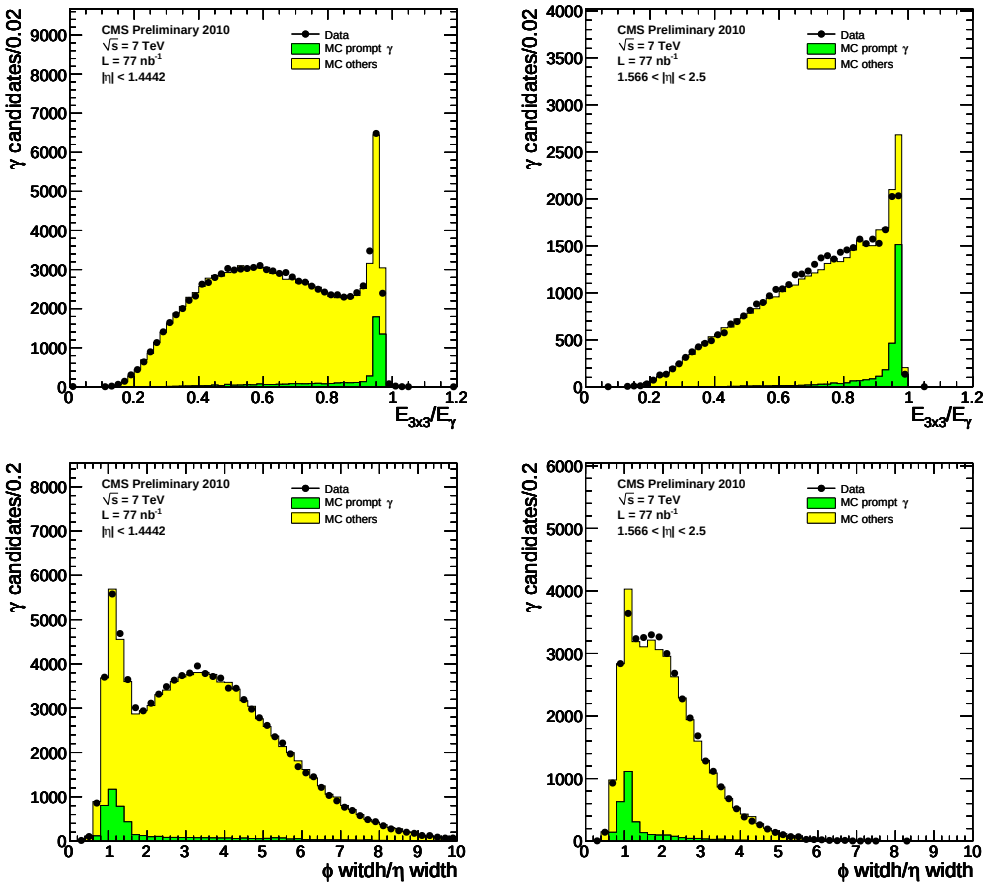


Figure 6.12: (top) Ratio of the energy of the 3x3 matrix of crystals surrounding the most energetic crystal to the total energy of the super-cluster, for barrel (left) and endcap (right) super-clusters. (bottom) Distribution of the *brem* variable described in the text for EB (left) and EE (right). The Monte Carlo results are normalised separately for barrel and endcap regions to the number of entries in the data histogram.

the so-called $N-1$ distributions. Overall, the simulation is observed to describe well the shape of the distributions.

Figure 6.13 shows the sum of $\mathrm{Iso}_{\mathrm{ECAL}}$, $\mathrm{Iso}_{\mathrm{HCAL}}$, and $\mathrm{Iso}_{\mathrm{TRK}}$ without any isolation cut applied. Figures 6.14-6.16 show the $N-1$ distributions for each isolation variable. In general the agreement of the simulation with the data is satisfactory. In some cases (e.g. the $\mathrm{Iso}_{\mathrm{TRK}}$) the observed isolation values are below what the simulation predicts. This effect is consistent with other studies performed by the CMS collaboration on soft particles produced in the underlying event [93, 26].

Figure 6.17 shows the $N-1$ plot for the pixel seed variable, where the fraction of candidates with a seed is consistent with the simulation, and the enhancement in signal purity for events without the seed can be seen. Figure 6.18 (left) shows the $N-1$ plot for the $\sigma_{i\eta i\eta}$ variable in EB. The signal contribution to the peak below 0.01 is clearly visible. Figure 6.18 (right) shows the $N-1$ plot for the $\sigma_{i\eta i\eta}$ variable in EE. Also here the signal contribution to the peak below 0.030 is clearly visible. A small shift of the distribution towards larger values is observed in data with respect to the simulation, which could

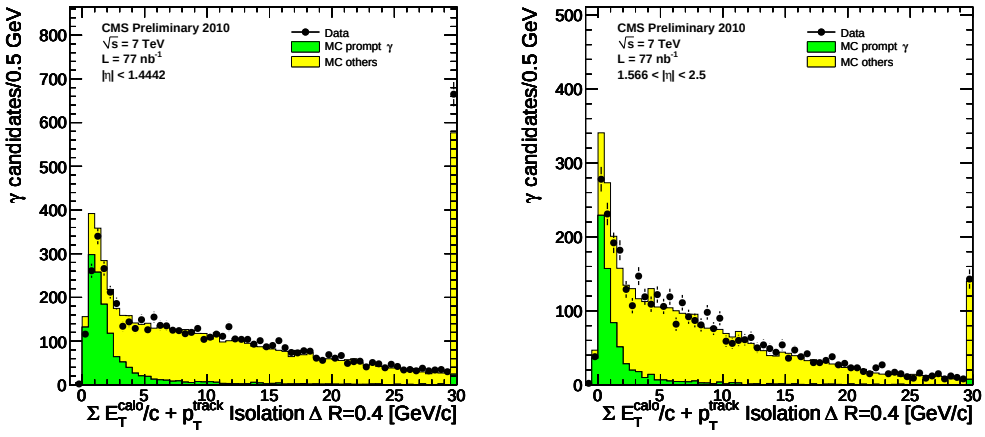


Figure 6.13: Sum of the isolation variables (ECAL, HCAL, and tracks) for barrel (left) and endcap (right) photon candidates, before applying photon isolation cuts. The Monte Carlo results are normalised separately for each plot to the number of entries in the data histogram.

be understood in terms of a miss-modeling of the detector noise or to the uncertainty in the channel-to-channel intercalibration constants. Finally, figure 6.19 shows the R_9 distribution with all selections applied. After photon ID the distribution has on average larger values, as expected from the simulation. Figure 6.20 shows the transverse energy spectrum of photon candidates in data with all selection criteria applied. The shape of the spectrum is observed to be well described by the simulation over a large energy range. Figure 6.21 shows the η distribution of the selected photon candidates. Also in this case, the MC seems to reasonably describe the distribution.

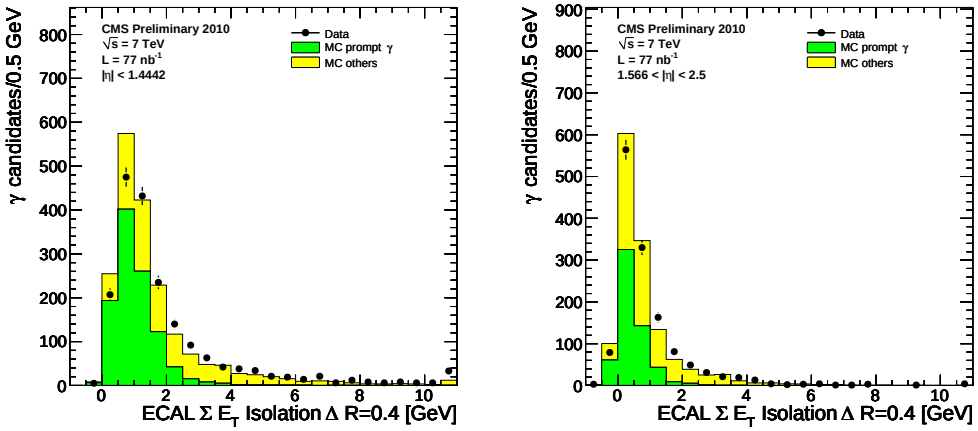


Figure 6.14: $N-1$ ECAL isolation distribution for data and MC, shown for barrel (right) and endcap (left). The Monte Carlo results are normalised separately for each plot to the number of entries in the data histogram.

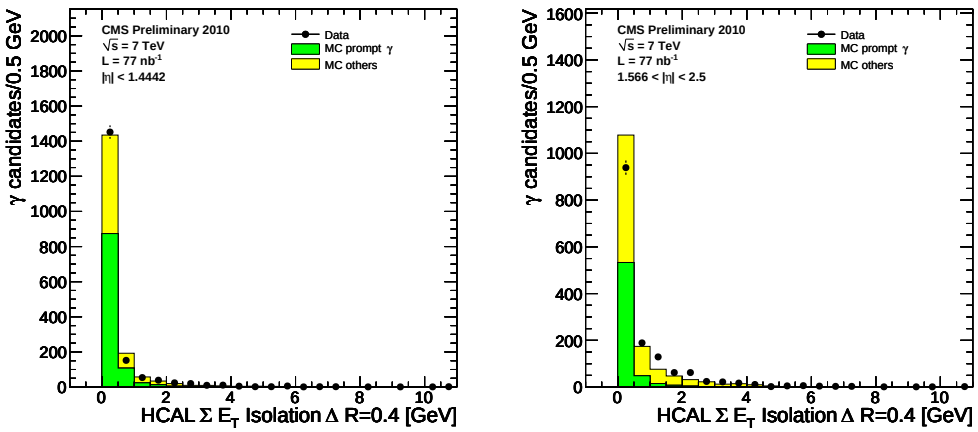


Figure 6.15: $N-1$ HCAL isolation distribution for data and MC, shown for barrel (right) and endcap (left). The Monte Carlo results are normalised separately for each plot to the number of entries in the data histogram.

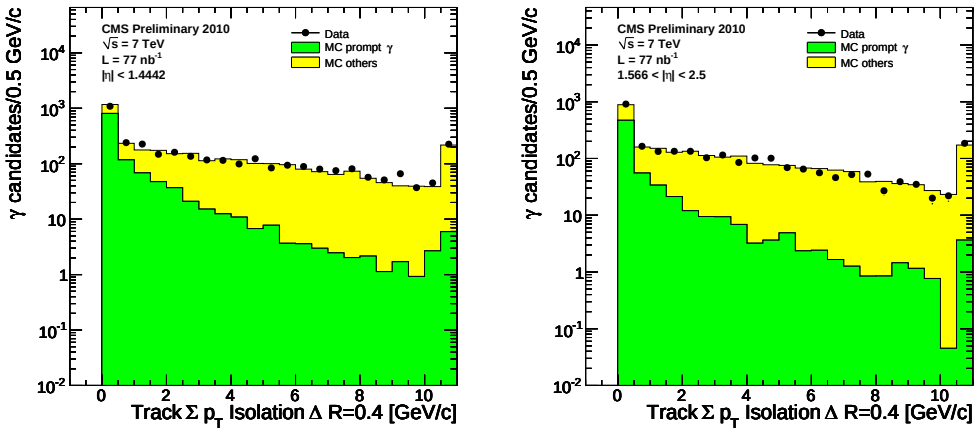


Figure 6.16: $N-1$ Track isolation distribution for data and MC, shown for barrel (right) and endcap (left). The Monte Carlo results are normalised separately for each plot to the number of entries in the data histogram.

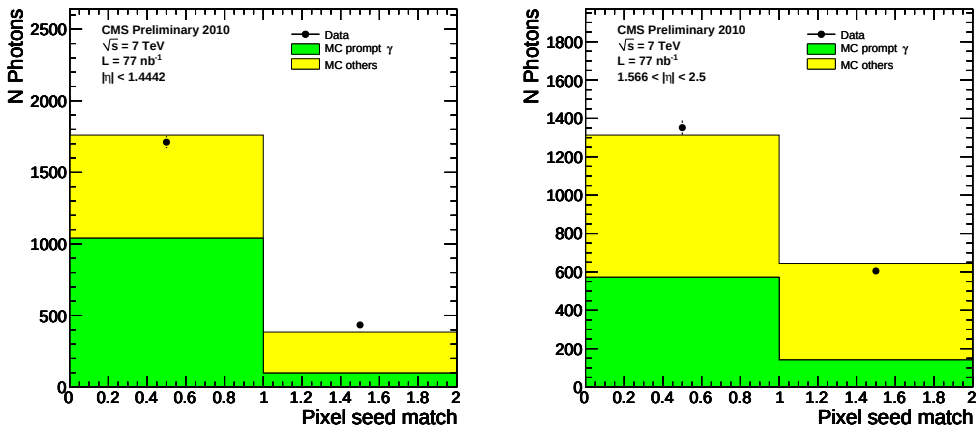


Figure 6.17: $N-1$ plot of the flag for the presence of a pixel seed for photons in EB (left) and EE (right), used in the selection in the text. The Monte Carlo results are normalised separately for each plot to the number of entries in the data histogram.

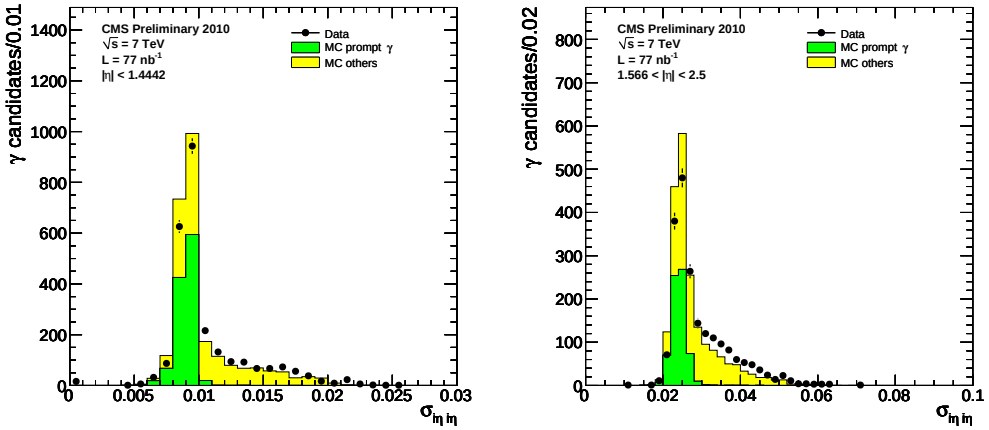
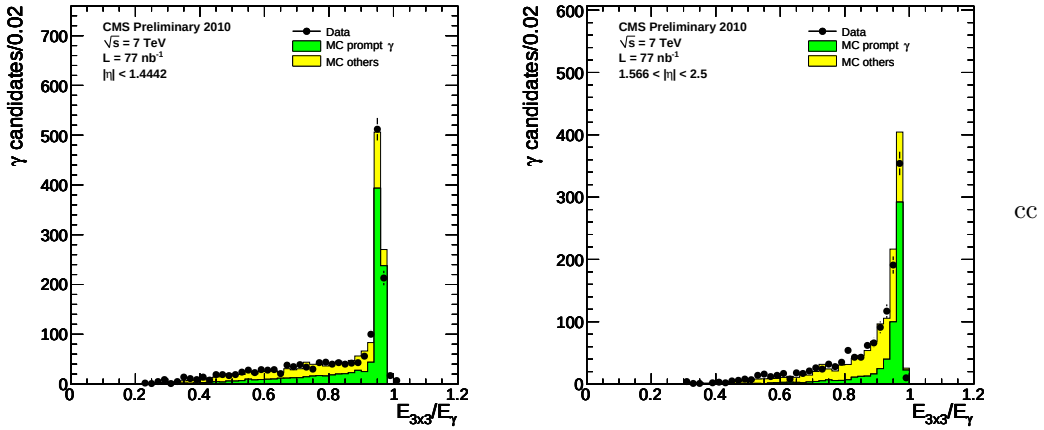


Figure 6.18: The $\sigma_{i\eta i\eta}$ shower shape variable for barrel (left) and endcap (right) photon candidates. The $N-1$ distributions are shown before cutting on the variables for photon identification. The Monte Carlo results are normalised separately for each plot to the number of entries in the data histogram.



cc

Figure 6.19: The R_9 variable of reconstructed photons in the selected sample in EB (left) and EE (right). The Monte Carlo results are normalised separately for each plot to the number of entries in the data histogram.

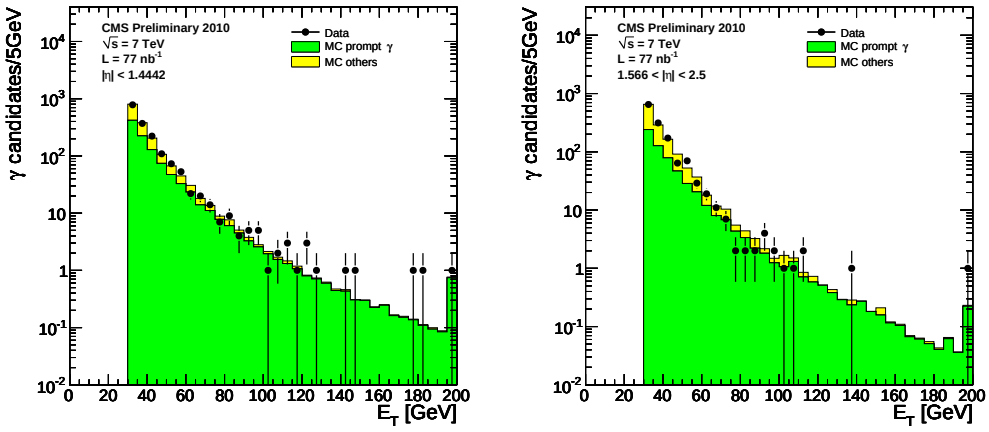


Figure 6.20: Transverse energies of EB (left) and EE (right) reconstructed photons in the selected sample. The purity increases with E_T. The Monte Carlo results are normalised separately for each plot to the number of entries in the data histogram.

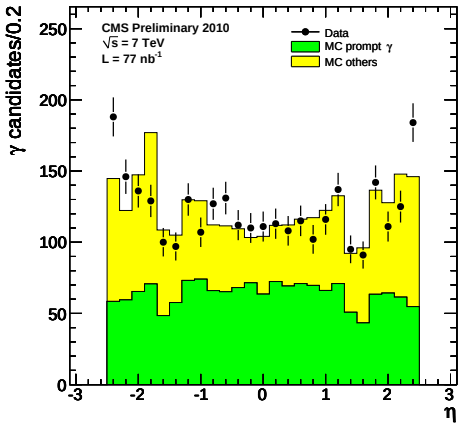


Figure 6.21: Pseudorapidity of reconstructed photons in the selected sample.

6.4 Conclusions

The photon reconstruction and identification techniques used in the CMS experiment were described in this chapter. The whole reconstruction process was reviewed and a photon identification selection aimed at early analyses was presented.
Using the first 7 TeV LHC data delivered to CMS, a series of studies on photon candidates was carried out. The basic reconstruction quantities for photons were compared to the Monte Carlo simulation and good agreement was observed.
Using the photon identification selection presented in this chapter, it was possible to enrich the sample in signal photons and suppress the background from jet production. It was then possible to verify that the photon reconstruction and identification performance is similar to that expected from simulation.
With this chapter the description of the experimental tools and techniques used in this thesis to study proton-proton collisions is completed. The last two chapters are devoted to the application of such techniques to the measurement of the inclusive photon production and associated Wγ production at the LHC.

Chapter 7

Inclusive photon production in pp collisions at $\sqrt{s} = 7$ TeV

Contents

Photon production in pp collision was introduced in Chapter 2. As stated there, the measurement of isolated photon production in pp collisions provides a test of perturbative quantum chromodynamics (pQCD) and the possibility to constrain the parton distribution functions (PDF). This complements deep inelastic scattering, Drell-Yan pair production and jet production measurements [94]. At the LHC, a significantly higher centre-of-mass energy than at the Tevatron collider [95] allows the exploration of new kinematic regions in the hard-interaction processes, providing additional constraints to pQCD and the PDFs [37].

This chapter presents a measurement of the E_{T}-differential inclusive isolated photon cross section in pp collisions at $\sqrt{s} = 7$ TeV. The analysed dataset corresponds to 2.9 ± 0.3 pb^{-1} of integrated luminosity [52] recorded by the CMS detector at the LHC up to September 2010. Isolated photons with a pseudo-rapidity $|\eta^{\gamma}| < 1.45$ and a transverse energy $E_{\mathrm{T}}^{\gamma} > 21$ GeV were investigated. The isolated photon production cross section was measured differentially in 11 bins in the photon E_{T}. The kinematic domain $0.006 < x_{\mathrm{T}} = 2E_{\mathrm{T}}^{\gamma}/\sqrt{s} < 0.086$ was probed, extending the region explored by previous experiments. Prompt photons were separated from neutral meson decays on a statistical basis, exploiting the different topologies of the electromagnetic energy deposits of the two populations.

In the following, the event selection and signal extraction process is described in detail. The sources of systematic uncertainties are then evaluated and the results are compared with pQCD NLO calculations.

7.1 Datasets

The full data sample processed for the analysis amounts to $2.9 \pm 0.3\,\mathrm{pb}^{-1}$. In these data, the lowest unprescaled photon trigger E_{T} threshold changed, following the LHC instantaneous luminosity increase. Only events selected by non-prescaled trigger paths were retained. Consequently, for the lowest E_{T} bins ($21 < E_{\mathrm{T}} < 26\,\mathrm{GeV}$) the last group of runs in Table 7.1 was not considered and the integrated luminosity analysed for these bins is effectively $2.1 \pm 0.2\,\mathrm{pb}^{-1}$.

The help of Monte Carlo simulations (MC) was necessary at several stages of the analysis. For this purpose, events generated with the PYTHIA 6.4 MC generator [17], using the D6T tune [92], were employed. Full detector response was simulated assuming start-up conditions for detector alignment and calibration.

The list of simulated samples used to model signal and the hadronic background is provided in Table 7.2. In some cases, the size of the MC samples used to model the background processes corresponds to an integrated luminosity considerably lower than the one analysed in data. In addition to those in Table 7.2, simulated samples with W and Z bosons decaying to electrons were used in some of the studies.

For computational reasons, the sample production was organised by splitting the different processes in several bins, according to the transverse momentum exchanged in the partonic interaction ($\hat{p_{\mathrm{T}}}$). In the analysis, bins were combined, weighing each according to the corresponding cross section. The overlaps between bins were removed and for each region of phase space only events from the bin with the lowest weight were employed.

Table 7.1: Trigger thresholds used in the analysis. The E_{T} ranges in which each trigger was used and the corresponding integrated luminosity are reported.

HLT path	L1^{thr} (GeV)	HLT^{thr} (GeV)	E_{T}^{γ} (GeV)	$\int Ldt$ (pb^{-1})
HLT_Photon10_L1R	5	10	21–300	3.46×10^{-6}
HLT_Photon15_L1R	8	15	21–300	1.29×10^{-2}
HLT_Photon15_Cleaned_L1R	8	15	21–300	4.20×10^{-1}
HLT_Photon20_Cleaned_L1R	8	20	21–300	1.72
HLT_Photon25_Cleaned_L1R	8	25	26–300	6.99×10^{-1}

7.2 Event selection

The offline photon reconstruction algorithms were discussed in detail in Chapter 6. Dedicated algorithms are implemented in the CMS trigger system to select events with large energy depositions in the ECAL. The e/γ first level trigger was extensively discussed in Chapters 3 and 5. In the high level trigger (HLT), energy deposits in ECAL are reconstructed using the same algorithms used offline. Events with energy deposits exceeding a programmable E_{T} threshold are stored for further analysis.
As explained in Chapter 5, anomalous interactions in the ECAL barrel have been observed in collision events. These type of events are consistent with a direct ionisation of the APD, leading to seemingly high energy deposits localised in a single crystal [59]. In order to limit the trigger rate, super-clusters due to anomalous interactions are rejected at the HLT level applying the "swiss cross" rejection criterion described in Chapter 5. The corresponding HLT paths are labelled as *Cleaned* in Table 7.1.
As anticipated, the analysed dataset was acquired with varying trigger conditions and for each period the unprescaled trigger path corresponding to the lowest threshold was employed. The summary of trigger thresholds used for the different periods is detailed in Table 7.1. For each photon E_{T} bin only fully efficient, non-prescaled, triggers were considered. The total integrated luminosity recorded with each trigger is reported in the same table along with the E_{T} bins to which it contributed data.
Events passing the trigger selection were required to satisfy an additional set of criteria. Only events with at least one good quality primary vertex were considered. The vertex was required to be within 18 cm of the nominal centre of the detector along the beam axis (3σ of the luminous region) and to have at least four degrees of freedom in the fit [96].
Finally, for events with more than 10 reconstructed tracks, at least 25% of them were required to be of good quality.

Table 7.2: Summary of MC samples used in the analysis. All samples were produced using PYTHIA with the D6T tune and full detector simulation was performed assuming start-up conditions for calibration and alignment.

photon-jet (PYTHIA process `MSEL=10`)			
$\hat{p_{\mathrm{T}}}$ range(GeV/c)	Num. events	σ (pb)	Equiv. $\int \mathcal{L}dt$ (pb^{-1})
15–∞	$1.2 \cdot 10^6$	192200	6
30–∞	$1.0 \cdot 10^6$	20070	51
80–∞	$3.2 \cdot 10^6$	556	5789
170–∞	$1.1 \cdot 10^6$	24	46754

di-jet (PYTHIA process `MSEL=1`)			
$\hat{p_{\mathrm{T}}}$ range(GeV/c)	Num. events	σ (pb)	Equiv. $\int \mathcal{L}dt$ (pb^{-1})
15–∞	$6.2 \cdot 10^6$	$8.8 \cdot 10^8$	$7.1 \cdot 10^{-3}$
20–∞	$1.3 \cdot 10^7$	$290 \cdot 10^6$	$2.2 \cdot 10^{-1}$
80–∞	$3.2 \cdot 10^6$	923800	4
170–∞	$3.2 \cdot 10^6$	25470	124

7.3 Isolated photons selection

Events passing the trigger and event selections were required to also contain at least one isolated photon candidate.
Only photon candidates reconstructed in the fiducial region of the ECAL barrel were considered in the analysis: the SC pseudorapidity, with respect to the CMS reference frame was required to be $|\eta_{SC}| <$

1.4442.

Candidates originating in anomalous interactions in the ECAL were removed requiring topological selection criteria to be satisfied and asking that the timing of the most energetic crystal in the SC was consistent with collision events. The topological selection applied at the trigger-level ($1 - E_4/E_1 < 0.95$) was enforced at analysis level. Furthermore, candidates were required to have $\sigma_{i\eta i\eta} > 0.001$ and $\sigma_{i\phi i\phi} > 0.001$ and the measured hit time of the most energetic crystal (*seed crystal*) was required to be consistent with the collision timing within 5 standard deviations. This set of selections are referred as *cleaning criteria* in the following.

In order to cope with a bias in the timing reconstruction of high energy signals, the timing cleaning criterion was applied only to candidates with reconstructed negative timing when the energy deposited in the seed crystal exceeds 130 GeV.

For the analysis presented here, the ECAL energy scale (G term in Equation 6.1) used in the data reconstruction was corrected in order to account for the observed difference between the true Z mass and the one reconstructed in $Z \to ee$ decays. Namely, the energy scale was shifted by +0.83% with respect to the nominal value used in the reconstruction software [97]. The uncertainty associated to this value ($\pm 0.64\%$) was taken into account as a source of systematic uncertainty, as discussed later.

The photon candidates' transverse energy was required to be above 21 GeV. The candidates where further requested to fulfil the isolation criteria and H/E requirement that were discussed in Chapter 6. No selection was applied on the $\sigma_{i\eta i\eta}$ variable. The selection thresholds applied are reported in Table 7.3.

In order to remove isolated electrons, the SC position was extrapolated to the interaction vertex and no match with hits in the pixel detector compatible with tracks from the hard interaction region was required (*pixel seed veto* criterion).

Table 7.3: Photon identification criteria.

Variable	Selection
Iso_{TRK}	$< 2.0\,\text{GeV} + 0.001 \cdot E_{\text{T}}^{\gamma}$
Iso_{ECAL}	$< 4.2\,\text{GeV} + 0.003 \cdot E_{\text{T}}^{\gamma}$
Iso_{HCAL}	$< 2.2\,\text{GeV} + 0.001 \cdot E_{\text{T}}^{\gamma}$
H/E	< 0.05

After all selection criteria were applied, a sample of 405 016 photon candidates was retained. The distribution of the selection variables prior to the application of the corresponding cut (*N-1 plot*) is reported in Figure 7.1. The MC prediction for the same distributions, obtained mixing photon-jet, di-jet and W$\to e\nu$ events according to the cross sections calculated by PYTHIA, is also shown in the plots. Data and MC distributions are in overall good agreement. As already noticed in Chapter 6, small discrepancies, linked to the simulation of calorimeter noise and low momentum particles, can be observed. Such small discrepancies, however, do not affect the accuracy of the measurement as a correction to the signal efficiency was derived using an electron control sample in data.

The E_{T} and η distributions of the selected candidates is shown in Figure 7.2. It can be seen that the kinematic distributions of the selected candidates are well reproduced by MC.

7.4 Signal extraction

Signal and background were estimated from an extended maximum likelihood two-component fit to the $\sigma_{i\eta i\eta}$ distribution in data. For each E_{T} bin, the signal component shape ($\mathcal{S}$) was taken from Monte Carlo and corrected for differences between data and MC observed in an electron control sample. The background component shape ($\mathcal{B}$) was derived from data.

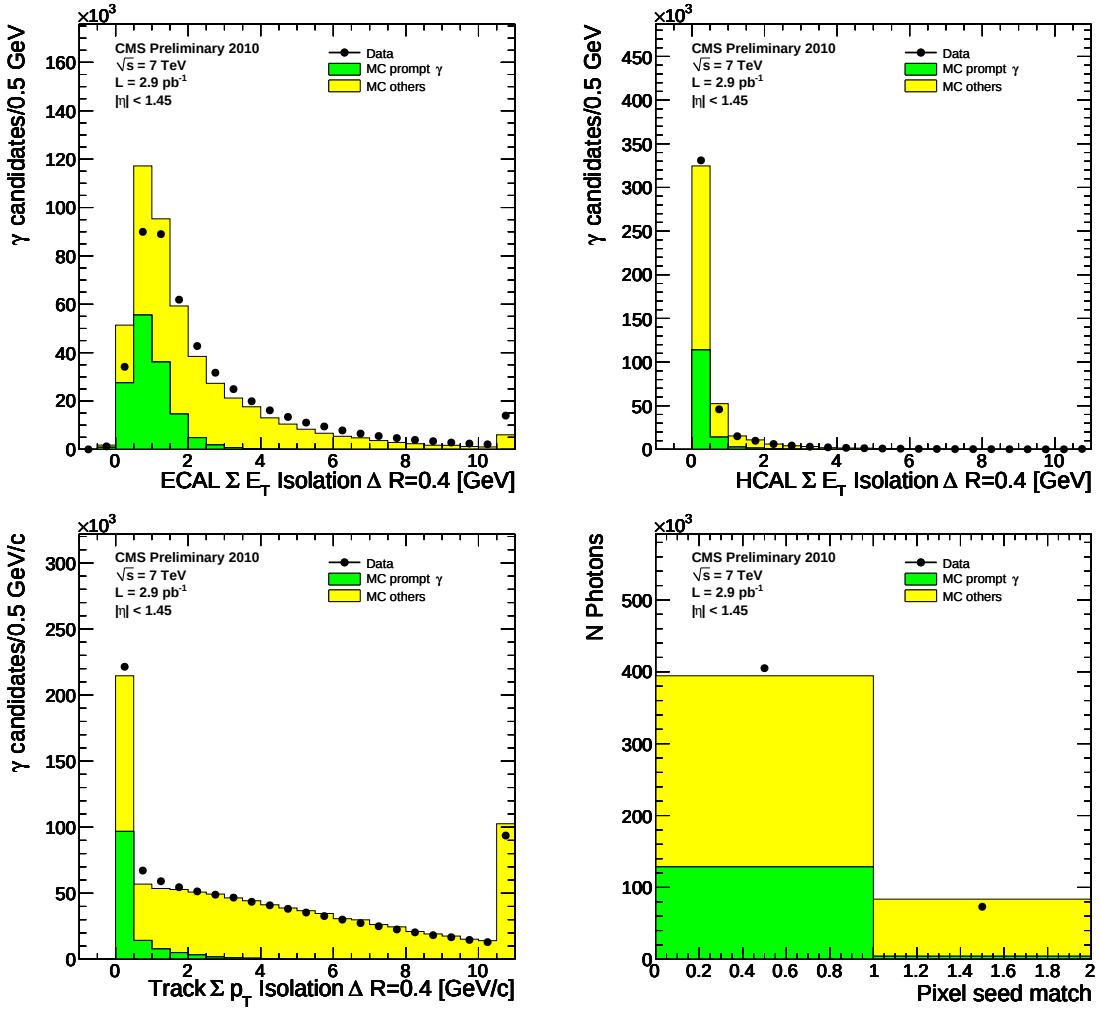


Figure 7.1: Distribution of variables used to select isolated photons. In each plot all the selection criteria are applied except the one involving the variable shown (*N-1 plots*). From left to right, top to bottom, Iso_{ECAL}, Iso_{HCAL}, Iso_{TRK} and the *pixel match* variable. The MC prediction is shown in each plot. Different MC components are normalised to each other according to the PYTHIA cross section. The total MC contribution normalisation is chosen to match the area of the corresponding histogram in data.

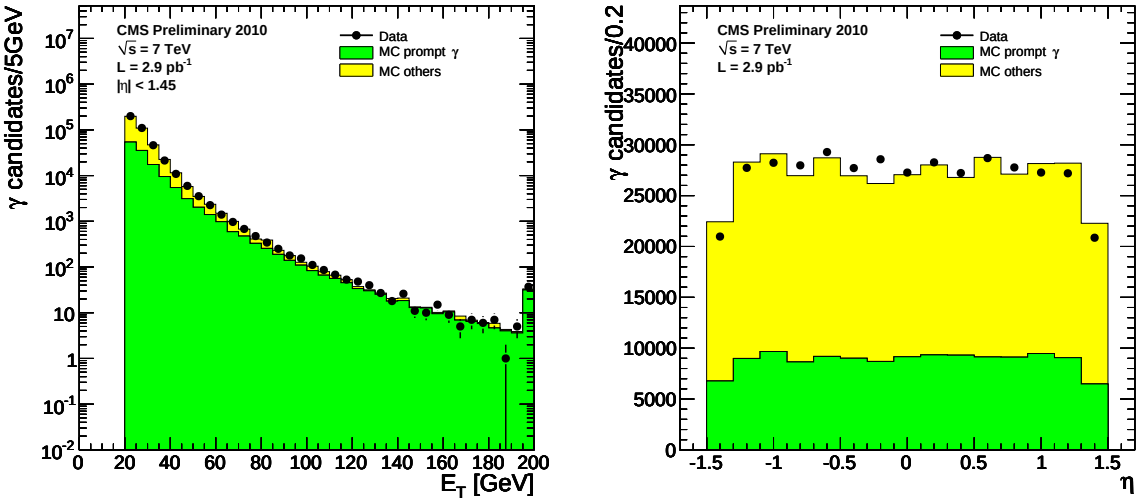


Figure 7.2: E_{T} spectrum (left) and η distribution (right) of selected photon candidates. The MC prediction is shown in each plot. Different MC components are normalised to each other according to the PYTHIA cross section. The total MC contribution normalisation is chosen to match the area of the corresponding histogram in data.

7.4.1 Signal component shape

The signal shape was derived from the PYTHIA Monte Carlo sample of photon-jet events, by taking the $\sigma_{i\eta i\eta}$ distribution of direct photons identified using MC truth information, as shown in Figure 7.3. Fragmentation photons were studied using the di-jet PYTHIA MC from which their $\sigma_{i\eta i\eta}$ distribution was extracted, as also shown in Figure 7.3.
Figure 7.3 shows that direct and fragmentation photons are identical in what regards $\sigma_{i\eta i\eta}$ and that this method effectively measures the sum of the two components.
In order to account for mis-modeling in the MC simulation, the $\sigma_{i\eta i\eta}$ distribution of isolated electrons was studied. For this purpose, events with a Z boson decaying to electrons were reconstructed using the photon reconstruction algorithms. Events with two "photon" candidates, required to match pixel hits consistent with tracks from the hard interaction and reconstructed in the ECAL barrel fiducial region, were selected. These candidates were required to fulfil the isolation and H/E criteria in Table 7.3. The purity of this sample was estimated to be $> 97\%$ in the mass range $60 < M_{ee} < 120\,\mathrm{GeV}/c^2$ (Figure 7.4, left).
As shown in Figure 7.4 (right), an overall shift was observed when comparing the distribution obtained in data to the MC prediction. The mean $\sigma_{i\eta i\eta}$ in data was found to be larger than the MC prediction by $(8 \pm 3) \times 10^{-5}$, corresponding to 0.9% of the $\sigma_{i\eta i\eta}$ mean. The effect was corrected by shifting the $\sigma_{i\eta i\eta}$ distributions from the photon-jet sample by $+8 \times 10^{-5}$.

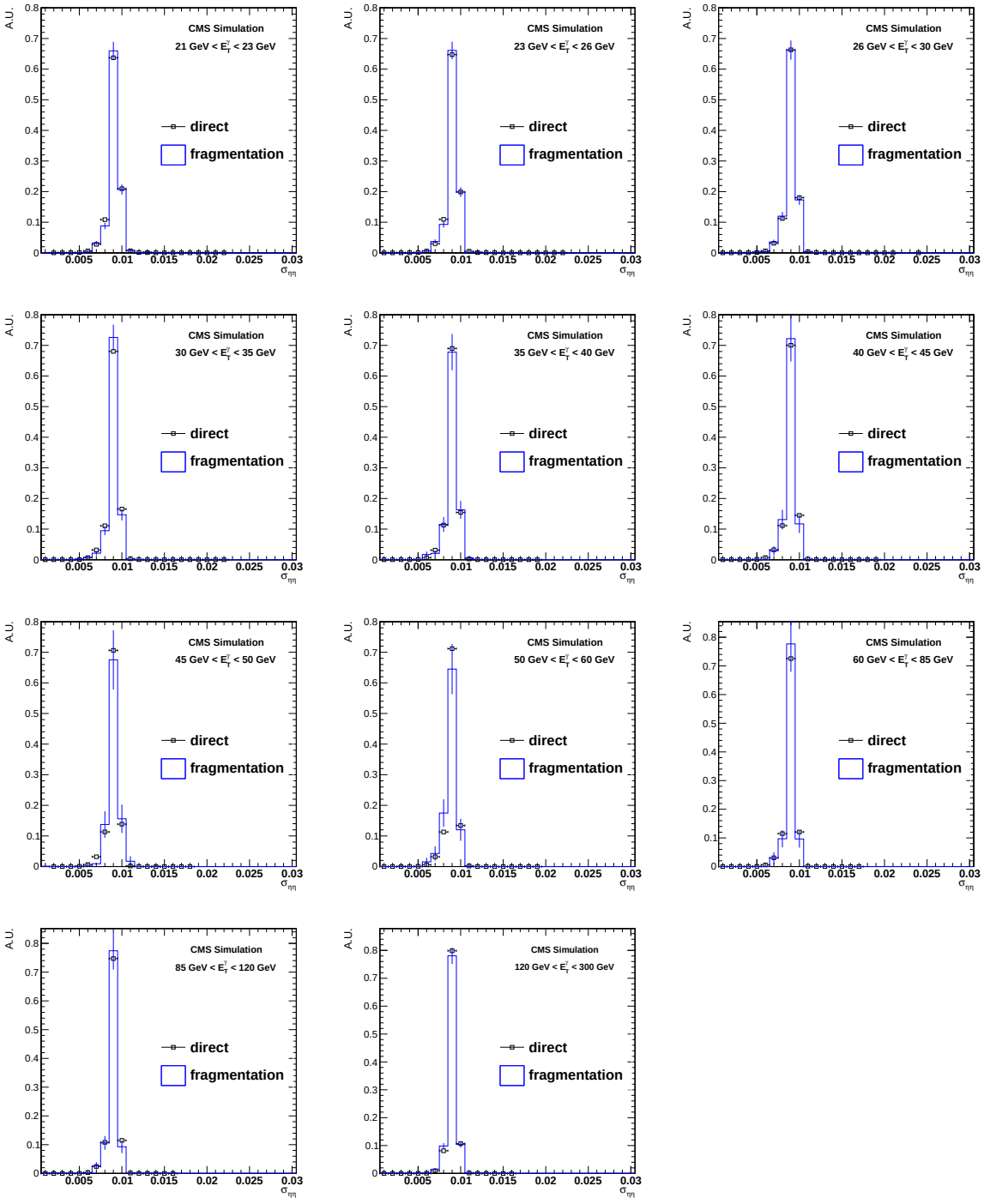


Figure 7.3: Distributions of $\sigma_{i\eta i\eta}$ for direct photons from a photon-jet MC sample and fragmentation photons from a di-jet MC sample. The two types of photons have the same $\sigma_{i\eta i\eta}$ distribution. The $\sigma_{i\eta i\eta}$ variable is labelled $\sigma_{\eta\eta}$ in the plots.

.

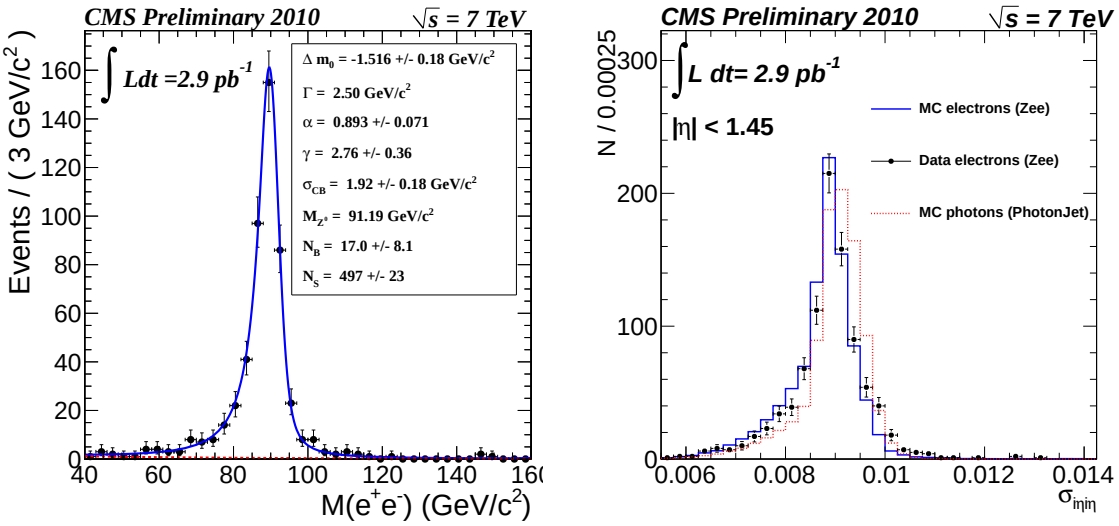


Figure 7.4: Isolated electron control sample used to correct the signal component shape. (left) Invariant mass of the selected electron pairs in data. (right) Comparison between $\sigma_{i\eta i\eta}$ values from electrons in Z decays between data and MC simulation.

7.4.2 Background component shape

The background shape was derived from data, by taking the $\sigma_{i\eta i\eta}$ distribution of a part of the photon candidates that fulfil all but the $\mathrm{Iso_{TRK}}$ selection criteria in Table 7.3, namely those with $2 < \mathrm{Iso_{TRK}} < 5\,\mathrm{GeV}/c$. This region can be thought of as a side-band to the signal one, $\mathrm{Iso_{TRK}} < 2\,\mathrm{GeV}/c$.
Figure 7.5 shows, as a function of E_T, the $\sigma_{i\eta i\eta}$ background component shapes obtained for:

- The side-band in data, labelled Data side-band.
- The side-band in MC, labelled MC side-band.
- The signal region in MC, requiring the photon candidates to fulfil the analysis isolation criteria and not to be direct or fragmentation photons, labelled MC background.

Comparing the two distributions obtained from MC simulations, it can be observed that the chosen side-band models well the background distribution. Thus the $\sigma_{i\eta i\eta}$ distribution obtained in data can be considered a good model of the background.

7.4.3 Two-component fit

For each photon E_T bin, data were fit with $f(\sigma_{i\eta i\eta}) = N_\mathcal{S}\mathcal{S}(\sigma_{i\eta i\eta}) + N_\mathcal{B}\mathcal{B}(\sigma_{i\eta i\eta})$, where $N_\mathcal{S}$ and $N_\mathcal{B}$ are the estimated number of signal and background events in the E_T bin.
The fit was performed using a binned extended maximum likelihood, by minimising:

$$\mathcal{L} = -\ln L = -(N_\mathcal{S} + N_\mathcal{B}) + \sum_{i=1}^{n} N_i \ln(N_\mathcal{S}\mathcal{S}_i + N_\mathcal{B}\mathcal{B}_i) \tag{7.1}$$

where n is the number of bins in $\sigma_{i\eta i\eta}$, N_i the observed number of events for the i-th $\sigma_{i\eta i\eta}$ bin and $\mathcal{S}_i$ and $\mathcal{B}_i$ the values of the corresponding components in that bin; N is the total number of data events in the given E_T bin.

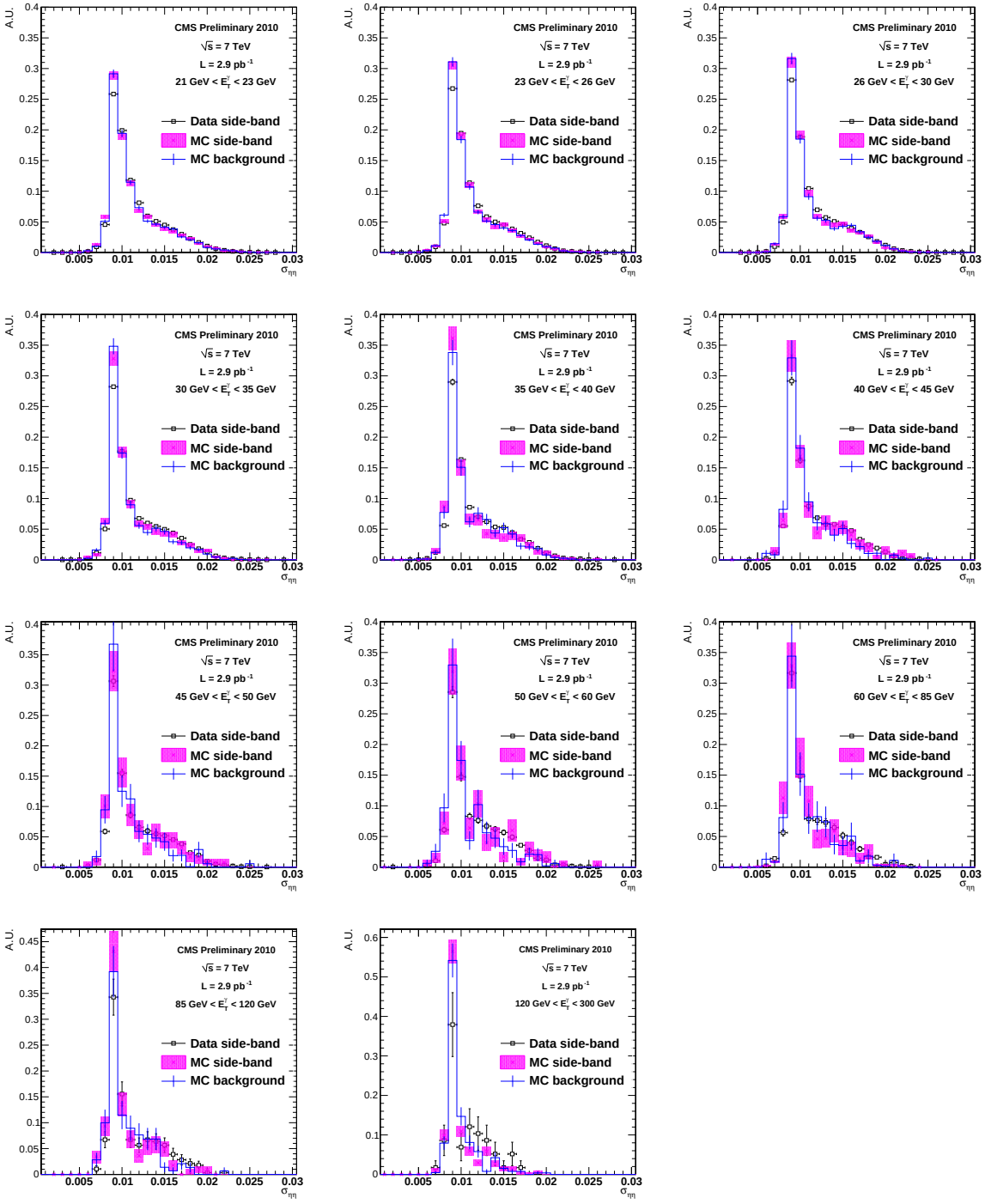


Figure 7.5: $\sigma_{i\eta i\eta}$ distributions for true background in MC (blue) and candidates from the $\mathrm{Iso_{TRK}}$ side-band in data (black) and MC (red) for each E_{T} bin. The $\sigma_{i\eta i\eta}$ variable is labelled $\sigma_{\eta\eta}$ in the plots.

The correction to the signal component shapes was let free to float independently for each E_T bin within the correction uncertainty. The two component fits were repeated for 5 values of the signal component shift (δ) in the range $5 \div 11 \times 10^{-5}$. The results obtained using the δ value that minimises a χ^2 test between the two-component fit results and the data was retained for each E_T bin. As discussed in Section 7.6.1, the full uncertainty on the signal-shape correction was taken into account when determining the cross section uncertainty.

Table 7.4: Estimated signal yield as a function of the photon E_T. For illustration, the fraction of isolated photons in the region $\sigma_{i\eta i\eta} < 0.01$ is also reported.

Photon E_T range (GeV)	Fitted signal yield	Signal fraction for $\sigma_{i\eta i\eta} < 0.01$
21–23	$(25.29 \pm 0.35) \times 10^3$	$33.9\% \pm 0.3\%$
23–26	$(24.24 \pm 0.32) \times 10^3$	$37.7\% \pm 0.3\%$
26–30	$(23.77 \pm 0.30) \times 10^3$	$41.6\% \pm 0.3\%$
30–35	$(15.38 \pm 0.22) \times 10^3$	$47.3\% \pm 0.4\%$
35–40	$(7.97 \pm 0.15) \times 10^3$	$51.6\% \pm 0.6\%$
40–45	$(4.73 \pm 0.11) \times 10^3$	$57.6\% \pm 0.9\%$
45–50	$(2.81 \pm 0.08) \times 10^3$	$60.7\% \pm 1.2\%$
50–60	$(3.04 \pm 0.08) \times 10^3$	$67.1\% \pm 1.1\%$
60–85	$(2.32 \pm 0.07) \times 10^3$	$71.5\% \pm 1.3\%$
85–120	635 ± 32	$79.2\% \pm 2.6\%$
120–300	249 ± 17	$95.2\% \pm 2.6\%$

The raw yield of isolated photons, obtained from the N_S for each E_T bin is shown in Figure 7.6 (left) and reported in Table 7.4. The results of the fits in all E_T bins are shown in Figure 7.7.
The fraction of isolated photons in data, for the isolated region ($\sigma_{i\eta i\eta} < 0.01$) was estimated as $N_S/(N_S + N_B)$ in that region and is shown in Figure 7.6 (right) and also reported in Table 7.4. The level of this fraction is a direct consequence of the loose selection used. The available dataset did not allow to determine the photon identification efficiency (ϵ_{ID}) fully from data. First, it was not possible to select a pure sample of isolated photons with sufficiently high statistics in data, thus the photon identification efficiency needed to be studied with isolated electrons from $Z \rightarrow ee$ decays. Furthermore, the number of available Z events was not sufficient to study the efficiency differentially in E_T and η. Thus, the measurement of the selection efficiency had to partially rely on the use of Monte Carlo simulation. In this scenario, a loose selection allowed the calculation of ϵ_{ID} (see Section 7.5.1) while minimising the dependence on the details of the isolation modeling in MC.

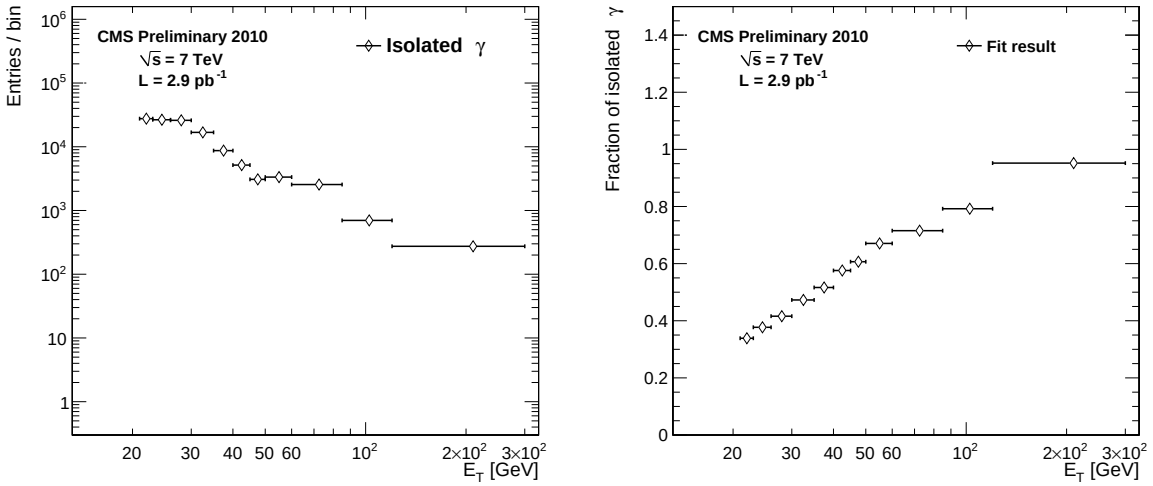


Figure 7.6: (left) Spectrum of isolated prompt photons as a function of E_{T} obtained from the two component fits. (right) Fraction of isolated photons in the signal region ($\sigma_{i\eta i\eta} < 0.01$) as a function of photon E_{T}.

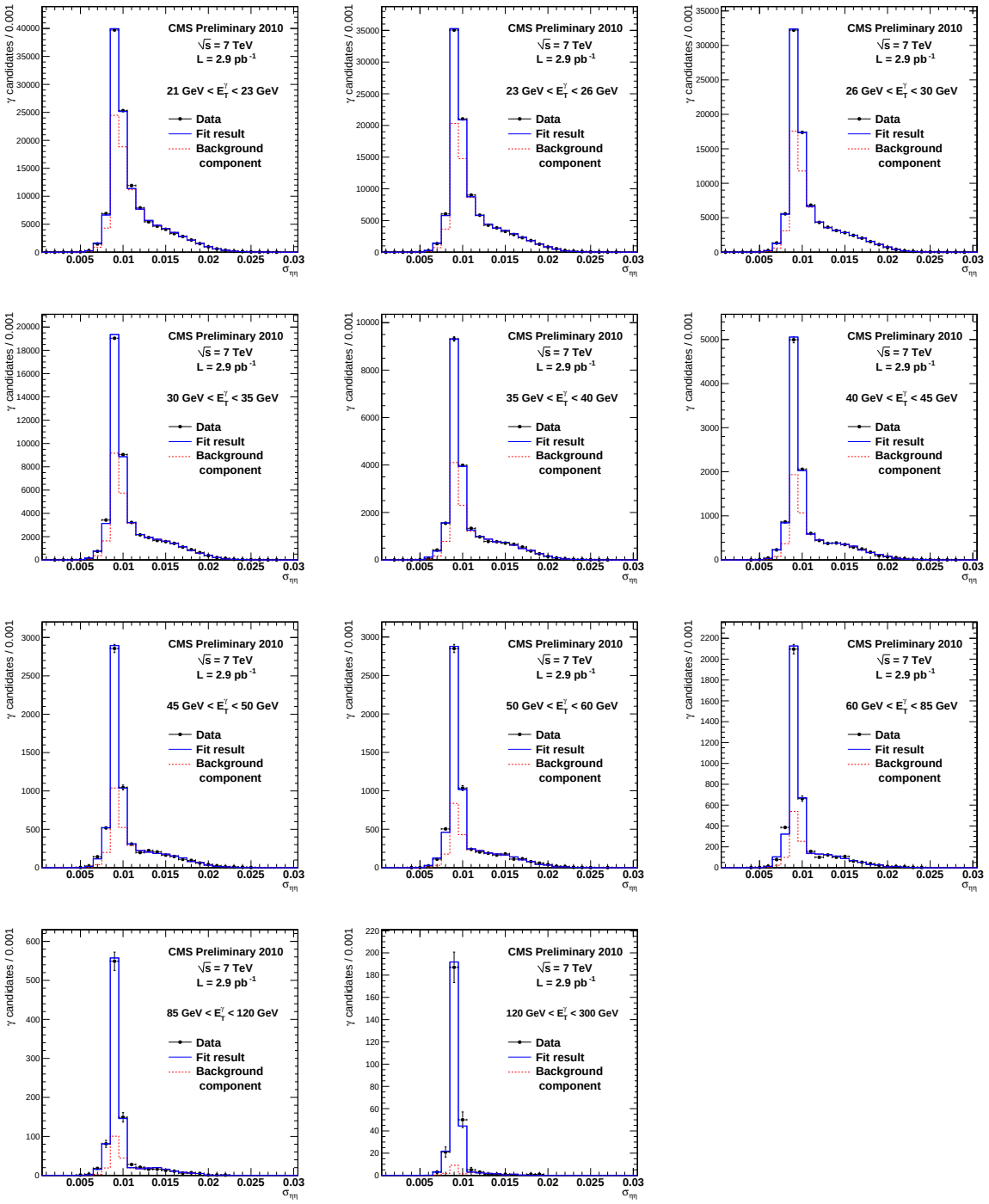


Figure 7.7: Measured $\sigma_{i\eta i\eta}$ distributions (points with error bars). The extended-ML fit result (solid blue) is overlaid in each plot and the component for background (dashed red) is shown separately. The $\sigma_{i\eta i\eta}$ variable is labelled $\sigma_{\eta\eta}$ in the plots.

7.4.4 (No) Contamination from anomalous interactions in ECAL

The residual contamination from anomalous interactions in the ECAL barrel was estimated extrapolating their contribution from a control region to the signal region. The study was restricted to the phase space region where the signal is concentrated, requiring all selection criteria used in the analysis to hold and additionally requiring $\sigma_{i\eta i\eta} < 0.01$.
To study the anomalous interactions, a sample of photon candidates was pre-selected requiring only the $1 - E_4/E_1 < 0.95$ cleaning. The latter removes the large majority of anomalous interaction events, which are in fact observed to produce very isolated signals in the calorimeter. A small subset of anomalous interactions, however, was observed to occur in adjacent crystals and are only removed by the timing requirement. This is well illustrated by Figure 7.8 (left) which shows the scatter plot between the reconstructed time of the most energetic crystal (t_{seed}) in the SC and E_{2nd}/E_8 variable. The latter is defined as the ratio between the energy deposited in the second most energetic crystal to the total energy deposited in the 8 crystals of the 3×3 crystal frame around the most energetic one. The small excess of candidates in the region $E_{2nd}/E_8 < 0.9$ and $t_{seed} > 1$ ns was verified to be due to a few channels with imperfect time calibration, as shown in Figure 7.8 (right). This excluded an anomalous origin.

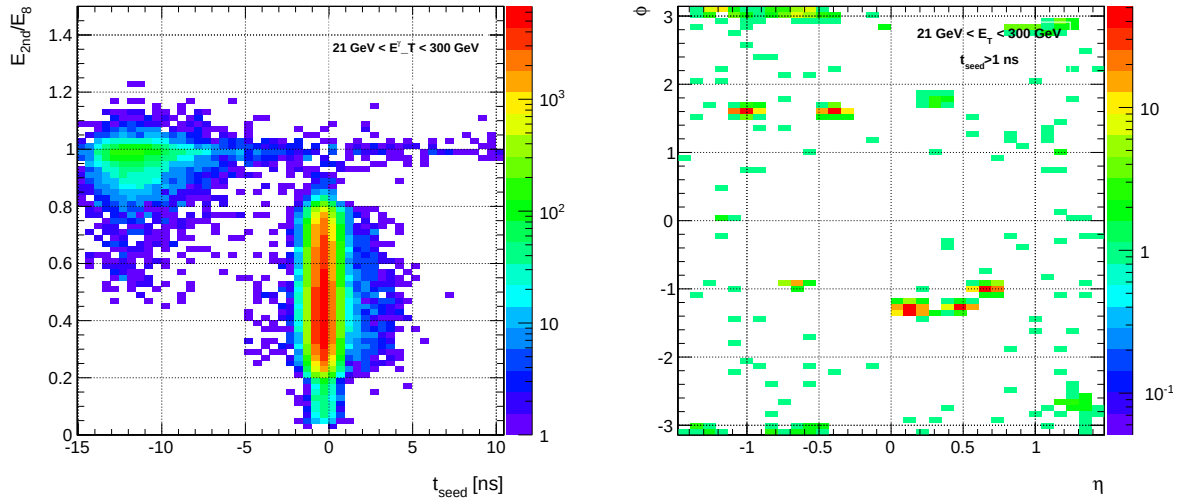


Figure 7.8: (left) Correlation between the measured seed crystal timing and the E_{2nd}/E_8 variable described in the text. The contribution from anomalous events is visible as the band at $E_{2nd}/E_8 \sim 1$. (right) $\eta - \phi$ map for events with $E_{2nd}/E_8 < 0.9$ and $t_{seed} > 1ns$ showing that the out-of-time contribution at positive times is due to mis-alignment of a few channels.

In order to estimate the remaining contamination after all cleaning criteria are applied, three control regions were defined:

- *Tagged out-of-time anomalous signals* were selected requiring $\sigma_{i\eta i\eta} < 0.001$ or $\sigma_{i\phi i\phi} < 0.001$ and asking the timing of the most energetic crystal not to be consistent with collisions timing, in particular the phase between the nominal collision time and the measured crystal timing was required to be between -12 and -8ns. The number of candidates in this region is labelled N_{out}^{tag}.

- *Non-tagged out-of-time anomalous signals* were required to have $\sigma_{i\eta i\eta} > 0.001$, $\sigma_{i\phi i\phi} > 0.001$ and timing not consistent with collision events. Their number is labelled N_{out}^{no-tag}.

- *Tagged in-time signals* were selected asking $\sigma_{i\eta i\eta} < 0.001$ or $\sigma_{i\phi i\phi} < 0.001$ and timing consistent with collision events. Their number is labelled as N_{in}^{tag}.

The signal region was finally defined by $\sigma_{i\eta i\eta} > 0.001$, $\sigma_{i\phi i\phi} > 0.001$ and timing consistent with collision events. The total number of candidates in this region is labelled N_{in}^{no-tag}.
The contamination from anomalous interactions to the signal region was estimated assuming that the timing spectrum of anomalous interaction events does not depend on their topology. Under this assumption, the contamination can be calculated as:

$$f_{anom} = \frac{N_{in}^{tag} \cdot N_{out}^{no-tag}}{N_{out}^{tag} \cdot N_{in}^{no-tag}}$$

The procedure was applied separately in E_{T} bins and is illustrated in Figure 7.9. The signal contamination from anomalous events in the different E_{T} regions is reported in Table 7.5. In order to obtain an upper limit on the contribution of anomalous interactions to the measured signal yield, the ratio of the estimated contamination to the purity extracted from the two component fit was calculated. It can be seen in Table 7.5 that the effect on the signal yield was estimated to be $\leq 0.21\%$ and is therefore negligible.

Table 7.5: Estimated contamination from anomalous events in the signal region ($\sigma_{i\eta i\eta} < 0.01$) after all selections. The second column shows the ratio between the estimated number of anomalous interaction events and the total number of events in each region. The fourth column shows the ratio between the estimated number anomalous interactions and signal events.

Photon E_{T} range (GeV)	Contamination	E_{T} range	Effect on signal
21–26	$0.041 \pm 0.002\%$	21–23	$0.15\% \pm 0.01\%$
		23–26	$0.14\% \pm 0.01\%$
26–30	$0.038 \pm 0.003\%$	26–30	$0.11\% \pm 0.01\%$
		30–35	$0.08\% \pm 0.01\%$
30–85	$0.04 \pm 0.02\%$	35–40	$0.07\% \pm 0.01\%$
		40–45	$0.06\% \pm 0.01\%$
		45–50	$0.06\% \pm 0.01\%$
		50–60	$0.05\% \pm 0.01\%$
		60–85	$0.05\% \pm 0.01\%$
85–120	$0.03 \pm 0.02\%$	85–120	$0.04\% \pm 0.02\%$
120–300	$0.09 \pm 0.10\%$	120–300	$0.10\% \pm 0.11\%$

7.4.5 Isolated electrons background

The background from isolated electrons was estimated measuring the probability for an isolated electron to satisfy all the identification criteria described in Section 7.3. Such probability was measured to be $1.9 \pm 0.5\%$ using an electron control sample from Z decays in data, with the tag and probe technique described in Section 7.5.1.
The fractional contribution from isolated electrons to the selected sample was estimated to be $< 10^{-3}$.

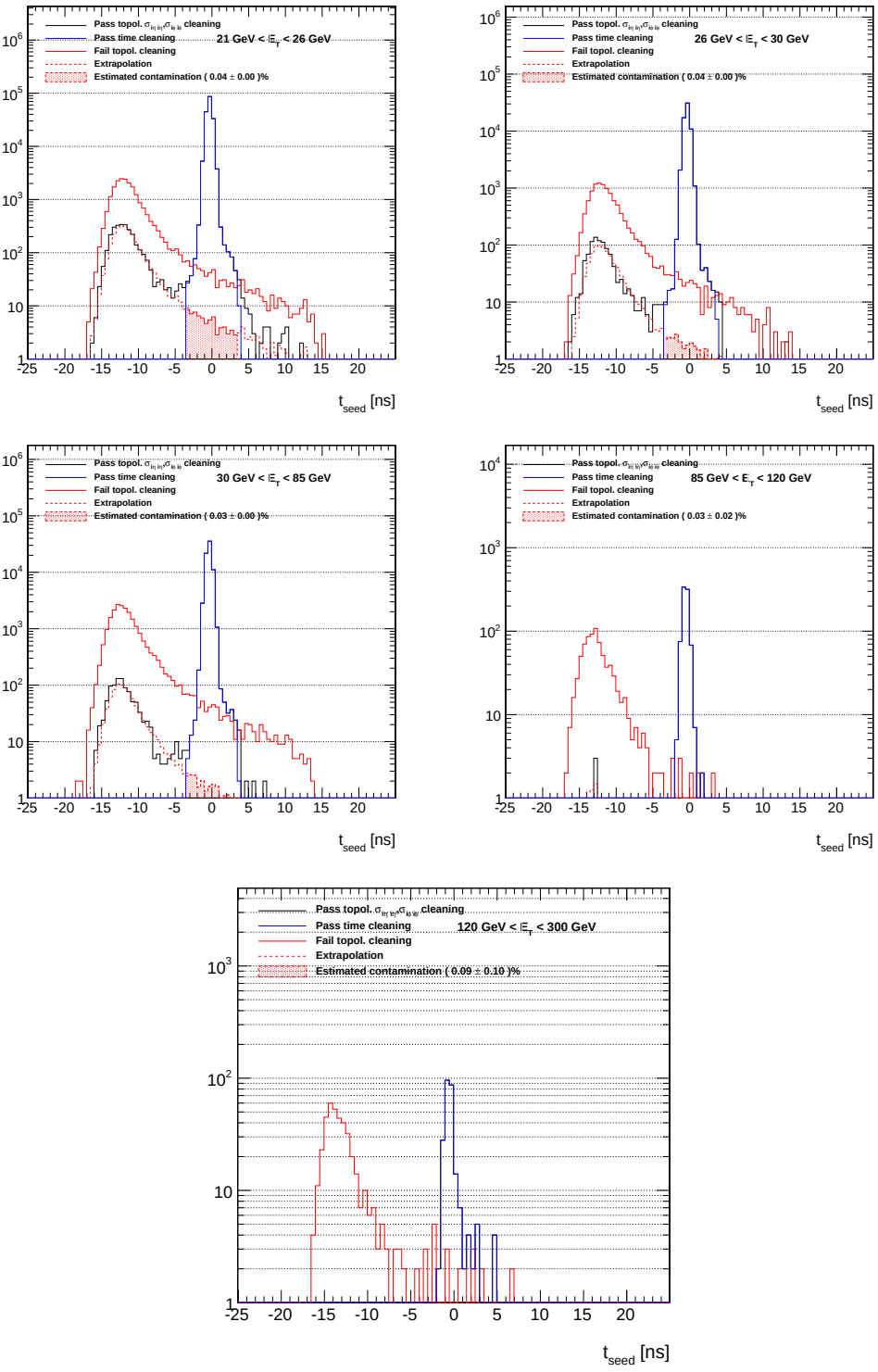


Figure 7.9: Illustration of the procedure to estimate the residual contamination from anomalous signals.

7.5 Cross section measurement

The differential photon production cross section was measured according to the following expression:

$$\frac{d^2\sigma}{dE_{\mathrm{T}}d\eta} = \frac{N_{\mathcal{S}}(\Delta E_{\mathrm{T}})}{\Delta E_{\mathrm{T}}\Delta\eta \cdot L(\Delta E_{\mathrm{T}}) \cdot \mathcal{U} \cdot \varepsilon(\Delta E_{\mathrm{T}})} \tag{7.2}$$

here $N_{\mathcal{S}}(\Delta E_{\mathrm{T}})$ is the number of prompt isolated photons, fulfilling a generator-level isolation criterion, observed in a given ΔE_{T} bin after all selection criteria are applied, $\varepsilon(\Delta E_{\mathrm{T}})$ is the reconstruction and selection efficiency, $\mathcal{U}$ the correction accounting for the smearing in the reconstructed energy, position and isolation quantities, and $L(\Delta E_{\mathrm{T}})$ the integrated luminosity corresponding to the analysed dataset.

The generator-level isolation criterion used to extract the isolated cross section was chosen in order to match the analysis level criteria as closely as possible. In particular, the isolated cross section was defined by requiring the isolation energy in a cone of $\mathcal{R} = 0.4$ around the photon to be below 5 GeV. As seen in Chapter 2, modern NLO generators are capable of implementing such a requirement in the calculation.

The efficiency and smearing corrections were obtained from data whenever possible. The details on the extraction of the two quantities follows.

7.5.1 Efficiency

The efficiency term in (7.2) can be factorised into several contributions, corresponding to the different steps of the selection process. In particular, it can be conveniently expressed as:

$$\varepsilon = \varepsilon_{RECO} \times \varepsilon_{Trig} \times \varepsilon_{ID} \tag{7.3}$$

where:

- ε_{RECO} represents the probability for a signal photon produced inside the detector geometrical acceptance to be reconstructed by the clustering algorithms;
- ε_{Trig} is the probability for a reconstructed signal photon to be selected by the trigger system;
- ε_{ID} is the probability for a reconstructed signal photon to pass the photon identification criteria described in Section 7.3.

All efficiencies were defined with respect to prompt isolated photons fulfilling the generator level isolation criteria defined above ($E_T^{max} = 5$ GeV in a cone $\mathcal{R} = 0.4$).
The reconstruction efficiency ε_{RECO} was obtained from MC simulations and found to be $> 98.8\%$.
The trigger efficiency was measured to be consistent with 100% in data (and MC) in the considered kinematic regions.
The term ε_{ID} was also measured from data. Given the limited size of the control samples available for this purpose, the E_{T} and η dependencies of the efficiency were modelled using MC simulation, while a global correction factor was extracted from data. Such factor is labelled as ρ_{ID} in the following. The photon identification efficiency ε_{ID} was then evaluated as:

$$\varepsilon_{ID}(E_{\mathrm{T}}, \eta) = \varepsilon_{ID}^{MC}(E_{\mathrm{T}}, \eta) \cdot \rho_{ID} \tag{7.4}$$

The photon identification efficiency in MC simulation, ϵ_{ID}^{MC}, is shown in Figure 7.10. The average efficiency for the mixture of direct and fragmentation photons predicted by PYTHIA is $89.9 \pm 0.2\%$. The variation in ϵ_{ID}^{MC} obtained by increasing and decreasing the fragmentation photons contribution by a factor of 2 is $\pm 0.25\%$, while the total spread over the region $|\eta| < 1.45$ is $\pm 0.5\%$.
In order to obtain the correction ρ_{ID} to the identification efficiency, the term was factorised in two parts. The first contains the isolation requirements and the cut on the H/E variable and is labelled as ε_{ISO}.

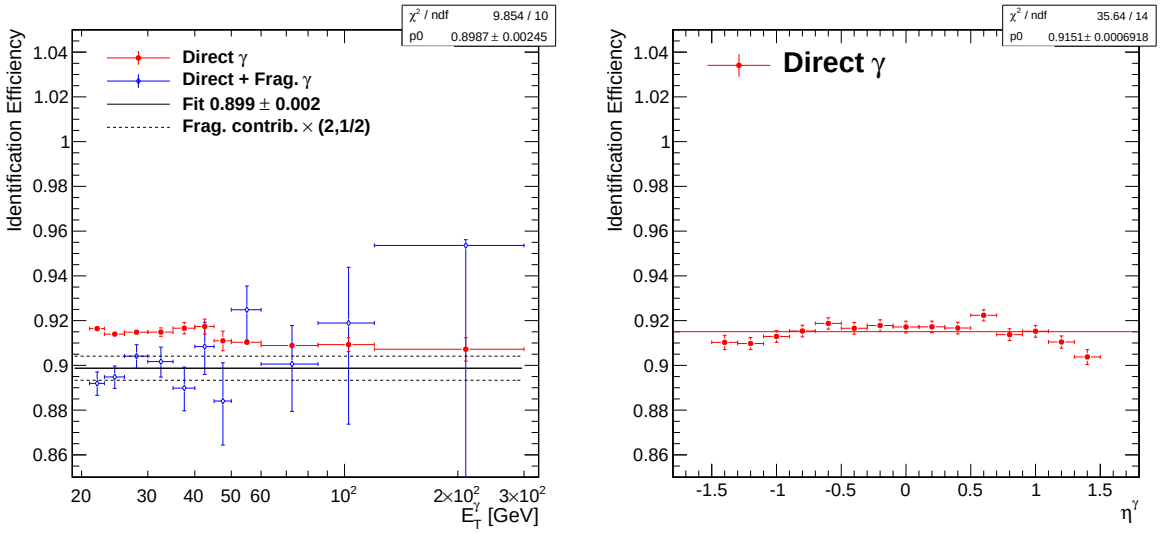


Figure 7.10: Efficiency of the photon identification criteria as a function of E_{T} and η.

The second contains the efficiency of the pixel seed veto criterion and is labelled ε_{Pix}. The two terms are largely uncorrelated, justifying the separate treatment.

The ε_{ISO} term was studied in data using an isolated electrons control sample, obtained from $Z \to ee$ decays. The average isolation efficiency for electrons from Z in MC was estimated to be $89.1 \pm 0.1\%$. The same quantity evaluated on a sample of direct photons (without applying any generator-level isolation criterion) in a similar kinematic range is $91.2 \pm 0.1\%$. The appropriateness of the comparison of the two quantities is guaranteed by the definition of the isolation cones, which reduces the sensitivity of the algorithms to differences between electrons and photons. This justifies the use of electron control samples to study photon isolation efficiencies.

Events with a Z decaying to electrons were studied using the tag and probe technique. A tight selection was imposed on one of the two electrons (the *tag*). Candidate Z events were selected using the invariant mass of the tag and a photon candidate (the *probe*), requiring it to be 60 GeV $< M_{ee} <$ 120 GeV. In this way a pure and unbiased sample of prompt electrons could be selected and used to measure the photon identification efficiency. The purity of the selected sample was estimated to be $> 99.2\%$ in MC simulations, as shown in Figure 7.11 (left).
The residual background contribution was subtracted using the yields predicted by the MC and a systematic uncertainty equal to the total predicted background was assigned to the efficiency measurement. The comparison between ε_{ISO} in data and MC in $Z \to ee$ events is shown in Figure 7.11 (right). A correction factor $\rho_{ISO} = 1.035 \pm 0.009(stat) \pm 0.014(syst)$ was estimated.

The term ε_{Pix} cannot be studied using electron control samples. The efficiency of this selection in MC is $\varepsilon_{Pix}^{MC} = (97.00 \pm 0.01)\%$ and is identical for direct and fragmentation photons. The variation over the E_{T} range considered in the measurement is $^{+0}_{-1}\%$, while the variation with η is $^{+0.5}_{-1.5}\%$.
Two effects contribute to the inefficiency of this criterion: photons converting before or in the first pixel layer are associated with hits in the pixel detector and thus get rejected; also random coincidences between photon and tracks from the interaction region may cause some prompt photons to be rejected. In MC simulations, the first effect accounts for most of the inefficiency and is labelled $\varepsilon_{ConvPix}$, while the second, labelled ε_{RndPix} accounts for a $\sim 0.5\%$ inefficiency.

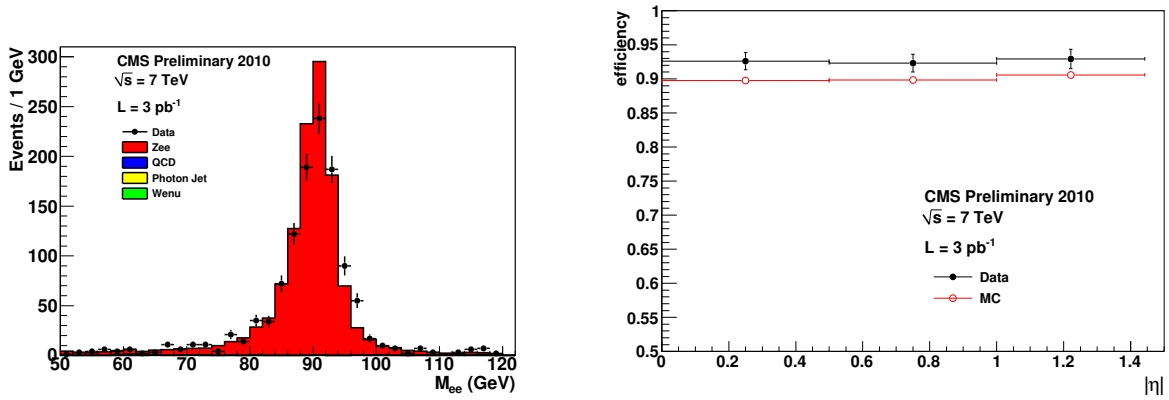


Figure 7.11: (left) Distribution of invariant mass of the tag-probe pairs used. (right) Comparison of ϵ_{ISO} in data and MC for the selected probes.

Table 7.6: Summary of efficiency contributions together with corrections estimated from data.

	MC value (%)	Correction from data (ρ)
ε_{RECO}	98.9	-
ε_{ISO}	92.7	1.035 ± 0.017
$\varepsilon_{ConvPix}$	97.5	-
ε_{RndPix}	99.5	$1 \pm (2\%/N_{vtx})$
Total ε	88.5	$1.035 \pm 0.017 \pm (2\%/N_{vtx})$

Due to the absence of a pure sample of prompt photons in data, the inefficiency due to early converting photons could only be estimated from MC simulation. The MC prediction for this quantity can be however considered reliable as the beam pipe and the pixel first layer material budget were verified to be accurately modelled, as discussed in [63].

The contribution due to random coincidences between photons and other tracks was studied in data. In order to do so, a sample of reconstructed conversions was selected. The same isolation criteria used in the analysis were applied and the conversion vertex fit was required to have good quality; the reconstructed conversion radius was required to be $r > 20$ cm in order to reject prompt photons converting in the pixel detector. The purity of the selected sample was estimated, with a two-component fit of the E/p distribution [98], to be $65 \pm 2\%$ for events with one reconstructed vertex and was seen to not vary with increasing number of primary vertexes.

To study the effect of pile-up, the pixel seed veto inefficiency was extracted as a function of the number of primary vertexes. The study was performed in two regions, one ($0.8 < E/p < 1.2$) with higher purity and one ($0 < E/p < 3$) with lower purity.
The MC simulation predicts that the measured inefficiency is dominated by background in both the higher and lower purity regions. Thus, the quantity taken into account in the study was the inefficiency increase per reconstructed primary vertex. The latter was estimated to be to be $1.74\% \pm 0.4\%$ in the higher purity region and $2.1\% \pm 0.5\%$ in the lower one. No correction was extracted from the study, but the measured inefficiency increase per reconstructed vertex was taken as an upper limit to the pile-up contribution to the pixel seed inefficiency and was assigned as a systematic uncertainty.

All terms contributing to the efficiency and corresponding corrections are summarised in Table 7.6. The total average efficiency used in calculating the cross-section is $\varepsilon = 91.6\%$.

7.5.2 Smearing corrections

The measured cross section was corrected in order to account for finite detector resolution, in terms of position, energy and isolation measurements. As shown below, this type of correction is very small, due to the high resolution in the measurement of the photon energy and position. Therefore, a simple correction, as opposed to more sophisticated unfolding techniques, can be justified.

As discussed in the following, the applied smearing correction is principally determined by losses in the reconstructed photon energy in the case of pair conversion. Since the description of the inner detector material budget was determined to be correct within 10% [99], the derived correction can be considered accurate enough for the precision attained by the measurement.

The correction was obtained from PYTHIA MC simulations and is represented by the $\mathcal{U}$ term in Equation (7.2). The $\mathcal{U}$ term is defined as the ratio between the number of prompt photons fulfilling the analysis-level criteria to those fulfilling the generator-level criteria for each E_{T} bin:

$$\mathcal{U} \cdot \varepsilon_{ID} = \frac{N^{\gamma}(E_{\mathrm{T}}^{RECO}, |\eta_{SC}| < 1.4442, ID^{RECO})}{N^{\gamma}(E_{\mathrm{T}}^{GEN}, |\eta^{GEN}| < 1.45, Iso^{GEN})}$$

where $N^{\gamma}(E_{\mathrm{T}}^{RECO}, |\eta_{SC}| < 1.4442, ID^{RECO})$ is the number of prompt photons fulfilling all the selection criteria outlined in Section 7.3, while $N^{\gamma}(E_{\mathrm{T}}^{GEN}, |\eta^{GEN}| < 1.45, Iso^{GEN})$ is the number of prompt photons fulfilling the generator level isolation criteria ($E_{\mathrm{T}}^{had} < 5\,\mathrm{GeV}$), i.e. the signal. The term ε_{ID} was defined in the previous section.
The smearing correction $\mathcal{U}$ obtained for a sample of direct photons is shown in Figure 7.12 (left). For comparison, the term ε_{ID} is also shown: it can be seen that the smearing correction is $\mathcal{U}$ considerably smaller than the efficiency correction.

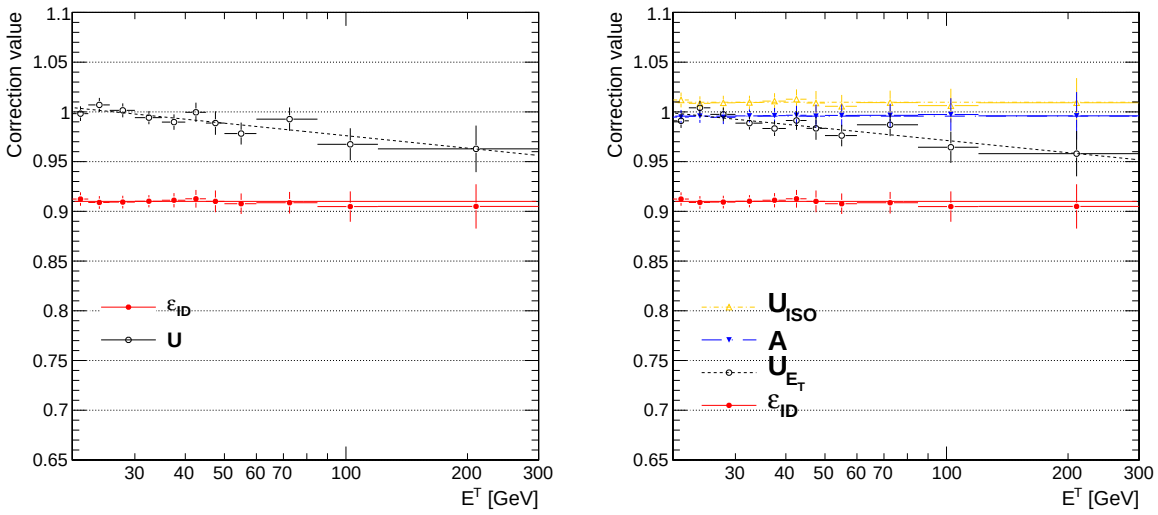


Figure 7.12: (*left*) Smearing correction term $\mathcal{U}$ as a function of E_{T}. For comparison, the photon identification efficiency correction ε_{ID} is also shown. (*right*) Different contributions to the smearing correction term.

In order to study the contribution of the different effects, $\mathcal{U}$ was factorised into three components: one containing the effect of the E_{T} smearing (U_{E_T}), one accounting for acceptance corrections due to smearing

in the η measurement (A) and a third accounting for smearing in the isolation quantities (U_{ISO}). They are symbolically defined as:

$$U_{E_T} = \frac{N^{\gamma}(E_{\mathrm{T}}^{\mathbf{RECO}}, \eta^{RECO}, Iso^{GEN})}{N^{\gamma}(E_{\mathrm{T}}^{\mathbf{GEN}}, \eta^{RECO}, Iso^{GEN})}$$

$$A = \frac{N^{\gamma}(E_{\mathrm{T}}^{RECO}, \eta^{\mathbf{RECO}}, Iso^{GEN})}{N^{\gamma}(E_{\mathrm{T}}^{RECO}, \eta^{\mathbf{GEN}}, Iso^{GEN})}$$

$$U_{ISO} \cdot \varepsilon_{ID} = \frac{N^{\gamma}(E_{\mathrm{T}}^{RECO}, \eta^{RECO}, ID^{\mathbf{RECO}})}{N^{\gamma}(E_{\mathrm{T}}^{RECO}, \eta^{RECO}, Iso^{\mathbf{GEN}})}$$

where the superscripts $RECO$ and GEN indicate whether a given selection criteria or variable is considered at reconstruction or generator level respectively.
The contribution of the three terms is shown in Figure 7.12 (right). It can be seen that the largest effect is due to the smearing in the energy reconstruction, which determines the observed 4% variation as a function of the photon E_{T}.

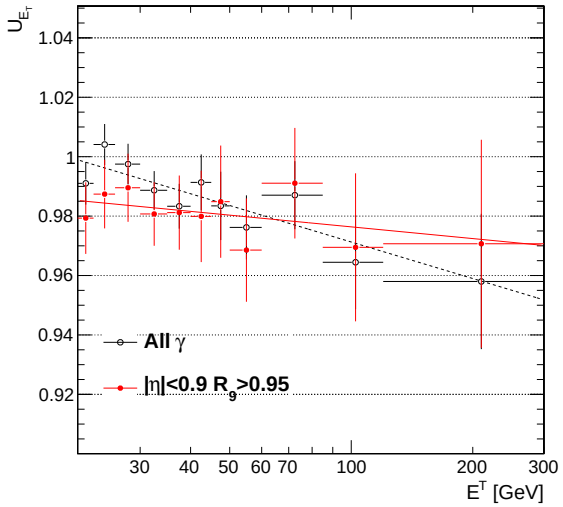


Figure 7.13: Comparison of the energy smearing correction for all photons and for a sample enriched in unconverted photons, obtained selecting candidates in the central barrel and requiring the SC to be narrow. It can be seen that in the latter case the observed variation as a function of E_{T} is greatly reduced.

The observed behaviour is mostly determined by imperfections in the super cluster energy corrections (discussed in Chapter 6) applied in order to recover energy losses due to pair conversion. Selecting photon candidates in the central barrel ($|\eta| < 0.9$) and enriching the sample in unconverted photons (requiring $R_9 > 0.95$) reduces the observed U_{E_T} variation to $\sim 1\%$ over the considered E_{T} range, as shown in Figure 7.13.

The terms U_{E_T} and A are common to direct and fragmentation photons, but this is not the case for the U_{ISO} term. Thus the U_{ISO} term was obtained separately for the two populations and the total smearing correction calculated as:

$$U = \frac{U^{dir}}{U^{dir}_{ISO}} \cdot \left(U^{dir}_{ISO} \cdot f^{dir} + U^{frag}_{ISO} \cdot f^{frag} \right)$$

here the superscripts dir and $frag$ refer to direct and fragmentation photons respectively, while f^{dir} and f^{frag} represent the fraction of direct and fragmentation photons in the isolated photons population.
In order to perform the linear combination above, the terms U^{dir}, U^{dir}_{ISO} and U^{frag}_{ISO} were fit to constant or to linear functions of $\log(E_{\mathrm{T}})$, obtaining the following results:

$$
\begin{aligned}
U^{dir}_{ISO} &= 1.0097 \pm 0.0003 \\
U^{frag}_{ISO} &= 1.037 \pm 0.006 \\
U^{dir} &= 1.0034 \pm 0.005 - (0.018 \pm 0.006) \cdot \log\left(\frac{E_{\mathrm{T}}}{22\,\mathrm{GeV}}\right)
\end{aligned}
$$

The bin-by-bin correction factors applied are reported in Table 7.7. It can be seen that the applied correction ranges between 1.01 and 0.97. A systematic uncertainty was derived changing the ratio f^{dir}/f^{frag} to twice and half of the value predicted by PYTHIA, obtaining a variation ≤ 0.004.

Table 7.7: Value of the applied smearing correction factor $\mathcal{U}$ as a function of the photon E_{T}.

E_{T} (GeV)	$\mathcal{U}$	E_{T} (GeV)	$\mathcal{U}$
21–23	1.011±0.007(stat)±0.004(syst)	45–50	1.000±0.005(stat)±0.004(syst)
23–26	1.011±0.007(stat)±0.004(syst)	50–60	0.997±0.005(stat)±0.004(syst)
26–30	1.007±0.006(stat)±0.004(syst)	60–85	0.992±0.004(stat)±0.003(syst)
30–35	1.004±0.006(stat)±0.004(syst)	85–120	0.984±0.004(stat)±0.004(syst)
35–40	0.999±0.005(stat)±0.004(syst)	120–300	0.971±0.004(stat)±0.001(syst)
40–45	0.995±0.005(stat)±0.004(syst)		

7.5.3 Result

The isolated photon cross section was obtained applying Equation (7.2). The number of signal events in each E_{T} bin was extracted from the two component fit and is reported in Table 7.4; the efficiency and smearing corrections were discussed above: the corrected average efficiency is reported in Table 7.6, while the smearing corrections are in Table 7.7; finally, the integrated luminosity used for each bin is reported in Table 7.1.
Combining these terms, the isolated photon cross production cross section was measured. Figure 7.14 shows the result of the procedure.

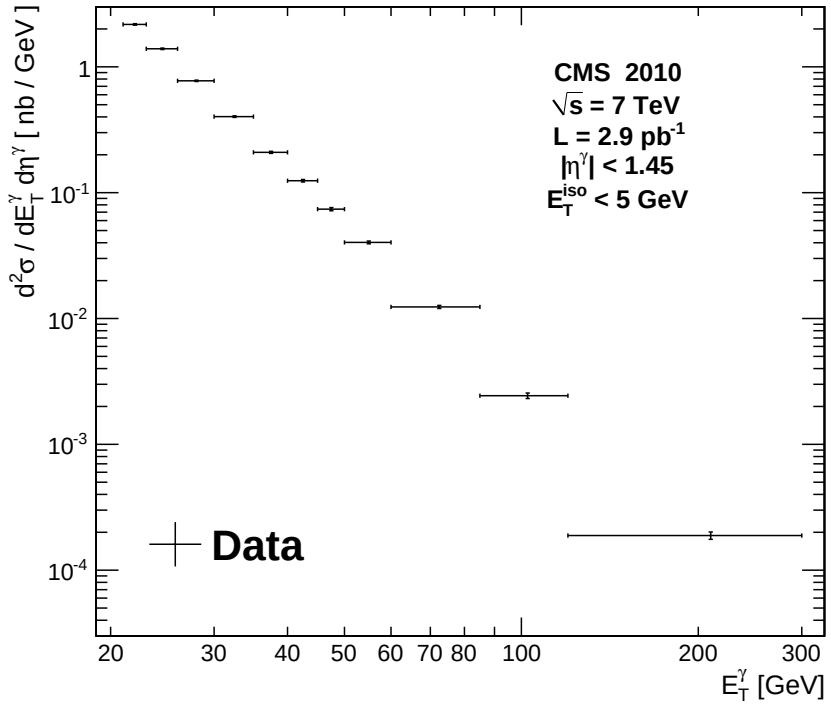


Figure 7.14: Measured isolated photon differential cross section. Only statistical errors are shown.

7.6 Systematic uncertainties

Several studies were performed to assess in which way quantities entering the cross-section determination can systematically affect the results.
In the following, all systematic uncertainties are expressed as variations with respect to the nominal values, $\delta = N_{\mathcal{S}\prime}/N_{\mathcal{S}} - 1$, where $N_{\mathcal{S}\prime}$ is the perturbed signal yield and $N_{\mathcal{S}}$ the nominal signal yield. Effects estimated to be smaller than 0.1% are not reported in the corresponding table.

7.6.1 Component shape fit

The signal yield was primarily extracted from the two-component fit of the $\sigma_{i\eta i\eta}$ distribution. For this reason an extensive investigation was performed to check all the quantities used in the procedure.

Signal Shape

As discussed in Section 7.4.1, the signal component shapes were corrected shifting the distributions by the difference in $\sigma_{i\eta i\eta}$ between electrons from Z decays in data and MC, 8×10^{-5}, letting the shift free to float bin-by-bin within the correction uncertainty.
A systematic uncertainty on the cross section was derived propagating the uncertainty on the $\sigma_{i\eta i\eta}$ correction. In order to do so, the shift of the signal component shape was varied in the range $5 \div 11 \times 10^{-5}$. The two component fits were repeated for 5 different working points within that interval and the maximum variations in the extracted signal yield taken as systematic uncertainty. The results are shown in Table 7.8. The effect was observed to be $\leq 3\%$.

Table 7.8: Effect of changing the signal component shape on the extracted signal yield.

Photon E_{T} (GeV)	Positive δ	Negative δ	Photon E_{T} (GeV)	Positive δ	Negative δ
21–23	+2.2%	-1.3%	45–50	+1.1%	-0.3%
23–26	+0.6%	-2.3%	50–60	+0.2%	-1.2%
26–30	-	-3.1%	60–85	-	-2.4%
30–35	+1.8%	-1.1%	85–120	0.4%	-1.0%
35–40	-	-3.2%	120–300	+0.1%	-0.1%
40–45	+0.6%	-2.6%			

Background Shape

The choice of the background component shape, namely the $\sigma_{i\eta i\eta}$ distribution for data in a $\mathrm{Iso}_{\mathrm{TRK}}$ side-band of $2 < \mathrm{Iso}_{\mathrm{TRK}} < 5\,\mathrm{GeV}/c$, may systematically affect the results.
In order to estimate the effect, a series of pseudo-experiments were performed, to compare the background model obtained from the $\mathrm{Iso}_{\mathrm{TRK}}$ side-band in MC (MC side-band in Figure 7.5) to the true $\sigma_{i\eta i\eta}$ distribution of neutral-meson background in the signal region (MC background in Figure 7.5).
In order to perform these studies, it was first verified that the correlation between $\sigma_{i\eta i\eta}$ and $\mathrm{Iso}_{\mathrm{TRK}}$ is well modelled by MC. Figure 7.15 shows how the average of the $\sigma_{i\eta i\eta}$ distributions evolves with $\mathrm{Iso}_{\mathrm{TRK}}$ for both data and MC. It can be seen that the MC reproduces very well the data up to large values of $\mathrm{Iso}_{\mathrm{TRK}}$. This shows, in particular, that for the chosen side-band ($2\ \mathrm{GeV}/c < \mathrm{Iso}_{\mathrm{TRK}} < 5\ \mathrm{GeV}/c$), the MC can be used to study the systematic uncertainty associated with the knowledge of the background shape.
The natural choice for the estimation of the systematic uncertainty would be to compare the true background from the di-jet samples to the background model extracted from a realistic side-band obtained mixing the di-jet and photon-jet samples.
As noted before, the size of the di-jet MC samples is limited. In particular, it did not allow to obtain smooth $\sigma_{i\eta i\eta}$ distributions of the MC true background for all the E_{T} bins used in the measurement.

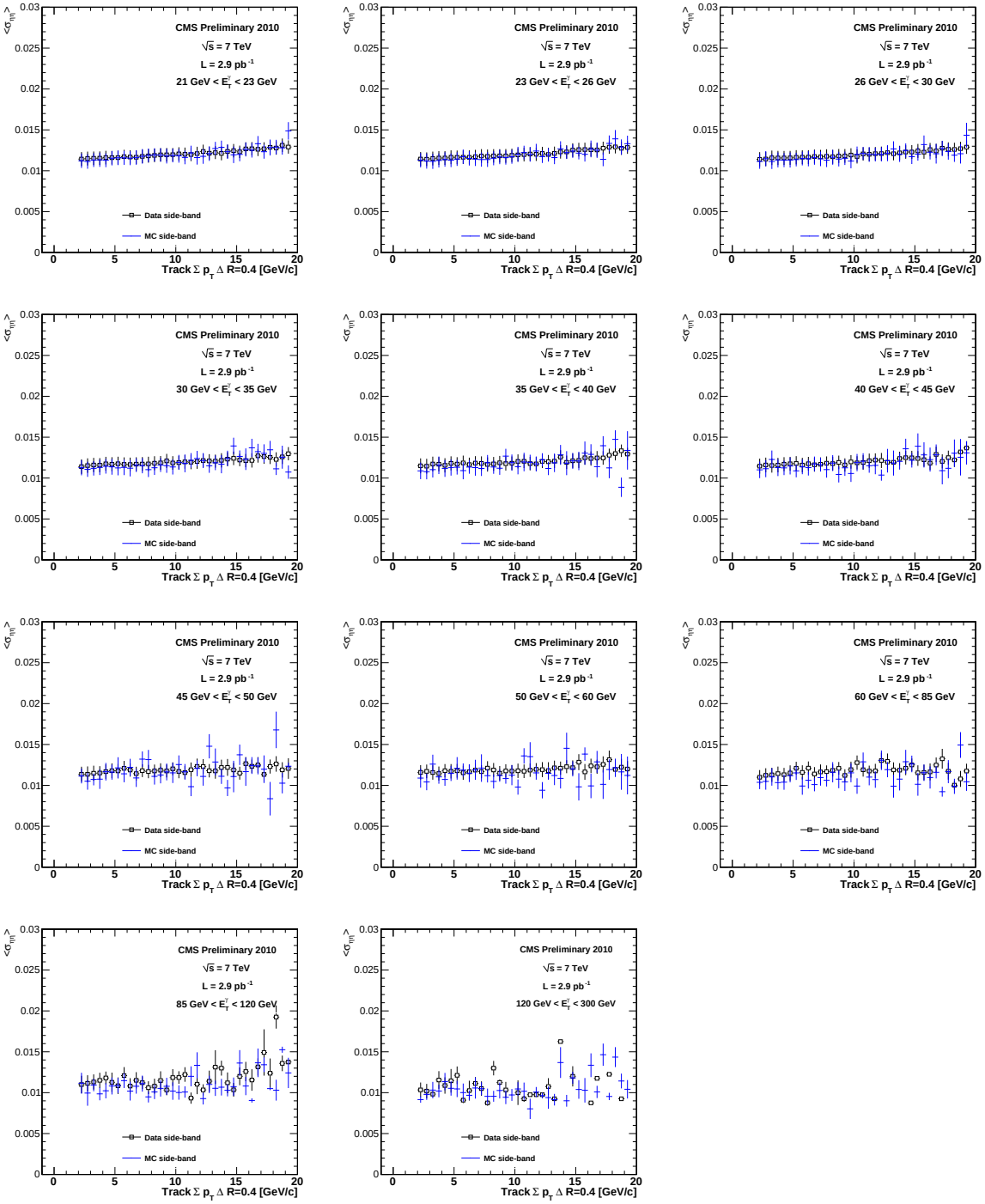


Figure 7.15: Average $\sigma_{i\eta i\eta}$ as a function of $\mathrm{Iso_{TRK}}$ for data (filled black) and MC (hollow red) in the $\mathrm{Iso_{TRK}}$ sideband. The $\sigma_{i\eta i\eta}$ variable is labelled $\sigma_{\eta\eta}$ in the plots.

.

In order to extract meaningful results, the comparison was performed in E_{T} sliding windows for $E_{\mathrm{T}} < 85\,\mathrm{GeV}$. For $E_{\mathrm{T}} > 85\,\mathrm{GeV}$, the samples with $\hat{p}_{\mathrm{T}} > 80\,\mathrm{GeV}$ provide adequate statistical power. The bins in E_{T} used for this study were 21–30, 23–35, 26–40, 30–45, 35–60, 40–85, 45–85, 50–85, 60–85, 85–120 and 120–300 GeV.
The effect on the extracted signal yield was studied. To do so, the pseudo-experiments were designed to reproduce as closely as possible the actual procedure with data.
Pseudo-data were generated in the following way:

- For signal, the shapes used to extract the yields were sampled $N_{\mathcal{S}}$ times, allowing for Poisson fluctuations.
- For background, the true background shapes obtained from the di-jet samples were sampled $N_{\mathcal{B}}$ times, allowing for Poisson fluctuations.

Then, the two component fits to the pseudo experiments were performed using the following shapes:

- For signal, the same shapes used for sampling.
- For background, the side-band background model from a realistic mix of di-jet and photon-jet samples.

The values of $N_{\mathcal{S}}$ and $N_{\mathcal{B}}$ used in sampling were those extracted from the fits to data. The pseudo-experiments were repeated 300 times in each E_{T} bin to obtain a distribution of $\delta = N'_{\mathcal{S}}/N_{\mathcal{S}} - 1$, the relative amount by which the nominal signal yield ($N_{\mathcal{S}}$) is affected. The result of the procedure is shown as black crosses in Figure 7.16. It can be seen that the systematic effect thus estimated is $< 10\%$.
Upon closer inspection, one can see that there are two competing effects at work in this procedure. The use of the side-band biases the background distribution, as it selects events with slightly different characteristics in terms of multiplicity and fraction of converted candidates. This leads to depression of the peaking part of the background shape and thus an overestimation of the signal.
On the other hand, the presence of non-isolated prompt photons in the side-band region leads to an enhancement of the peaking part of the distributions and thus to an underestimation of the signal.
The comparison was therefore broken down into two steps to study the individual effects separately. Two other sets of pseudo-experiments were produced, changing the background model used in the generation and in the fit.

1. Side-band bias for the background, red points in Figure 7.16:
   - Sample from the **true background** from di-jet samples.
   - Fit using the **background-only side-band** distribution, obtained by excluding non-isolated prompt photons from the sample.
2. Signal contamination in the side-band, blue points in Figure 7.16:
   - Sample from the **background-only side-band** (last step above).
   - Fit using the background model from the **realistic side-band**, obtained by mixing background and non-isolated prompt photon contributions.

In Figure 7.16 it can be seen that the effects are indeed of opposite sign and similar magnitude. Since the degree to which this cancellation really happens depends on the details of the MC simulation, a conservative approach was taken. Given the magnitude of the individual effects, the systematic uncertainty assigned to the knowledge of the background component shape, takes them into account independently. The estimated uncertainty is represented by the solid lines in Figure 7.16 and shown in Table 7.9. An uncertainty of $^{+12.5\%}_{-15.0\%}$ ($^{+12.5\%}_{-6.5\%}$) was assigned for E_{T}^{γ} above (below) 85 GeV.

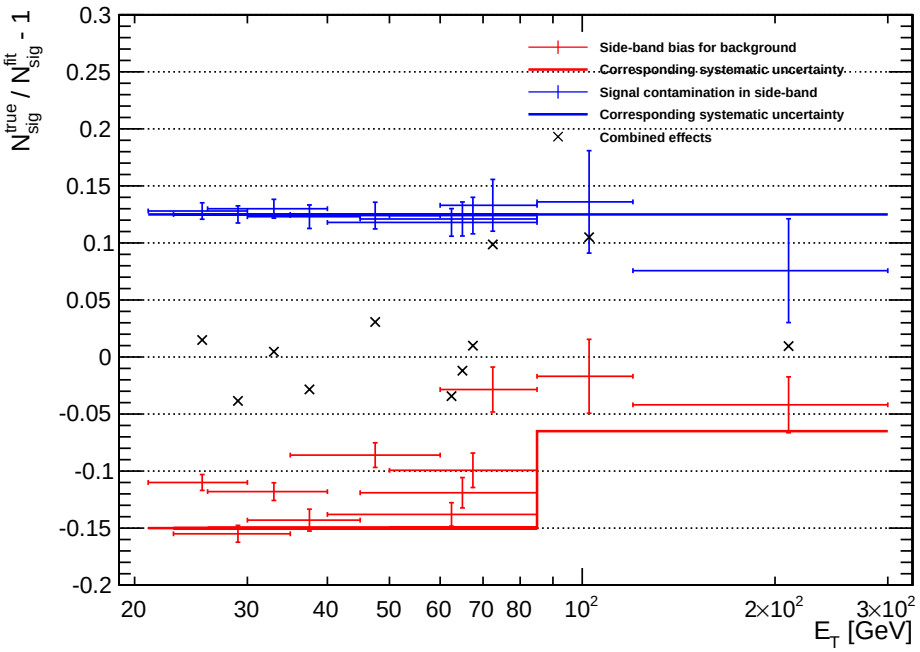


Figure 7.16: Relative effect on the number of signal events due to the difference between the true background shape and the one coming from the $\mathrm{Iso}_{\mathrm{TRK}}$ side-band. By themselves, the side-band bias (red) and the contribution of signal to the side-band (blue) are sizable. Since the effects have different signs, they tend to cancel each other out when taken together (black).

Table 7.9: Estimated systematic uncertainty on the extracted signal yield due to the background component shape.

Photon E_{T} (GeV)	Positive δ	Negative δ	Photon E_{T} (GeV)	Positive δ	Negative δ
21–23	+12.5%	-15.0%	45–50	+12.5%	-15.0%
23–26	+12.5%	-15.0%	50–60	+12.5%	-15.0%
26–30	+12.5%	-15.0%	60–85	+12.5%	-15.0%
30–35	+12.5%	-15.0%	85–120	+12.5%	-6.5%
35–40	+12.5%	-15.0%	120–300	+12.5%	-6.5%
40–45	+12.5%	-15.0%			

Data-driven verification of background shape systematic uncertainty

In order to verify whether the background shape effects are correctly modelled by the MC simulation, the $\mathrm{Iso}_{\mathrm{TRK}}$ side-band composition was tested directly in data. To do so, the definition of the side-band was varied in such way to enhance each of the two effects affecting the background shape and the two-component fit was repeated with the modified background model.
In particular, the size of the side-band window was varied as follows:

- In order to test the effect of the intrinsic differences between side-band and true background, the lower limit was raised in such a way as to remove the isolated photons contribution. Two working points were tested: $3 < \mathrm{Iso}_{\mathrm{TRK}} < 5\,\mathrm{GeV}/c$ and $4 < \mathrm{Iso}_{\mathrm{TRK}} < 7\,\mathrm{GeV}/c$.

- In order to test the effect of the signal contamination, the upper limit was decreased in order to enrich the side-band in prompt photons. The tested working points were $2 < \mathrm{Iso}_{\mathrm{TRK}} < 4\,\mathrm{GeV}/c$ and $2 < \mathrm{Iso}_{\mathrm{TRK}} < 3\,\mathrm{GeV}/c$.

The signal yields estimated using the alternative working points were compared to the nominal fits. The relative difference between the two is shown in Table 7.10.
It can bee seen that the sign of the variations matches the prediction from MC: enriching the side-band in isolated photons leads to an underestimation of the signal yield, while accepting less isolated background candidates leads to an overestimation.
The magnitude of the observed variations are also in good agreement with the predictions from MC, thus corroborating the treatment of the background shape systematic uncertainty.

Table 7.10: Verification of background systematic uncertainty in data. The relative difference between signal yield fitted using the nominal and modified side-band definition is shown as a function of E_{T}. Sign and magnitude of the variations match well the MC predictions.

	$\mathrm{Iso}_{\mathrm{TRK}}$ side-band limits (GeV)			
	2–4	2–3	3–5	4–7
Photon E_{T} (GeV)	Relative difference in fitted signal yield			
21–23	-2.39%	-7.56%	+5.84%	+14.44%
23–26	-9.42%	-4.56%	+5.67%	+14.20%
26–30	-5.79%	-2.27%	+4.63%	+11.00%
30–35	-7.10%	-2.48%	+5.14%	+8.91%
35–40	-5.11%	-2.82%	+4.59%	+6.92%
40–45	-5.86%	-2.67%	+3.37%	+5.60%
45–50	-7.80%	-2.69%	+4.74%	+10.29%
50–60	-2.13%	-1.31%	+0.45%	+3.63%
60–85	-5.40%	-3.51%	+3.64%	+5.37%
85–120	-0.71%	+2.50%	+1.72%	+2.79%
120–300	-1.43%	-1.76%	+1.62%	+4.48%

Pile-up effects

The effect of pile-up on $\sigma_{i\eta i\eta}$ was evaluated looking at the dependence of the $\sigma_{i\eta i\eta}$ distribution on the number of reconstructed primary vertexes.
To do so, events were classified according to the number of reconstructed vertexes passing the quality criteria listed in Section 7.2. In order to minimise the ambiguity in the determination of the event vertex multiplicity, events with one or more reconstructed vertexes failing the quality criteria were not considered in this study. This requirement rejected $\sim 10\%$ of the events.

The $\sigma_{i\eta i\eta}$ distribution was plotted separately for each multiplicity bin and for each photon E_T bin. An example is shown in Figure 7.17 (left). The mean value of the $\sigma_{i\eta i\eta}$ distribution as a function of the number of primary vertexes was then obtained. The resulting profile plot was fitted with a straight line for each E_T bin and the pile-up effect was evaluated taking the ratio of the line slope to the intercept calculated for $N_{VTX} = 1$. The result of the procedure is shown in Figure 7.17 (right), where the estimated effect can be seen to be negligible.

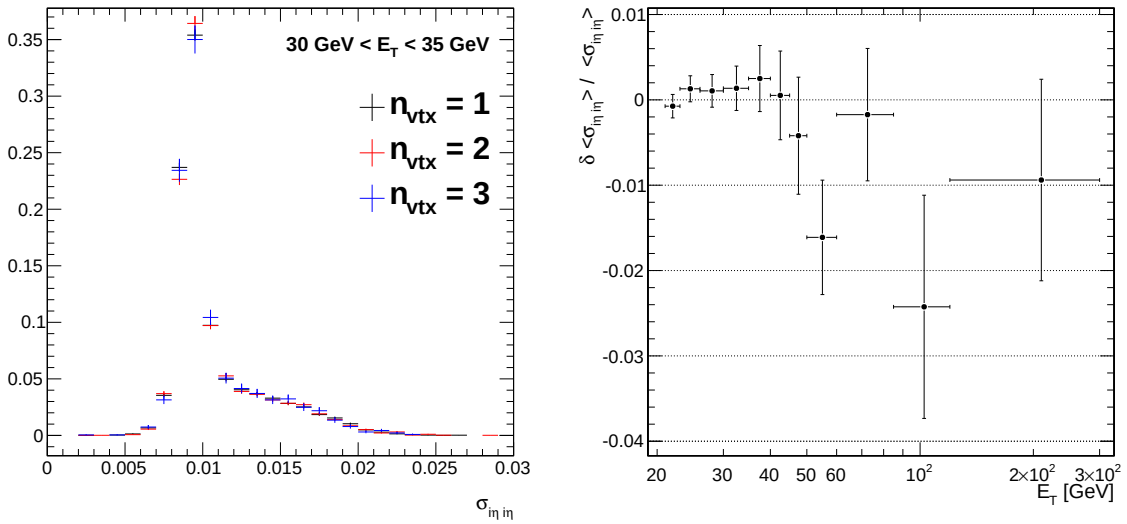


Figure 7.17: (*left*)$\sigma_{i\eta i\eta}$ distribution obtained in events with different multiplicities for photon candidates with E_T between 30 and 35 GeV. (*right*) Estimated fractional increase of the average $\sigma_{i\eta i\eta}$ per number of reconstructed vertexes as a function of the photon E_T. The $\sigma_{i\eta i\eta}$ variable is labelled $\sigma_{\eta\eta}$ in the plots.

Component shape statistics

The shapes used to model the signal and background components have a statistical uncertainty which, in some cases, can not be dismissed.
To estimate this effect, the component shapes were modified within their intrinsic statistical uncertainties. In practise, pseudo-experiments were performed where each bin in each shape was replaced by a number extracted, at random, from a Poisson distribution with mean equal to the contents of the bin. The two-component fit to data was then repeated with the perturbed shapes and the extracted yield ($N'_{\mathcal{S}}$) compared with the original yield ($N_{\mathcal{S}}$).
The procedure was performed 300 times in each E_T bin and the mean (μ) and RMS (s) of the $N'_{\mathcal{S}}/N_{\mathcal{S}}$ distribution (observed to be Gaussian) were calculated. An uncertainty covering the interval $\mu \pm s$ was used to estimate the systematic uncertainty due to this effect. The results are shown in Table 7.11. The effect was found to be $< 1\%$ in the lower part of the E_T spectrum and $\lesssim 2 - 4\%$ in the higher part.

Binning

The $\sigma_{i\eta i\eta}$ distributions were binned in equal bins of width 0.001. In order to understand the effect introduced by this choice, the fitting procedure was repeated varying the bin size to twice and half of the nominal one. Larger bins led to a higher signal yield ($N_{\mathcal{S}\prime} > N_{\mathcal{S}}$) while smaller bins had the opposite effect ($N_{\mathcal{S}\prime} < N_{\mathcal{S}}$). The results are reported in Table 7.12. The size of the effect was found to be $< 4\%$.

Table 7.11: Effect of varying the signal and background component shapes within their intrinsic statistical uncertainties on the extracted signal yield.

Photon E_{T} (GeV)	Positive δ	Negative δ	Photon E_{T} (GeV)	Positive δ	Negative δ
21–23	+0.8%	-0.5%	45–50	+1.0%	-1.3%
23–26	+0.6%	-0.8%	50–60	+1.1%	-1.4%
26–30	+0.7%	-0.7%	60–85	+1.8%	-1.8%
30–35	+0.7%	-0.9%	85–120	+2.7%	-3.6%
35–40	+0.8%	-1.0%	120–300	+1.5%	-2.0%
40–45	+0.8%	-1.0%			

Table 7.12: Effect of changing the binning of the $\sigma_{i\eta i\eta}$ distributions on the extracted signal yield.

Photon E_{T} (GeV)	Positive δ	Negative δ	Photon E_{T} (GeV)	Positive δ	Negative δ
21–23	+0.1%	-3.1%	45–50	+0.1%	-0.5%
23–26	+0.7%	-2.8%	50–60	+0.4%	-
26–30	-	-3.8%	60–85	+0.1%	-2.3%
30–35	+1.1%	-2.5%	85–12 0	-	-1.3%
35–40	-	-3.5%	120–300	-	-2.3%
40–45	-	-3.7%			

7.6.2 Efficiency and Smearing corrections

The studies performed to obtain the efficiency term in Equation 7.2 were described in detail in Section 7.5.1.

For the isolation efficiency the systematic uncertainty was derived from the total uncertainty on the ρ_{ISO} correction from data. For the pixel seed efficiency, a systematic uncertainty equal to the total predicted MC inefficiency was applied. An additional 2% uncertainty per each additional vertex was added in order to account for pile-up effects. Also, the effect due to the variation in the fragmentation photon contribution by a factor of 2 more (or less) was included. Finally, a systematic uncertainty equal to $1 - \varepsilon_{RECO}$ was assigned to the absolute reconstruction efficiency.
The total systematic uncertainty ascribed to the knowledge of the efficiency factors was $\pm 3.8\%$.

In order to estimate the uncertainty due to the smearing correction U, the total uncertainty in Table 7.7 was propagated to the cross section and an additional uncertainty equal to $1 - \mathcal{U}$ added.
The effect of the difference between predicted and simulated detector response to low energy particles [100] contributing to the isolation sums was studied by scaling the track isolation sum by $\pm 2\%$ and the calorimetric isolation sums by $\pm 5\%$ in the MC simulation. The effect on the smearing correction and identification efficiency was found to be negligible.
The total uncertainty associated to the smearing correction is reported in Table 7.13 and is $\leq 3\%$.
Figure 7.18 shows the value of the ε and U corrections applied in the cross section measurement together with their associated uncertainties.

Table 7.13: Uncertainty associated to the smearing correction.

Photon E_T (GeV)	Positive δ	Negative δ	Photon E_T (GeV)	Positive δ	Negative δ
21–23	+0.9%	-1.3%	45–50	+0.6%	-0.7%
23–26	+0.8%	-1.4%	50–60	+0.7%	-0.7%
26–30	+0.8%	-1.0%	60–85	+1.0%	-0.6%
30–35	+0.7%	-0.8%	85–120	+1.7%	-0.6%
35–40	+0.7%	-0.7%	120–300	+3.1%	-0.5%
40–45	+0.8%	-0.7%			

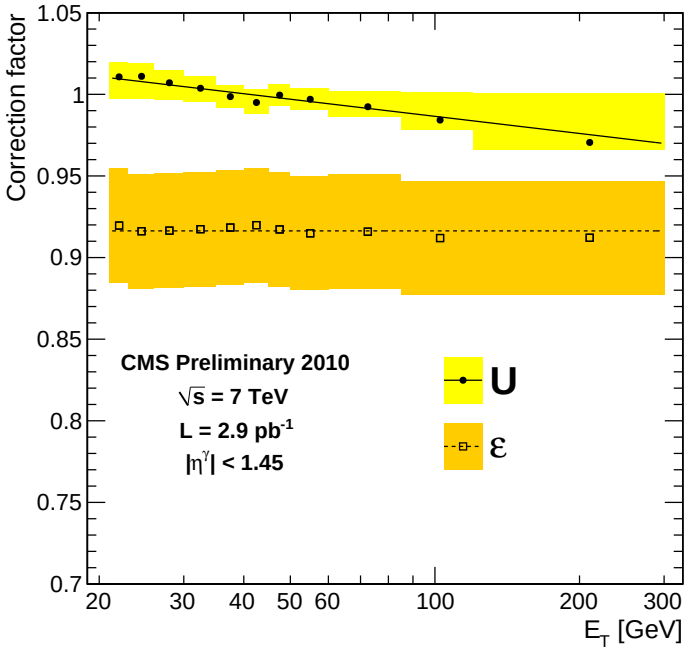


Figure 7.18: Efficiency and smearing corrections applied.

7.6.3 Photon energy scale

The uncertainty in the reconstructed photon E_{T} directly affects the extracted signal yield. The effect is mostly due to migration of events between bins, or falling out of the E_{T} range studied.
The energy scale was varied by ±0.68%, taking into account the energy scale correction uncertainty (±0.64%) and the differences between reconstructed and generated photons in photon-jet MC samples (±0.2%).
The latter effect is shown in Figure 7.19 (left), where the distribution of the generated to reconstructed photon transverse energy ($E_{\mathrm{T}}^{\mathrm{RECO}}/E_{\mathrm{T}}^{\mathrm{GEN}}$) is shown for a particular photon E_{T} bin. In order to asses the energy scale of photon candidates, the peak position was determined as a function of E_{T} through a Gaussian fit to the peaking part of the distribution.
Since the startup SC energy corrections were obtained using electrons, the difference between the fitted value and unity represents a measure of the difference between the electron and photon energy scales which was estimated to be ±0.2% (as shown in Figure 7.19, right).

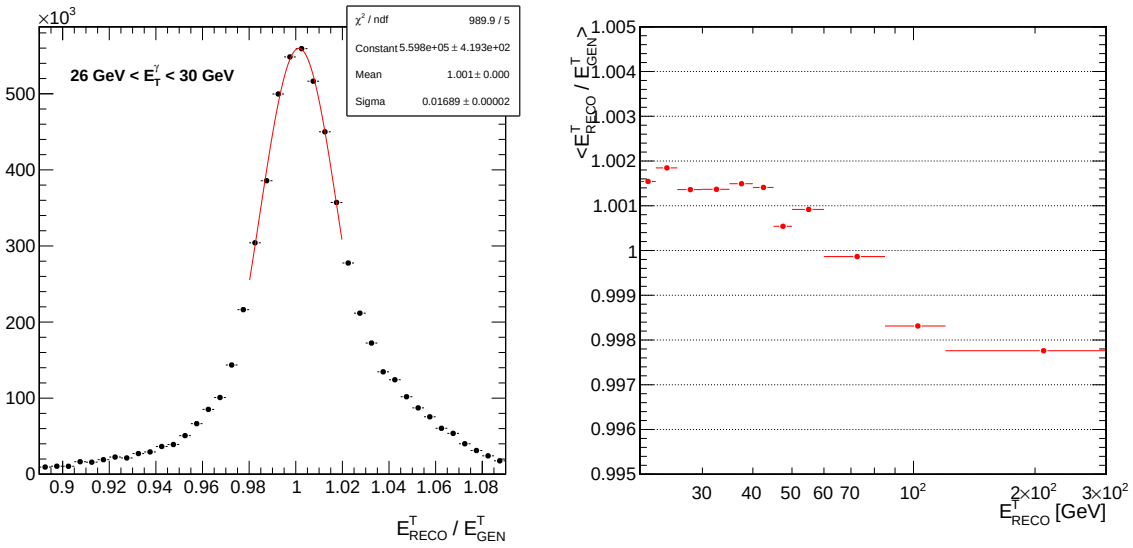


Figure 7.19: (left) Distribution of the ratio of reconstructed to generated photon transverse energy ration in a particular E_{T} bin. A gaussian fit to the peak region is shown. The non-gaussian tails are due to the imperfect recovery of radiative losses. (right) Position of the fitted gaussian mean as a function of the reconstructed E_{T}.

In order to evaluate the uncertainty due to the knowledge of the photon energy scale, the analysis was repeated shifting the latter by the full ±0.68% uncertainty. The resulting effect in the extracted yield is shown in Table 7.14. The magnitude of the estimated uncertainty is 1–3%.

Table 7.14: Effect of the photon energy scale uncertainty on the extracted signal yield.

Photon E_T (GeV)	Positive δ	Negative δ	Photon E_T (GeV)	Positive δ	Negative δ
21–23	+0.8%	-1.1%	45–50	+1.2%	-3.2%
23–26	+1.8%	-1.8%	50–60	+2.1%	-0.7%
26–30	+2.4%	-2.2%	60–85	+3.4%	-2.7%
30–35	+2.0%	-2.5%	85–120	+2.6%	-1.9%
35–40	+2.9%	-1.5%	120–300	+1.8%	-2.8%
40–45	+2.3%	-2.4%			

7.6.4 Summary of systematic uncertainties

Figure 7.20 summarises the estimated systematic uncertainties and, in black, shows the total systematic uncertainty. The latter was obtained by adding in quadrature (and separately) the positive and negative parts of each of the contributions.
The largest contribution to the systematic uncertainty on the cross-section measurement comes from the knowledge of the shape of the background component and is of the order of 12–15%. The contribution due to the knowledge of the efficiency correction was estimated to be 3.8% and is dominated by the uncertainty in the efficiency of requiring no hits in the pixel detector. All other uncertainties were estimated to be $\leq 3\%$.

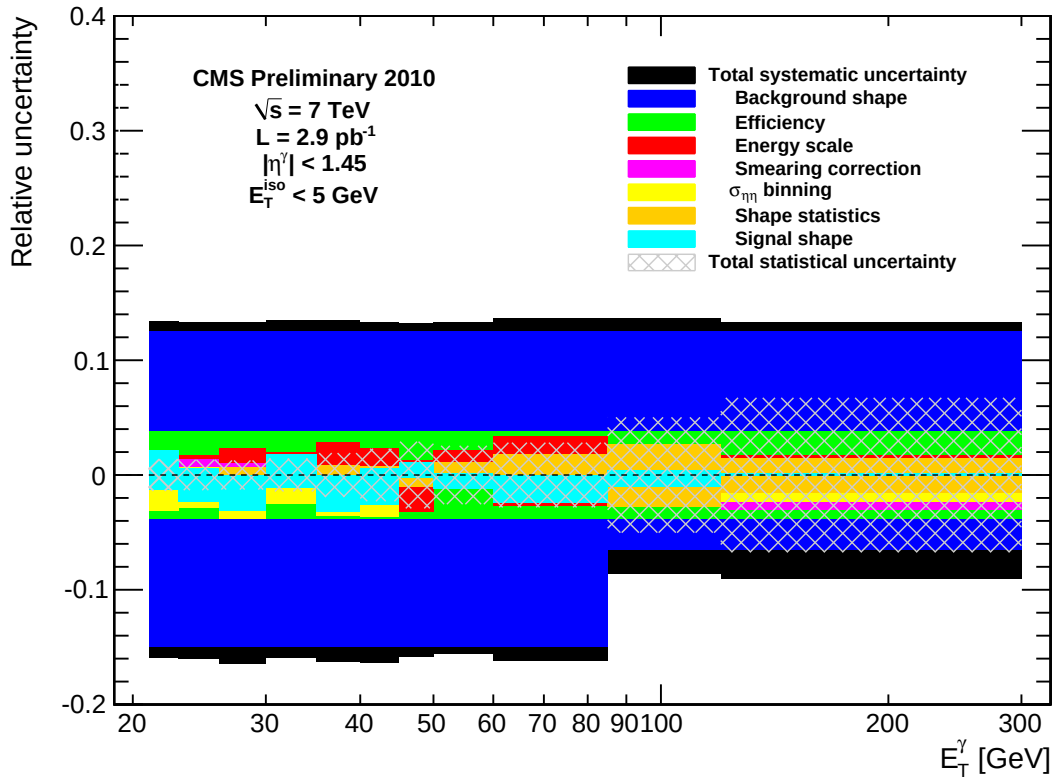


Figure 7.20: Summary of the systematic uncertainties, compared with the statistical uncertainty in each E_T bin.

7.7 Results and comparison to theoretical predictions

The measured isolated photon cross-section with the associated uncertainties is shown in Figure 7.21 and reported in Table 7.15.

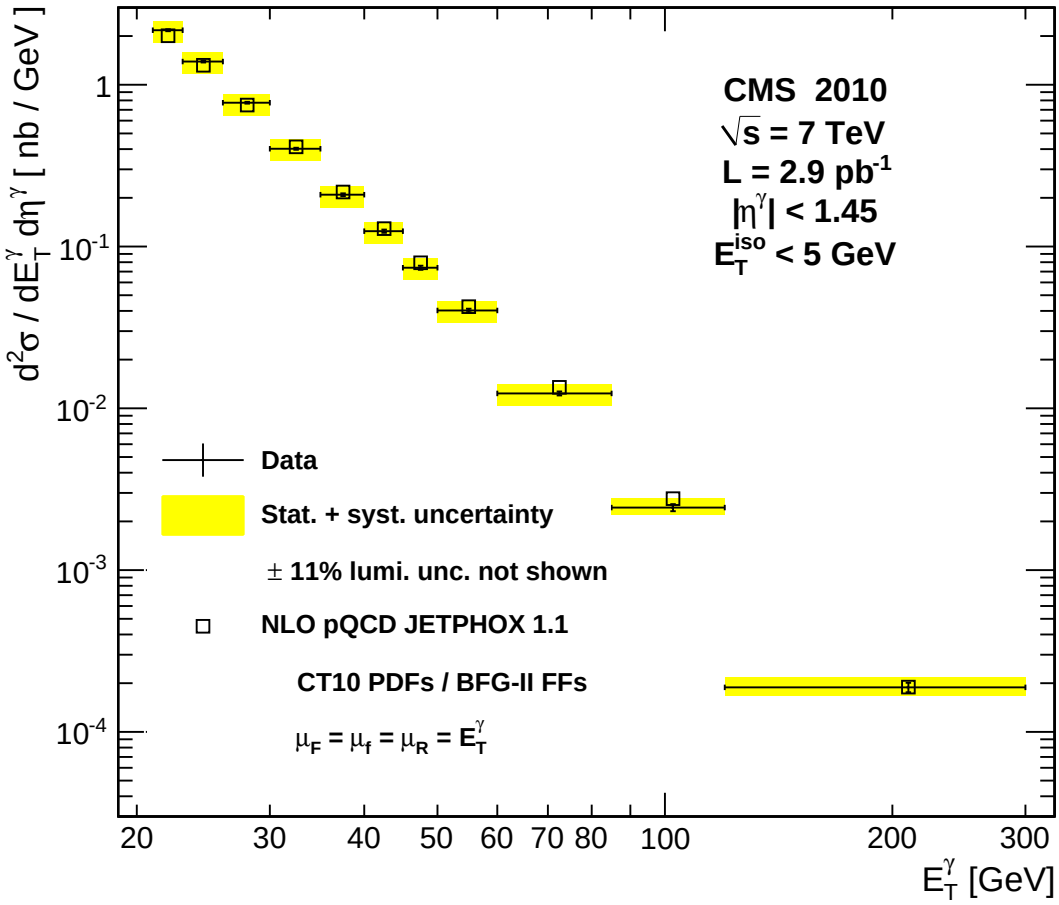


Figure 7.21: Differential isolated photon cross section.

The data were compared to NLO pQCD predictions from JETPHOX 1.1 [29] using the CT10 PDFs [30] and the BFG-II set of fragmentation functions (FFs) [31]. The normalisation, factorisation and fragmentation scales were set to E_{T}^{γ} and photon isolation was required as for the data, namely at most 5 GeV within $R < 0.4$. Setting all theoretical scales to $E_{\mathrm{T}}^{\gamma}/2$ and $2E_{\mathrm{T}}^{\gamma}$ changes the predictions by 6% to 11% with increasing E_{T}^{γ}. The uncertainty on the predictions due to the PDFs was estimated from the envelope of predictions obtained using three global fit parametrisations, CT10, MSTW2008 [101] and NNPDF2.0 [102], as recommended by the PDF4LHC working group [103]. This uncertainty was estimated to be about 6% over the considered E_{T} range, as shown in Figure 7.22. Predictions obtained using the CTEQ6.1M PDFs [104], extensively used in previous comparisons with data, are consistent with those obtained with CT10 to within 3%. Finally, using the BFG set I of FFs with softer gluons instead of BFG set II yielded negligible differences in the predictions.

The theoretical predictions include an additional correction factor $C(E_{\mathrm{T}}^{\gamma})$ to account for the presence of non-pQCD contributions from the underlying event and parton-to-hadron fragmentation, which tend to

Table 7.15: Measured differential cross section and associated uncertainties. An additional 11% uncertainty associated to the luminosity measurement has to be added to the quoted uncertainty.

E_T^γ (GeV)	$d^2\sigma/dE_T^\gamma d\eta^\gamma$ (nb/ GeV)	Syst. Unc. (%)
21–23	2.17 ± 0.03	$+13.4, -15.9$
23–26	1.39 ± 0.02	$+13.3, -16.0$
26–30	0.774 ± 0.010	$+13.3, -16.4$
30–35	0.402 ± 0.006	$+13.4, -15.9$
35–40	0.209 ± 0.004	$+13.4, -16.3$
40–45	$(124.4 \pm 2.8) \times 10^{-3}$	$+13.3, -16.3$
45–50	$(74.0 \pm 2.1) \times 10^{-3}$	$+13.2, -15.9$
50–60	$(40.3 \pm 1.0) \times 10^{-3}$	$+13.3, -15.6$
60–85	$(12.36 \pm 0.35) \times 10^{-3}$	$+13.6, -16.2$
85–120	$(2.43 \pm 0.12) \times 10^{-3}$	$+13.6, -08.6$
120–300	$(0.188 \pm 0.013) \times 10^{-3}$	$+13.3, -08.9$

increase the energy in the isolation cone [105]. Using Monte Carlo PYTHIA events, C was defined as the ratio between the isolated fraction of the cross section at the hadron level to the same fraction obtained after turning off both multiple parton interactions and hadronisation.

Four different PYTHIA tunes (Z2, D6T, DWT and Perugia-0 [106]) were considered. The correction factors obtained with the four different tunes are shown in Figure 7.23. The average among parameter sets, $C = 0.97 \pm 0.02$, has little E_T^γ dependence and was taken as the correction. As expected, the correction reduces the predicted cross section, since the presence of non-perturbative activity results in some photons failing the isolation requirements.
The ratio between the observed and the NLO prediction is shown in Figure 7.24. The measured cross section is roughly 5% larger than the predicted value for $E_T < 30$ GeV and 5% smaller for $E_T > 30$ GeV. The measured values agree with theoretical predictions bin-by-bin well within the experimental uncertainties.

The level of agreement of theoretical predictions to the observation is overall better than that found in experiments at lower $\sqrt{s}$, where an excess of the measured cross section with respect to theoretical predictions was typically observed in the lowest part of the x_T range [37].

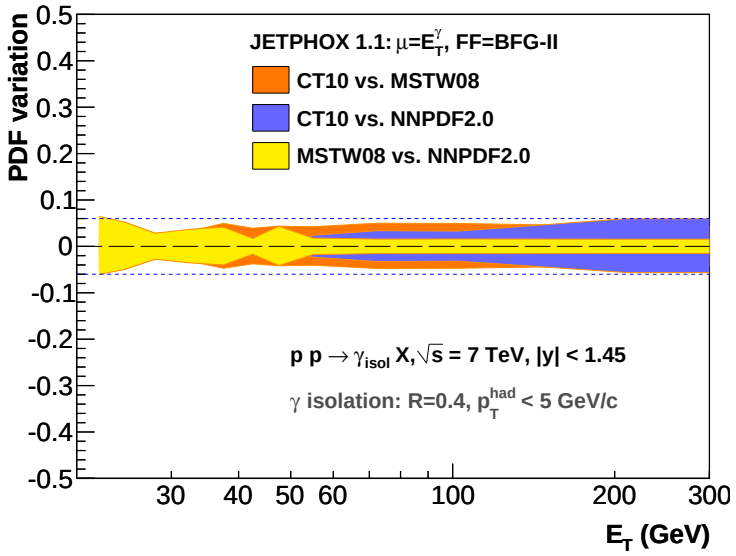


Figure 7.22: Fractional differences between the isolated photon spectrum in pp collisions at $\sqrt{s} = 7$ TeV obtained with JETPHOX (scales fixed at $\mu = E_T^\gamma$, FF set to BFG-II) using three different PDF parametrizations (CT10, MSTW08 and NNPDF2.0). The dashed lines indicate a ±6% uncertainty.

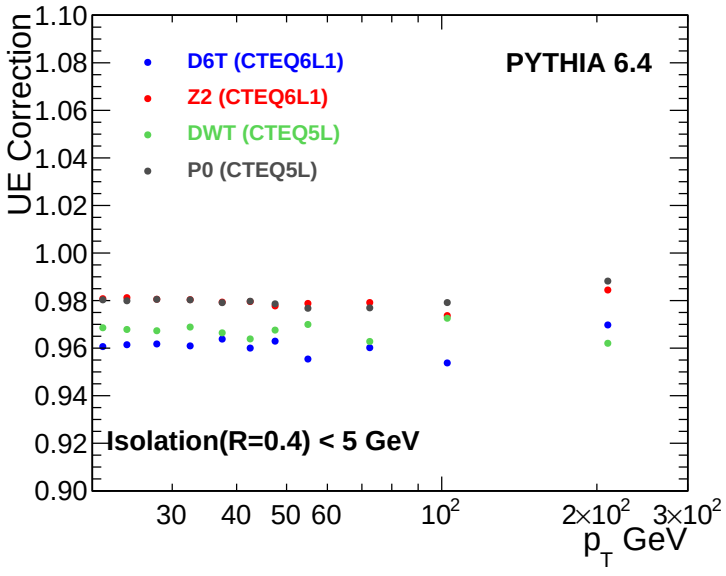


Figure 7.23: Estimation of non-perturbative effects not included in the theoretical calculations. Using PYTHIA-generated events, the estimate is computed as the ratio between the isolated fraction of the cross section at the hadron level and the same ratio obtained after turning off multiple parton interactions and hadronization. The average effect, $C = 0.97 \pm 0.02$, has no E_T^γ dependence.

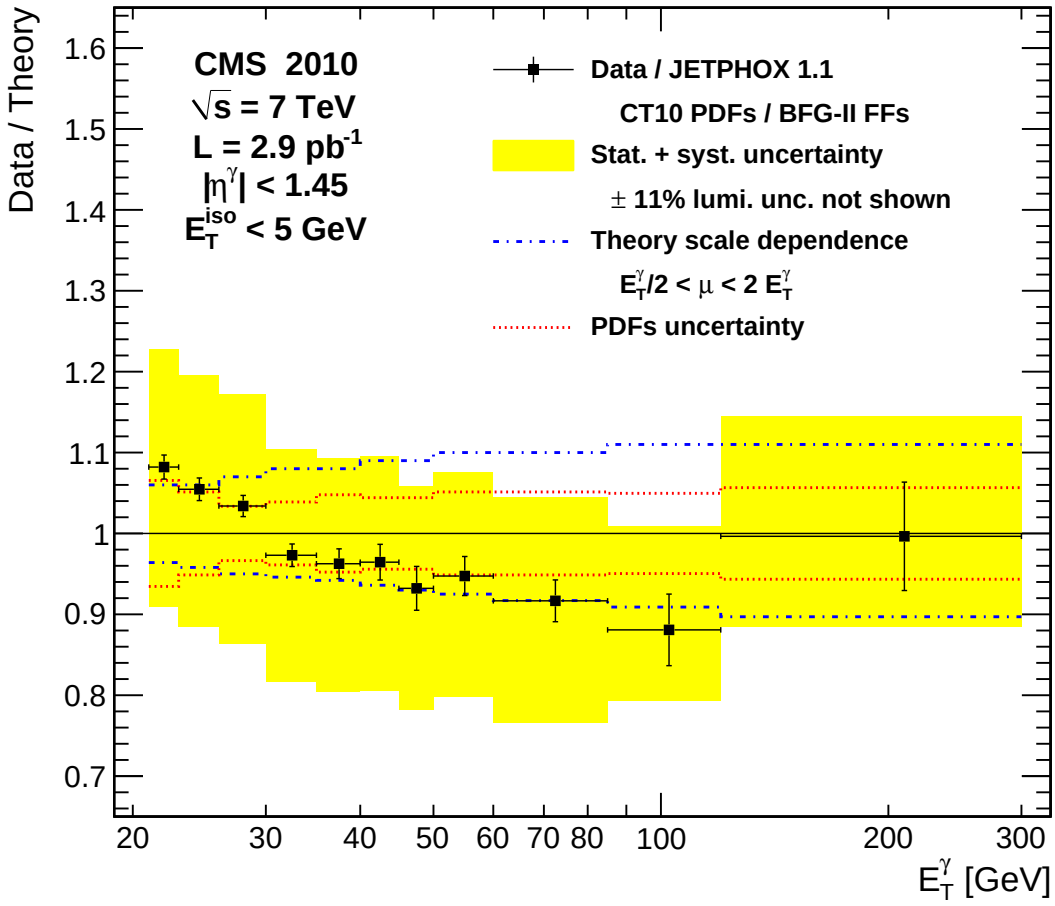


Figure 7.24: Comparison of data to pQCD predictions, including uncertainties on the latter due to scale variations and PDF uncertainties. A correction for underlying event effects ($C = 0.97 \pm 0.02$) is applied to the pQCD predictions..

7.8 Conclusions

The first measurement of the cross section for the production of isolated prompt photons with $21 < E_{\mathrm{T}}^{\gamma} < 300$ GeV and $|\eta^{\gamma}| < 1.45$ in pp collisions at $\sqrt{s} = 7$ TeV, using data corresponding to 2.9 ± 0.3 pb^{-1} was presented.
A data-driven technique was employed to extract the signal yield in data. The measurement relied only on a small set of variables that are easy to model already at the LHC start-up.
An extensive study of the systematic effects was performed. The dominant effect affecting the observation was found to be linked to the modeling of the background which was otherwise derived in a fully data-driven way.
This measurement was performed in the kinematic regime $0.006 < x_{\mathrm{T}} < 0.086$, probing a previously unexplored region at low x_{T}, and agrees with NLO pQCD predictions in the whole x_{T} range. This measurement established a benchmark for photon identification and background estimation, and constrained the rate of one of the background processes affecting searches for new physics involving photons.
The analysis presented here, represents the first measurement involving isolated high-energy final state photons performed at the LHC. The results were approved by the CMS collaboration for public release and a scientific communication was submitted for publication on the journal *Physical Review Letters*.

Chapter 8

Wγ production in pp collisions

Contents

This final chapter presents a feasibility study performed to establish the sensitivity of the CMS detector to associated Wγ production in pp collisions with early LHC data. The study was performed before the LHC start-up and was targeted for a centre of mass energy of 10 TeV. In particular, this study preceded all studies performed on data that were presented in previous chapters. Consequently, this analysis does not benefit from many of the lessons learnt with the first LHC data.

Even though it could not profit from the deeper understanding of the detector that was achieved with the LHC 2010 run, the impact of such limitations on the outcome of the study is small and the conclusions can be considered valid in light of the results obtained in collision events.

The Physics of Wγ associated production in pp collisions was discussed in Chapter 2. As seen there, the study of this final state provides information on the structure of the SM gauge group and in particular on the coupling between the photon and W fields.

Deviations from the SM prediction may be found looking at observables sensitive to the high energy limit of the Wγ scattering amplitudes. The most easily accessible of such observables is the photon E_{T} spectrum. In the presence of anomalous triple gauge couplings (aTGC), an enhancement in the production cross section with respect to the SM predictions is expected in the high E_{T} region of the spectrum ($E_{\mathrm{T}} \gtrsim M_W$).

The study presented below focused on establishing a simple event selection that would allow to measure the Wγ associated production cross section and the photon E_{T} spectrum, using events where the W boson decays to a $\mu - \nu_\mu$ pair, with early LHC data.

8.1 Event simulation

The study was performed on simulated events. Full detector simulation was performed assuming start-up detector conditions. The signal was generated using PYTHIA and a representative set of background processes were considered: W, Z and $\mathrm{t\bar{t}}$ inclusive production as well as multi-jet final states. Other backgrounds, such as associated Wγ production with other W decay modes and di-boson final states, were also checked and found not to be relevant. The standard CMS reconstruction software was employed to reconstruct the events.

In order to properly scale the event rates, k-factors, defined as the ratio between the NLO and LO cross section values, were calculated using the MCFM [107] Monte Carlo generator and applied to all background sample cross sections.

The signal production cross section was calculated at NLO in pQCD using the WGAMMA_NLO [41] event generator. The latter is a matrix element generator that calculates the Wγ production cross section at parton level with NLO accuracy. Since the generator does not calculate the final state evolution to hadronic states, it was simply used to calculate the cross-section k-factor to be applied to the PYTHIA generated samples. The kinematic region defined by the cuts listed in Table 8.1 was used to estimate the k-factor. The CTEQ6.1M [104] PDF set was used in the calculation.

An average k-factor of 1.5 was calculated. The uncertainty due to the knowledge of parton distribution functions (PDF) was evaluated, using the "Hessian" method [108], to be of the order of 10%. The PYTHIA (LO) cross-section rescaled by the k-factor extracted from WGAMMA_NLO is $\sigma^{\mathrm{rescaled}} = 18.3 \pm 1.8$ pb for each leptonic decay channel of the W boson.

In order to avoid double-counting, events containing prompt photons were removed from inclusive W sample using MC-truth information. In particular, all events with a generated photon radiated by a final state muon or by an initial or final state quark were removed from the W+jet sample.

A summary of all samples used in the analysis is reported in Table 8.2. Events from different samples were weighted according to the corresponding cross sections in order to simulate the expected event yield for an integrated luminosity of 50 pb^{-1}.

Table 8.1: Generator-level cuts used to evaluate the Wγ k-factor.

Variable	selection
$p_{\mathrm{T}}(\mu)$	> 5 GeV/c
$\lvert\eta(\mu)\rvert$	< 8
$p_{\mathrm{T}}(\gamma)$	> 5 GeV/c
$\lvert\eta(\gamma)\rvert$	< 8
$p_{\mathrm{T}}(\nu)$	> 5 GeV/c
$\lvert\eta(\nu)\rvert$	< 8
$\Delta\mathrm{R}(\mu,\gamma)$	> 0.05
$\mathrm{M_T}(\mu,\gamma;\nu)$	> 10 Gev/c^2
$\mathrm{E_T^{had}}(\mathcal{R}=0.4)/\mathrm{E_T^{\gamma}}$	< 0.15

Table 8.2: Samples used in this analysis. Cross-section values correspond to NLO predictions.

Process	Generator	σ[pb]	n. events	$\int \mathcal{L}\cdot dt[\mathrm{pb}^{-1}]$
W($\mu\nu$)γ	PYTHIA	18.3	103,122	$9.37\cdot 10^3$
W(τ/eν)γ	PYTHIA	36.6	103,122	$9.37\cdot 10^3$
t $\bar{\mathrm{t}}$ jet	MADGRAPH	375	792,896	$2.11\cdot 10^3$
W jet	MADGRAPH	$3.56\cdot 10^4$	9,645,360	$2.71\cdot 10^2$
Z jet	MADGRAPH	$3.54\cdot 10^3$	1,054,278	$2.98\cdot 10^2$
QCD	PYTHIA	$1.21\cdot 10^5$	6,265,808	52

8.2 Event reconstruction and selection

In this study, the $W \to \mu\nu$ decay channel was used to probe Wγ production. The final state includes one high-p_{T} muon, missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$), together with a high-$E_{\mathrm{T}}$ photon.
The algorithms used to reconstruct the final state particles in the CMS detector are described below.

Muons

Muon reconstruction combines tracking information from the inner silicon tracker and from the outer muon spectrometer [109]. Muon tracks reconstructed in the two sections of the detector are combined into a single object known as *global muon*. The track matching quality is evaluated through a χ^2 variable that takes into account the difference in the reconstructed variables from the inner tracker and the muon spectrometer, using the ratio of charge-to-momentum (q/p) and the extrapolated track position at the matching surface (located between the Tracker and the Muon detectors). Muons were selected as follows:

- p_T >20 GeV/c;
- $|\eta| < 2.1$;
- χ^2_{match}/n.d.o.f. < 5;
- number of tracker hits > 12;
- distance of closest approach to the vertex in the transverse plane < 0.4 mm.

In addition, a muon isolation criterion was defined by the requirement $\sum \mathrm{p_T^{track}} < 0.09 \cdot \mathrm{p_T^{\mu}}$, where the track index runs over all tracks within 1(2) mm from the interaction vertex in the $x-y$ transverse plane (z-direction) reconstructed by the silicon tracker in a cone of radius $\mathcal{R} = 0.4$ around the muon direction, excluding the region at $\mathcal{R} < 0.01$.

Missing transverse energy

Missing transverse energy was measured from the imbalance in the uncorrected calorimetric energy deposits in projective calorimetric towers [110]. This quantity was corrected for several effects [110, 111]. Two types of corrections were used by the reconstruction algorithm: i) "jet energy scale" corrections to take into account the differences between the raw and the corrected jet energies, and ii) "muon" corrections to account for the presence of muons in the event, correcting the missing transverse energy for the muons' momenta and the energy that the muons deposit in the calorimeters.
The algorithm described above has been improved in versions of the CMS reconstruction software more recent than the one used for this analysis. A considerable improvement in resolution was obtained from including information from the silicon tracker in the computation of missing energy [112]. In light of this progress, the result of the study presented here, though valid, can be expected to improve.

Photons

The photon reconstruction and identification algorithms were described in details in Chapter 6. The photon identification variables used for this study were the same described there, namely the $\mathrm{Iso_{ECAL}}$, $\mathrm{Iso_{HCAL}}$, $\mathrm{Iso_{TRK}}$ and H/E variables. Since the thresholds used in the computation of the isolation sums had slightly different values with respect to what was described in Chapter 6, the exact definitions are reported below. The isolation sums were defined in an cone of radius $\mathcal{R} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$.

- $\mathrm{Iso_{TRK}}$ was defined as the scalar sum of the p_T of the tracks consistent with the primary vertex of the interaction. An exclusion region consisting of an inner cone of radius $\mathcal{R} = 0.05$ was defined.

- $\mathrm{Iso_{ECAL}}$ was defined as the sum of transverse energy deposited in the electromagnetic calorimeter crystals. In order to remove the contribution due to the photon candidate, an exclusion region consisting of an inner cone of radius $\mathcal{R} = 0.06$ and a strip of dimensions $\Delta\eta \times \Delta\phi = 0.04 \times 0.4$ was defined. In the barrel (endcap) region, energy (transverse energy) deposited in a single crystal was required to be larger than 80(100) MeV.

- $\mathrm{Iso_{HCAL}}$ was defined as the sum of the energy deposited in the hadronic calorimeter towers, excluding an inner cone of radius $\mathcal{R} = 0.01$.

- The H/E ratio was defined as the ratio between the energy deposited in the HCAL in a cone of radius $\mathcal{R} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.01$ around the SC direction and the energy of the photon candidate.

The distributions of these variables are compared in Figure 8.1 for signal (prompt photons in Wγ events) and background (photons from meson decays in W+jet events) events. Slight differences with respect to the plots shown in Chapter 6 can be observed and are due to small differences in the underlying detector simulation and to the different thresholds applied in the computation of the isolation sums. The photon identification criteria applied were:

- $\mathrm{Iso_{ECAL}} < 5\ \mathrm{GeV} + 0.0075 \cdot \mathrm{E_T}$
- $\mathrm{Iso_{HCAL}} < 2\ \mathrm{GeV} + 0.002 \cdot \mathrm{E_T}$
- $\mathrm{Iso_{TRK}} < 5\ \mathrm{GeV/c}$
- $\mathrm{H/E} < 0.05$
- No pixel hits consistent with tacks from the hard interaction matching the photon candidate position were allowed.

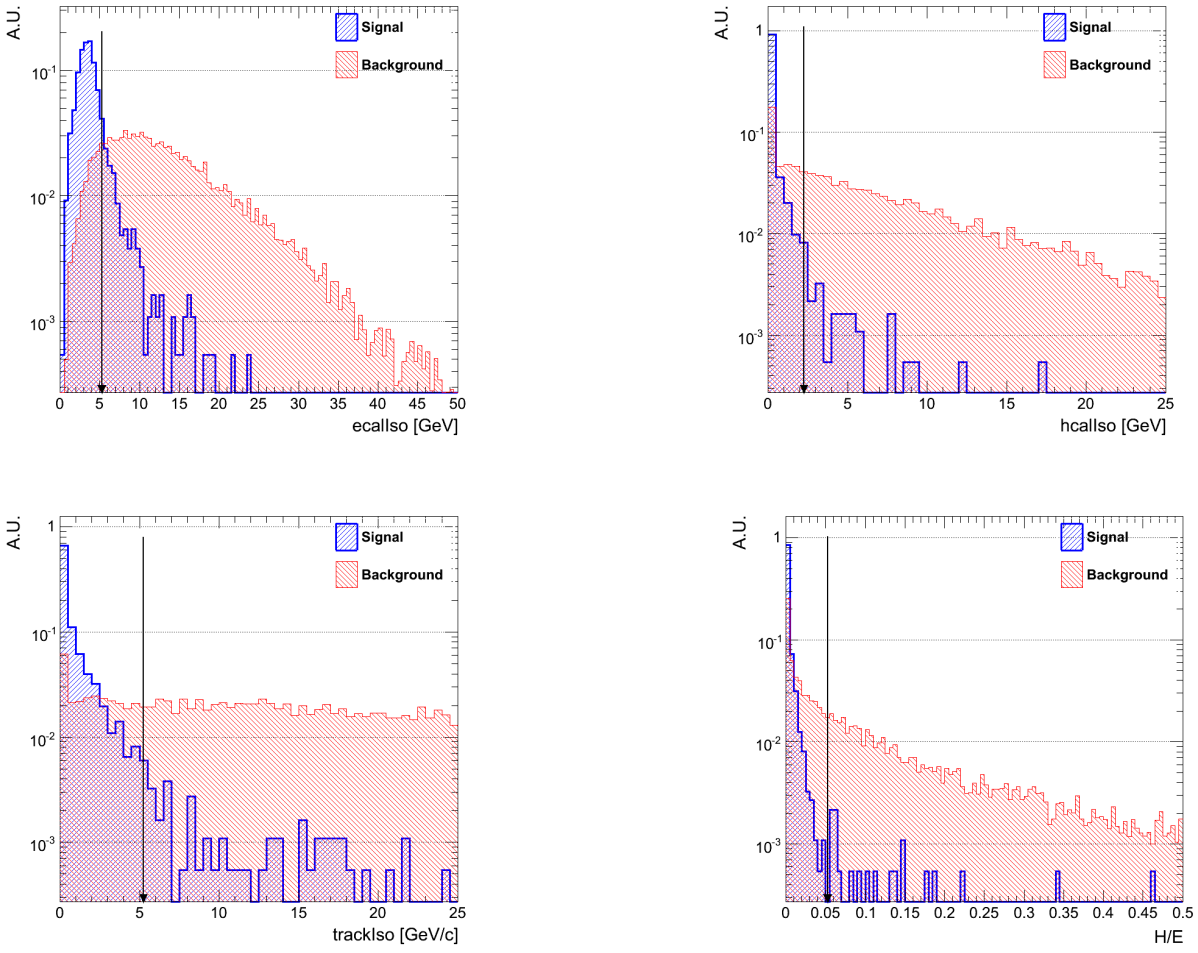


Figure 8.1: Distributions used in photon identification for signal (Wγ, blue) and background (W+jets, red): ECAL isolation, HCAL isolation, tracking isolation, and H/E. The arrows indicate the values of the selection cuts used.

Event selection

The simulated events were required to satisfy the selection used for the measurement of the inclusive W production cross section [113] and in addition the presence of a reconstructed isolated photon was required.
The simulated events were required to pass the HLT_Mu15 trigger, that requires a single non-isolated muon with $p_T > 15$ GeV/c to be reconstructed by the HLT.
Events satisfying the trigger criterion were further preselected requiring the presence of at least one reconstructed global muon with $p_{\rm T} > 15$ GeV/c and at least one reconstructed photon with $E_{\rm T} > 15$ GeV. Furthermore, the following criteria were applied:

1. The presence of one isolated muon candidate with $p_{\rm T} > 20$ GeV/c and $|\eta| < 2.1$ was required. Events with two or more muon candidates with $p_{\rm T} > 10$ GeV/c were rejected;
2. Events were required to have $\not\!\!E_{\rm T} > 20$ GeV;
3. The transverse mass of the $\mu+\not\!\!E_{\rm T}$ system, defined as $m_T = \not\!\!E_{\rm T} p_{\rm T}^{\mu}[1 - \cos\Delta\phi(\not\!\!E_{\rm T}, \mu)]$, had to be larger than 50 GeV/c^2;
4. The presence of one photon with $E_{\rm T} > 20$ GeV and $\eta < 1.45$, satisfying all the criteria outlined in Section 8.2 was required. Events with two or more photon candidates with $E_{\rm T} > 10$ GeV/c were rejected;
5. The photon was required to be separated from the muon, in order to suppress the contribution from the radiative W decays. In particular the separation between the reconstructed photon and muon was required to be $\mathcal{R} > 0.7$.

A summary of the event selection criteria is reported in Table 8.3, and the distribution of the variables involved is shown in Figure 8.2.

Table 8.3: Summary of the event selection criteria.

Variable	selection
$p_{\rm T}(\mu)$	> 20 GeV/c
$E_{\rm T}^{\rm miss}$	> 20 GeV
$m_{\rm T}(\mu, \not\!\!E_{\rm T})$	> 50 GeV/c^2
$p_{\rm T}(\gamma)$	> 20 GeV/c
$\Delta{\rm R}(\mu, \gamma)$	> 0.7

The event yield expected after each selection cut for an integrated luminosity of 50 pb^{-1} is reported in Table 8.4. The selection criteria allow one to improve the signal to background ratio by four orders of magnitude, from 10^{-4} to ~ 1. After full event selection, the number of expected signal events was found to be 44 ± 7, with a total number of expected background events of 50 ± 3. The largest background was found to be due to W+jet events, where a jet fragments mainly into a neutral meson that mimics an isolated photon. Background from jet events was found to be $\sim 6\%$ of the total. The expected photon $E_{\rm T}$ distribution for the final sample is shown in Figure 8.3. It is important to note that the purity of the selected sample is expected to improve for $E_{\rm T}^{\gamma} > 30$ GeVc, even if the total number of events becomes smaller: a total of 36 signal events are expected to remain with a signal to background ratio of ~ 1.2.

8.3 Background determination

Most of the physical information on the WWγ vertex structure can be extracted from the shape of the photon $E_{\rm T}$ spectrum. In particular, non-SM couplings are expected to enhance the number of events

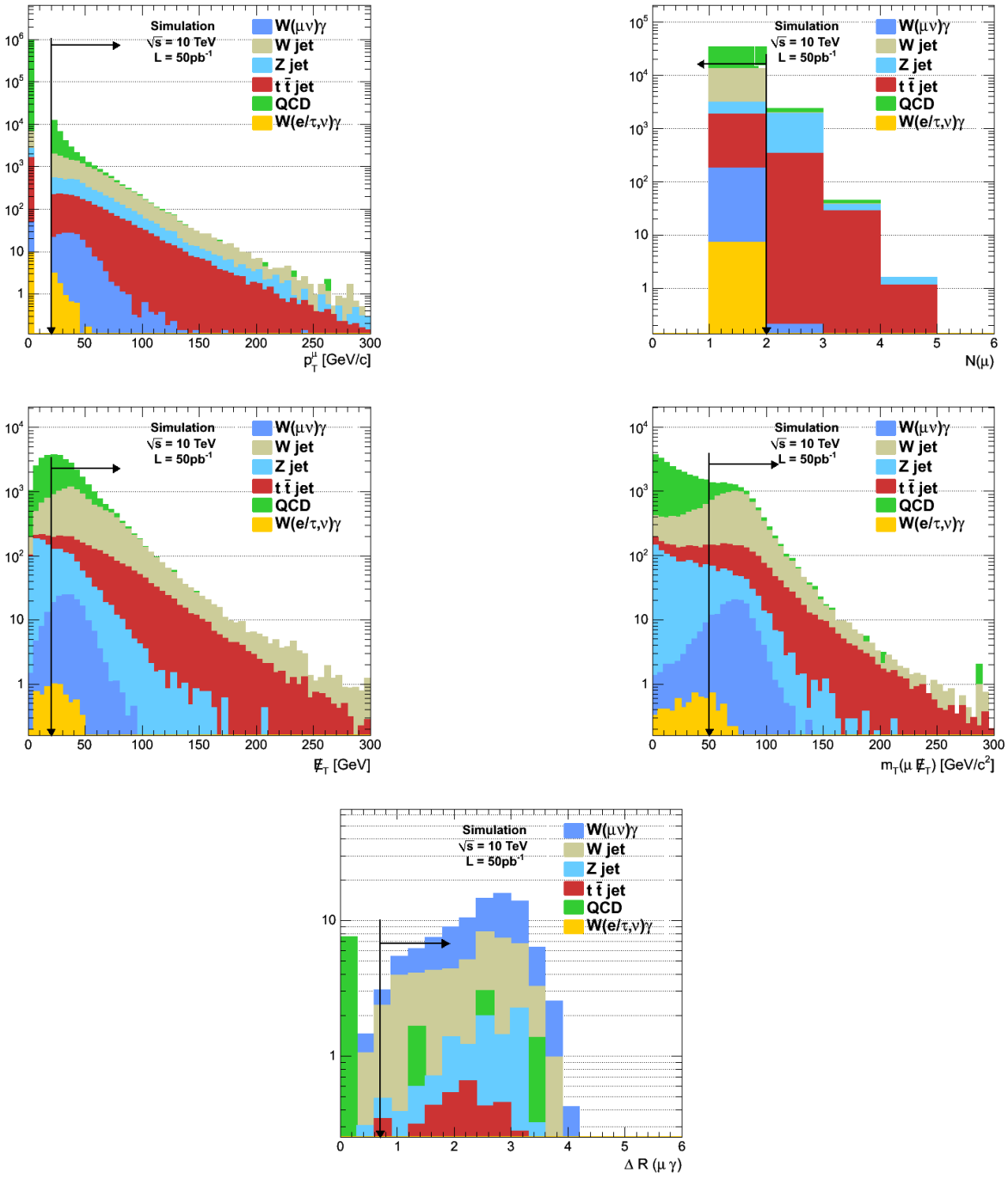


Figure 8.2: Event selection: distribution of the variables used. From left to right, top to bottom, each plot includes the cuts of all previous ones. Histograms from different samples are stacked. (top left) p_{T} distribution for the most energetic isolated muon. Events with no isolated muon contribute to the first bin; (top right) number of non-isolated muons; (middle left) $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution; (middle right) distribution of the $(\not{E}_{\mathrm{T}}, \gamma)$ transverse mass; (bottom) distribution of $\Delta\mathrm{R}(\mu, \gamma)$.

Table 8.4: Expected number of events in each sample after each selection cut, for a luminosity of 50 pb^{-1} and $\sqrt{s} = 10$ TeV.

	W($\mu\nu$)γ	W + jets	Z + jets	t $\bar{\text{t}}$ + jets	QCD(μ)	total B	S/B	S/$\sqrt{\text{B}}$
trigger	211	14,201	3,797	3,611	919,712	941,338	10^{-4}	0.2
1 isolated μ	172	10,206	2,773	2,030	20,976	35,992	10^{-3}	0.9
2 μ veto	172	10,140	1,203	1,663	20,594	33,606	10^{-3}	0.9
$\not{E}_{\text{T}}$	147	8,751	639	1,538	12,310	23,242	10^{-3}	1.0
$m_t(\mu, \not{E}_{\text{T}})$	135	6,697	288	1,094	2,491	10,571	10^{-2}	1.3
γ selection	55 ± 7	2,645	89 ± 9	492	940	4,167	10^{-2}	0.9
γ ID	45 ± 7	37 ± 6	7 ± 3	3 ± 2	10 ± 3	59 ± 4	0.8	5.8
Δ R(μ, γ)	44 ± 7	36 ± 6	7 ± 3	3 ± 2	3 ± 2	50 ± 3	0.9	6.3

produced at large values of the photon transverse energy [41]. It is therefore of primary importance to carefully characterise and subtract all the sources of background that may prevent a correct assessment of the photon E_{T} spectrum.
Two types of background leading to a distortion of the photon E_{T} spectrum were identified. The largest background was found to be due to neutral mesons produced mainly in W+jet events and amounts to $> 85\%$ of the predicted background. The remaining 15% of the background is made by prompt photons, mostly coming from associated Zγ production (contained in the Z+jet sample).

8.3.1 Neutral meson background

As said above, the largest background to Wγ production is expected to come from neutral mesons in jets. This contribution was estimated using the two component fit employed to measure the isolated photon cross section in pp collisions and described in the previous chapter.
For this study, the square of the $\sigma_{i\eta i\eta}$ variable was used to perform the two component fits. The background component shape was extracted from a Iso_{TRK} side-band. In particular, events with $5 < \text{Iso}_{\text{TRK}} < 10\,\text{GeV}/c$ were selected to model the background. For the signal component shape, a slightly different approach with respect to the one applied for the isolated photon cross section measurement was employed and the signal component shape used was directly extracted from simulated Z→ee events.
In order to perform the two component fits three bins in the photon E_{T} were used: 20–30, 30–50 and 50–120 GeV. Figure 8.4 shows the distribution of the $\sigma^2_{\eta\eta}$variable for four samples:

- Direct photons from $\gamma + jet$ production in PYTHIA, labelled $\underline{\gamma \text{ MC}}$.
- Isolated electrons from Z→ee events, labelled $\underline{\gamma \text{ data-driven}}$.
- Photon candidates from hadron decays in jet events, labelled $\underline{\text{MC background}}$.
- Photon candidates from the side-band selection, labelled $\underline{\text{Background data-driven}}$.

In Figure 8.5, the average of the component shapes distributions is plotted as a function of the photon candidates pseudorapidity.
"Data-driven" and MC distributions show good agreement overall. Isolated electrons appear to have slightly lower values of $\sigma^2_{\eta\eta}$with respect to isolated photons and also a small bias of the background shape can be observed, due to the use of a relatively large side-band in this study. In light of the performances obtained in the measurement of the isolated photon cross section described in Chapter 7, a better modeling of the component shapes can be expected.
The results of the two component fits are shown in Figure 8.6, In Figure 8.7 the prompt photon spectrum determined by the fit is compared to the prompt photon spectrum obtained using MC-truth information. The latter can be observed to be reproduced within uncertainties. The number of isolated photons in the

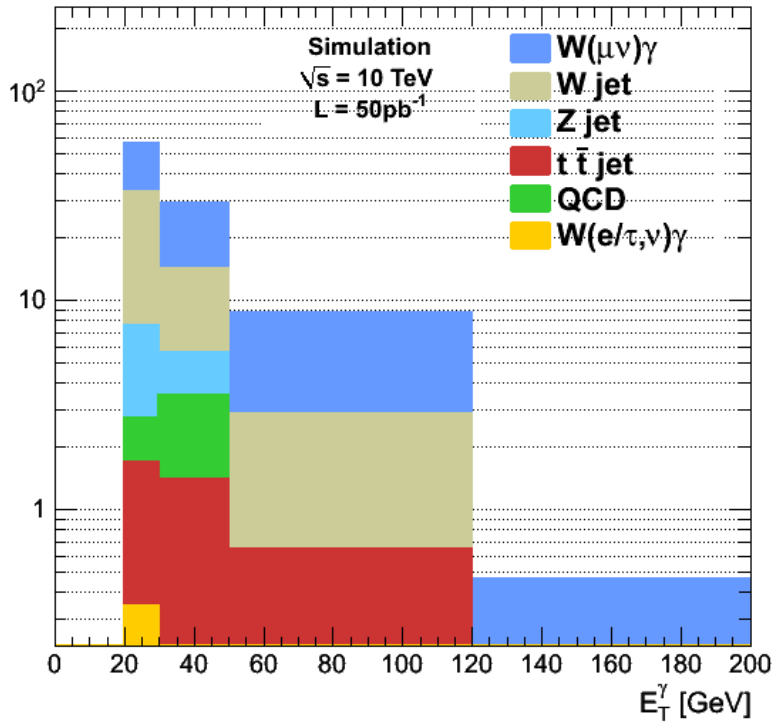


Figure 8.3: Photon E_{T} distribution after the full event selection. Histograms from different samples are stacked. The dominant background comes from W+jet events.

final sample was 54, while the two component fit estimated a total of 60 ± 10 events. The true content of isolated photons could therefore be correctly estimated within statistical uncertainties.

8.3.2 Isolated photon background

After subtracting the fake photon contribution (by means of the two component fits), the fraction of isolated photons coming from Wγ events was determined in order to be able to measure the Wγ production cross section.
According to MC, approximately 90% of the selected isolated photon candidates are from Wγ and the remaining comes from Zγ events.
Associated Zγ production is an almost irreducible background to Wγ production if one of the two leptons from the Z falls outside the detector acceptance and it is not reconstructed. On the other hand, Zγ events are produced though an electroweak process and the MC modeling of such events can be expected to be accurate. In the following, it was thus assumed that this background will be subtracted using MC information.

8.3.3 Systematic uncertainties

The main systematic uncertainties due to the background estimation were evaluated.
The uncertainty due to the knowledge of the component shapes was estimated from the difference between the signal yield estimated using the MC shapes in Figure 8.4 and the "data driven" ones. A total systematic uncertainty of $\sim 10\%$ was estimated.

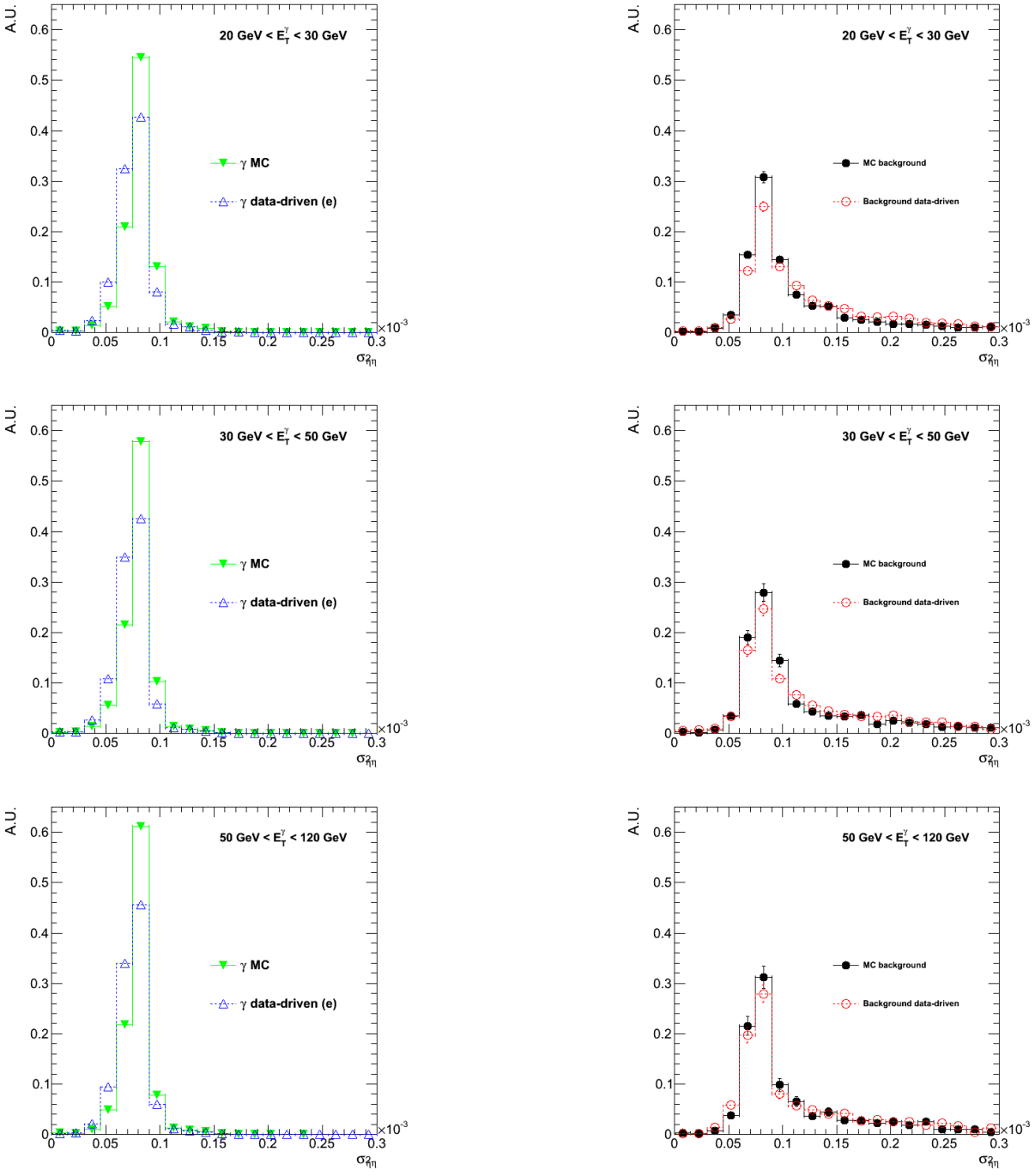


Figure 8.4: Comparison of the $\sigma^2_{\eta\eta}$ component shapes obtained: MC truth (solid) and "data-driven" (open) samples defined in the text. Signal events are in the left column: MC-matched isolated photons from γ+jet production (green, solid triangles) and "data" from electrons from Z→ee decays (blue, open triangles). Background events are in the right column: photon candidates from hadron decays (black, solid circles) and "data" obtained by inverting the isolation requirement as described in the text (red, open circles).

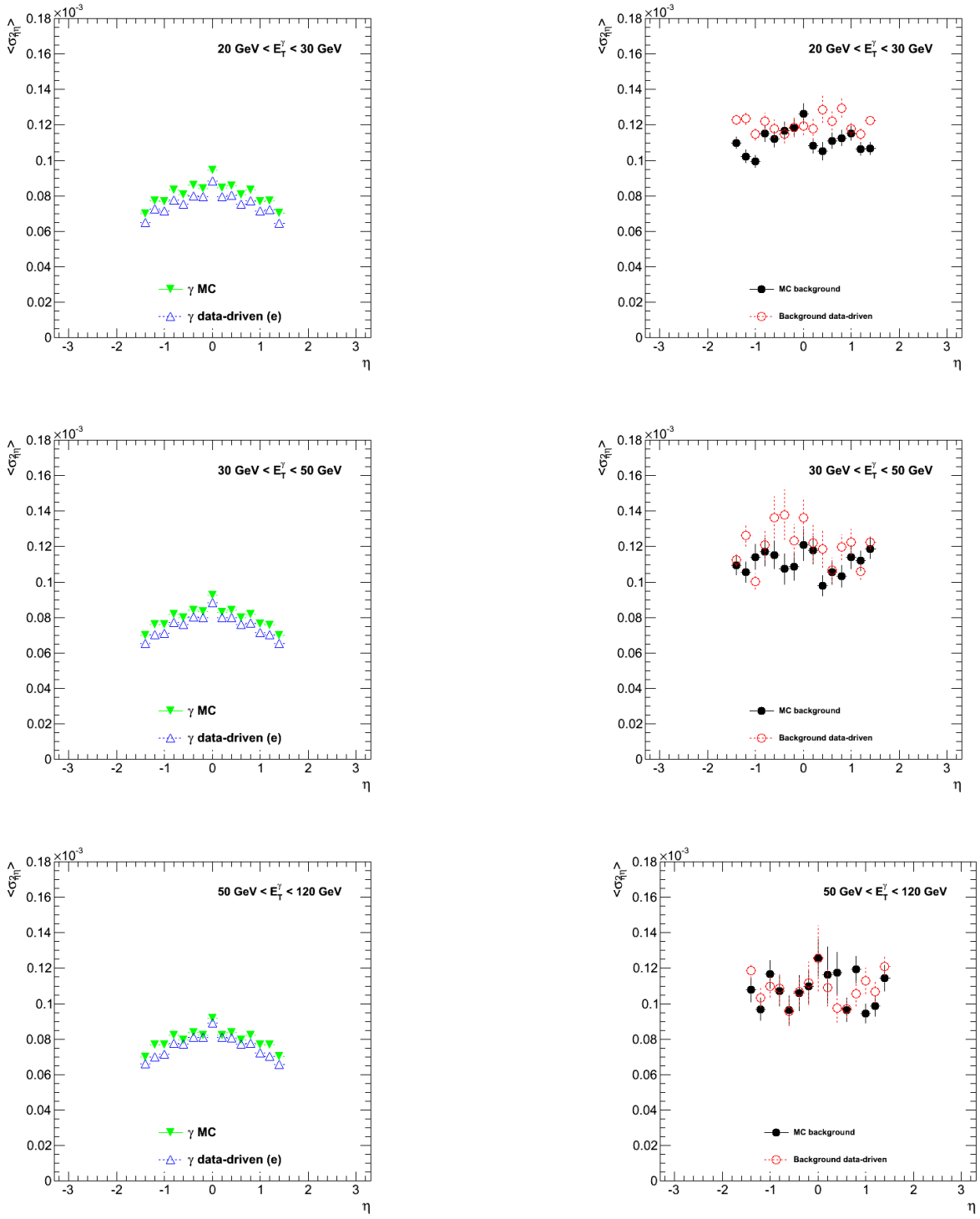


Figure 8.5: Comparison of the η-dependence of the $\sigma^2(\eta\eta)$ variable for the MC and the "data-driven" component shapes for signal (left) and background (right).

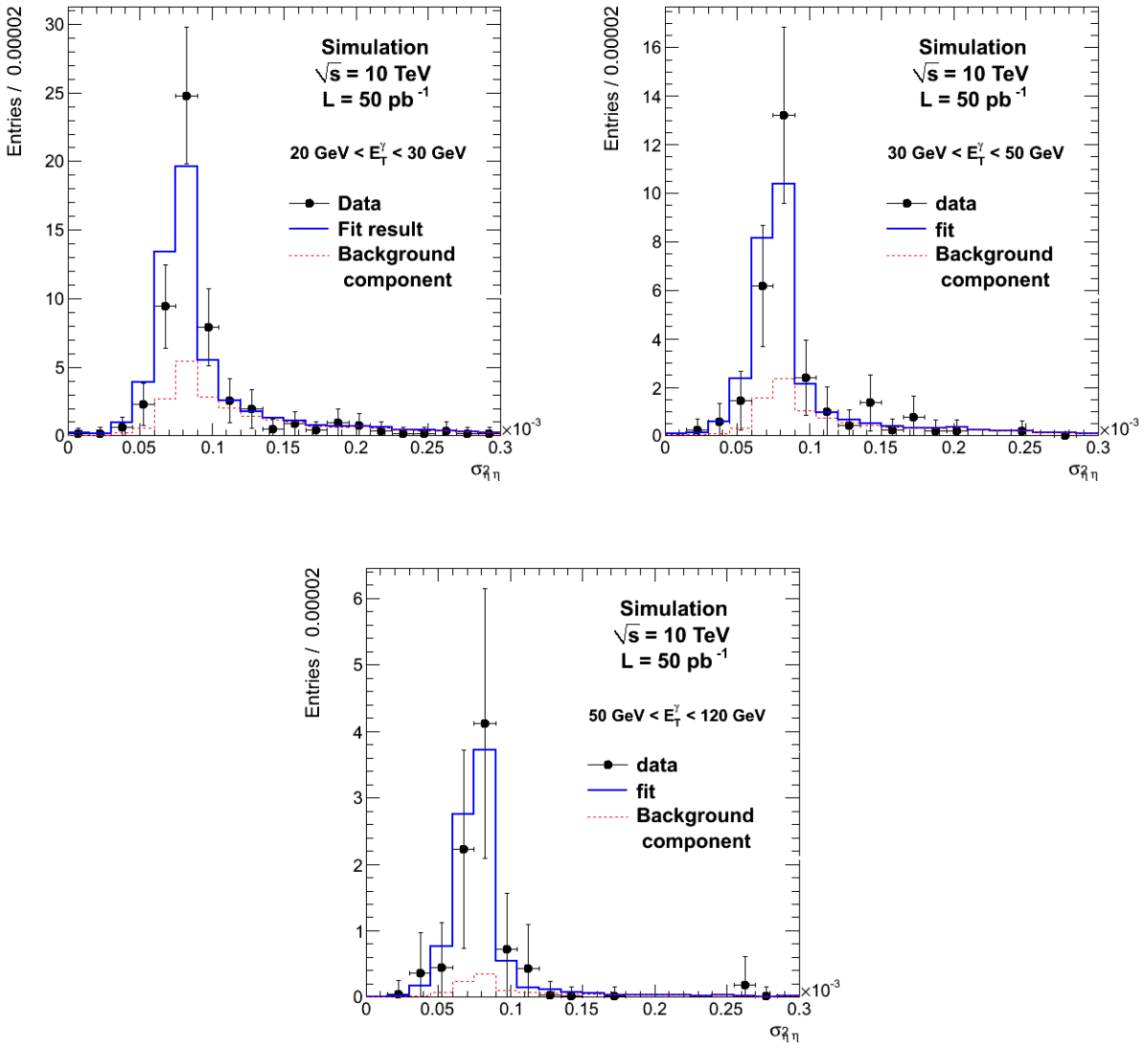


Figure 8.6: Template method fits for three $p_T(\gamma)$ bins, as defined in the text.

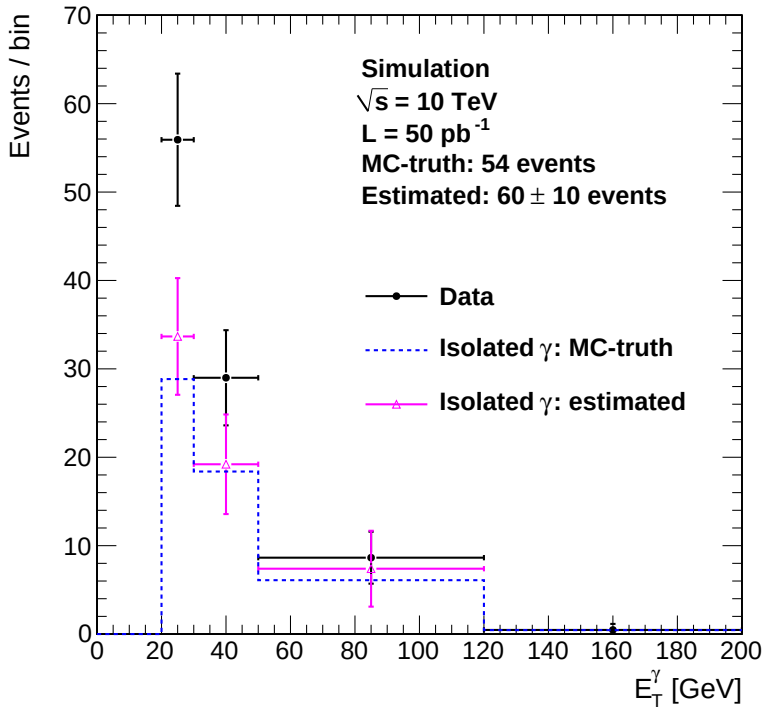


Figure 8.7: Expected photon E_{T} distribution for all W+γ candidate events in 50 pb^{-1} (solid histogram). The isolated photon yield is estimated, as a function of E_{T}, using the two component fit described in the text and compared to the true yield in MC.

The systematic uncertainty linked to the choice of the binning used for the two-component fit was studied varying the bin size by factors of 2, 3, 4 and 5. The number of isolated photons estimated from the template fit changed by $^{+0\%}_{-8\%}$.
Finally, the dependence of the results on size of the electron control sample was studied. The analysis was repeated extracting the signal component shape from Z samples equivalent to 12, 25 and 50 pb^{-1}. The total signal yields estimated for the different size of the Z sample are reported in Table 8.5. The effect can be seen to be negligible.

Table 8.5: Total signal yield estimated using Z samples corresponding to different integrated luminosities to extract the signal component shape.

Z $\int \mathcal{L} \cdot dt[pb^{-1}]$	N(Z $\rightarrow$ ee)	estim. $N_{\mathrm{Iso}}(\gamma)$
12	5987	60 ± 8
25	11945	60 ± 8
50	23839	61 ± 9

The total systematical uncertainty was estimated bin-by-bin by taking the largest variation in the fitted value. The estimated photon E_{T} spectrum including the systematic uncertainties is shown in Figure 8.8. A total systematic uncertainty of ($\pm$15%) was estimated.

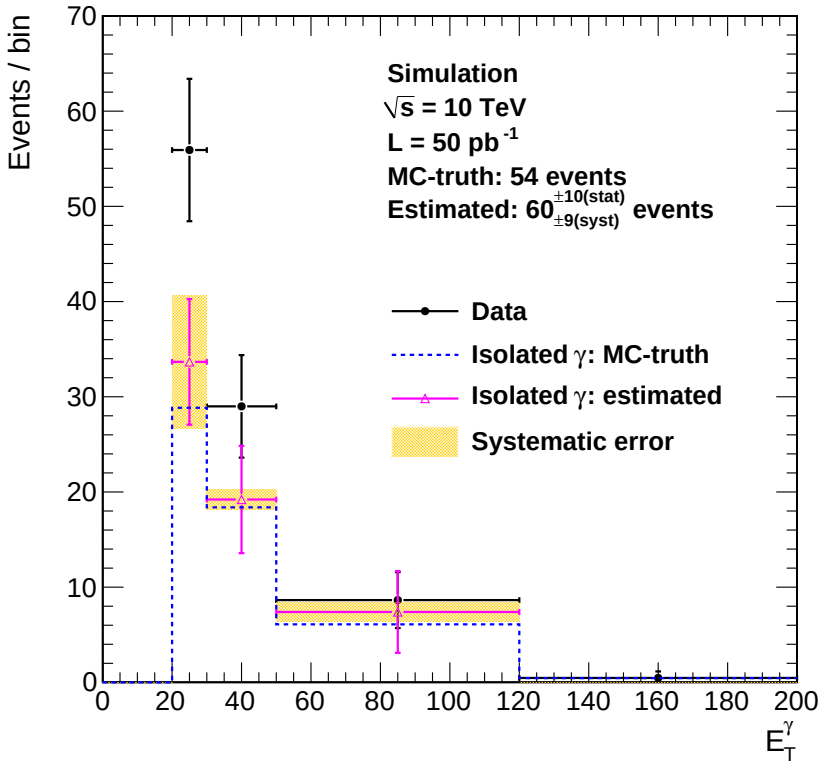


Figure 8.8: Photon E_{T} distribution. Same as Figure 8.7, including estimated systematic uncertainties.

8.4 Cross section measurement

The cross-section for Wγ production was extracted using the following expression:

$$\sigma\left(pp \to W\gamma\right) \cdot BR\left(W \to \mu\nu\right) = \frac{N_{tot} - N_{bkg}}{A \cdot \varepsilon \cdot \int \mathcal{L} \cdot dt} \tag{8.1}$$

where $\mathrm{N_{tot}}$ is the total number of events containing a W$(\mu\nu)\gamma$ candidate after full event selection, $\mathrm{N_{bkg}}$ is the number of background events, $\int \mathcal{L} \cdot dt$ the total integrated luminosity, and $A \cdot \epsilon$ the product of the geometric and kinematic acceptances and the selection efficiency. To reflect the way in which the various quantities were estimated, the expression can be rewritten as:

$$\sigma\left(pp \to W\gamma\right) \cdot BR\left(W \to \mu\nu\right) = \frac{N(isolated\ \gamma) - N(Z\gamma)}{A \cdot \varepsilon \cdot \int \mathcal{L} \cdot dt} \tag{8.2}$$

where N(isolated γ) is the number of isolated photon events estimated with the two-component fit, and $N(Z\gamma)$ is the number of Zγ background events estimated from MC.

8.4.1 Efficiency and acceptance

In order to estimate the signal efficiency and acceptance, signal events were split in two sub-samples: one, containing 2/3 of the events was used as the "data" sample, while the second, containing 1/3 of the events, was used to measure acceptances and efficiencies. The $\mathrm{A} \cdot \varepsilon$ term was factorised as:

$$A \cdot \varepsilon = A \cdot \varepsilon_{reco} \cdot \varepsilon_{trg} \cdot \varepsilon_{ID} \cdot \varepsilon_{veto} \tag{8.3}$$

where:

1. $\mathrm{A} \cdot \varepsilon_{\mathrm{reco}}$ is the probability that a W$(\mu\nu)\gamma$ event is reconstructed in the geometric and kinematic regions defined by the analysis selection criteria. It contains: the requirements on photon, muon and missing transverse energy (η and p_{T} cuts), the muon-photon separation and the muon-$E_{\mathrm{T}}^{\mathrm{miss}}$ transverse mass requirement;

2. $\varepsilon_{\mathrm{trg}}$ is the trigger efficiency;

3. $\varepsilon_{\mathrm{ID}}$ is the efficiency of the photon and muon particle identification criteria;

4. $\varepsilon_{\mathrm{veto}}$ is the efficiency due to the requirement of no additional photons and muons in the event.

Such parametrisation allows one to separate those terms that can be measured directly from data from those which rely on the use of MC. In particular, terms 2 and 3 can be measured directly from data, while for terms 1 and 4 some MC inputs are required. Muon reconstruction and trigger efficiencies can be measured in events where di-muon resonances are produced using the "tag-and-probe" technique. For this aspect, the results of the W inclusive analysis [113], showing that an unbiased measurement of these efficiency can be obtained from Z decays in muon channel, were assumed. In order to measure photon efficiencies from data, the "tag-and-probe" technique can be used on events with di-electron resonances, as discussed in Chapter 7.
The numeric values of all terms entering the acceptance computation are reported in Table 8.6. The total $\mathrm{A} \cdot \varepsilon$ was estimated to be 0.041 ± 0.002(stat). The uncertainty related to the knowledge of the muon, photon and missing transverse energy scales was estimated, assuming the start-up uncertainties expected at the time of the study, to be $^{+0.07}_{-0.06}$(syst) as reported in Table 8.7.

Table 8.6: Acceptance and efficiency values of the different selection criteria.

Criterion	relative eff.	cumulative eff.
$A \cdot \epsilon_{reco}$		5.1%
$\gamma, \mu, \not{E}_T$ (p_T and η cuts)	7.2%	
$\Delta R(\mu, \gamma)$	98%	
$m_T(\mu, \not{E}_T)$	80%	
ϵ_{trg}	90%	
$A \cdot \epsilon_{reco} \cdot \epsilon_{trg}$		4.6%
$A \cdot \epsilon_{reco} \cdot \epsilon_{trg} \cdot \epsilon_{ID}$		4.2%
μ ID	100%	
γ ID	82%	
$A \cdot \epsilon_{reco} \cdot \epsilon_{ID} \cdot \epsilon_{trg} \cdot \epsilon_{veto}$		4.1%
μ veto	100%	
γ veto	97%	

Table 8.7: Acceptance uncertainties linked to muon, photon and missing transverse energy scales. The relative energy scale variation, and the corresponding relative shifts in the acceptance are listed.

	variation	Δ^+A/A	Δ^-A/A
MET	± 10%	-4.6%	6.0%
Muon	± 2%	-0.6%	1.1%
Photon	± 2%	-2.9%	2.9%
Total		-5.5%	6.8%

8.4.2 Results

By taking into account the values and the uncertainties reported in Table 8.8, the following estimation of the Wγ production cross section was obtained:

$$\sigma(pp \to W\gamma) \cdot BR(W \to \mu\nu) = 26.0 \pm 5.2(\text{stat}) \pm 4.8(\text{syst}) \text{ pb} \tag{8.4}$$

The input MC cross-section of 18.3 pb was reproduced within uncertainties. The uncertainty linked to the neutral meson background subtraction was estimated to be of the order of 15%, while the one due to the knowledge of the energy scales to be of the order of 6–7%. Finally, an additional uncertainty of approximately 10% on the measurement of the integrated luminosity was assumed.

Table 8.8: Summary of all quantities used in the cross-section measurement and their uncertainties.

Quantity	value	uncertainty	
N(isolated γ)	60	$\pm 10(\text{stat}) \pm 9(\text{syst})$	Sections 8.3.1 and 8.3.3
N(Zγ)	7	$\pm 2.6(\text{syst})$	Section 8.3.2
$A \cdot \varepsilon$	0.041	$\pm 0.002(\text{stat}) \pm 0.003(\text{syst})$	Section 8.4.1
$\int \mathcal{L} \cdot dt$	50 pb^{-1}	±10%	

8.4.3 Extrapolation to $\sqrt{s} = 7$ TeV

The results obtained for proton-proton collisions at $\sqrt{s} = 10$ TeV were scaled to 7 TeV taking into account the cross sections evolution [24]. Both the signal and the main background cross sections are

reduced to 2/3 going from $\sqrt{s} = 10$ TeV to 7 TeV. About 20% more luminosity was thus estimated to be needed at 7 TeV in order to have the same accuracy in the cross section measurement.

8.5 Conclusions

The sensitivity of the CMS detector to W($\mu\nu$)γ final states in pp collision was studied using simulated data. Signal and background processes were simulated using shower MC. A full simulation of the detector response was performed, assuming start-up conditions for detector calibration and alignment. The event rate for signal and background was normalised to calculations with NLO precisions in pQCD.

Photon and muon candidates were selected with transverse momenta above 20 GeV. Muons were accepted in the region $|\eta| < 2.1$ and photons in the region $|\eta| < 1.45$. For a centre of mass energy of 10 TeV and an integrated luminosity of 50 pb^{-1}, the observation of approximately 100 Wγ candidate events in the accepted kinematic region, with a signal-to-background ratio $S/B \sim 1$ was predicted. The dominant background is expected to come from neutral mesons produced in W+jet events.

A data-driven estimate of the dominant background exploiting the different shape of electromagnetic energy deposits between signal and background was employed.

The main sources of systematic uncertainty for the measurement were studies. The largest effect, leading to an uncertainty of $\sim 15\%$ on the cross section measurement, was found to be due to the knowledge of the background.

Given these performances, corroborated by the results obtained on collision data, an integrated luminosity of 50 (60) pb^{-1} at a centre of mass energy of 10 (7) TeV was estimated to be sufficient to measure the Wγ production cross section in pp collisions with a precision $\sim 20\%$. Such precision is comparable to the ones presently obtained at other hadron collider experiments and could be improved upon taking into account the experience that meanwhile has been gained with data.

With the technique presented, the measurement of the photon E_{T} spectrum is also possible. This will allow the triple gauge W-photon coupling to be constrained and thus the gauge sector of the SM Lagrangian to be tested at the LHC.

Chapter 9

Conclusions

This manuscript presented some of the aspects of the Physics of proton-proton collisions at the LHC, focusing on the measurement of final states with isolated photons, which represent an important part of the LHC research program.
The scope of the thesis covered several facets of the study of Physics processes at hadron colliders. The work aimed at the early phases of the LHC Physics program, devoted to the measurement of known processes, which is an essential prerequisite to the search for new phenomena.

An important contribution to the detector commissioning process was given. In particular in what concerns the trigger and data acquisition system of the CMS ECAL, through the development of the online software driving the system and through the measurement of the trigger performance with cosmic ray events. Furthermore, a contribution to the development of photon reconstruction and identification techniques was given. Photon identification techniques for early analyses were studied and their performance assessed using the first pp collision data delivered by the LHC. Using data acquired during the 2008 CMS cosmic ray run, the electromagnetic trigger was probed using radiative losses from cosmic ray particles. The energy and position resolution of the L1 e/γ trigger algorithms were found to be consistent with expectations. The algorithm was shown to be fully efficient and the E_{T} turn-on points for different trigger thresholds were measured.
Photon reconstruction and identification algorithms were tested using collision data after the LHC startup. Using the first proton-proton collision data collected with the CMS detector in 2009 at a centre of mass energy of 900 GeV, the behaviour of such algorithms was assesed in minimum bias collisions and compared to the predictions from Monte Carlo (MC) simulations. Several kinematic and identification variables were considered and good agreement between data and MC observed, leading to the conclusion that the response of the detector is well modelled in the simulation and that the algorithms designed and optimised in the simulation show a behaviour consistent with what is expected.
Using the first 7 TeV LHC data delivered to CMS, a deeper study of the reconstruction and identification algorithms was performed. Basic reconstruction quantities for photons were compared to the Monte Carlo simulation and a good agreement was observed.
Using the photon identification selection developed as part of this thesis, it was possible to enrich the sample in isolated photons and suppress the background due to jet production. It was then possible to verify that the photon reconstruction and identification performances are similar to those expected from simulation.

The first measurement of the cross section for the production of isolated prompt photons with $21 < E_{\mathrm{T}} < 300$ GeV and $|\eta| < 1.45$ in pp collisions at $\sqrt{s} = 7$ TeV was presented.
The measurement of isolated photon production in pp collisions provides a test of perturbative quantum chromodynamics (pQCD) and the possibility to constrain the parton distribution functions (PDF). Such measurements complement deep inelastic scattering, Drell-Yan pair production and jet production measurements. A significantly higher centre-of-mass energy than at the Tevatron collider allowed the exploration of new kinematic regions in the hard-interaction processes, providing additional constraints to pQCD and the PDFs.
The cross section was measured, differentially in the photon E_{T}, using 2.9 ± 0.3 pb^{-1} of data delivered by the LHC during the 2010 pp run and it relied on a small set of variables that are easy to model already at the LHC start-up.
A data-driven technique was employed to separate, on a statistical basis, the isolated photon signal from the background due to neutral meson decays. The difference in the topology of signal and background electromagnetic energy deposits was exploited for this purpose.
The efficiency of the selection criteria was also studied on data using electron candidates from Z decays. An extensive study of the systematic effects was performed. The dominant effect affecting the observation was found to be linked to the modeling of the background which was otherwise derived in a fully data-driven way.
The precision of the measurement reached a level comparable to those of previous hadron collider exper-

iments. It was performed in the kinematic regime $0.006 < x_{\mathrm{T}} < 0.086$, probing a previously unexplored region at low x_{T}, and agrees with NLO pQCD predictions in the whole x_{T} range. This measurement established a benchmark for photon identification and background estimation, and constrained the rate of one of the background processes affecting searches for new physics involving photons.
The analysis presented here, represents the first measurement involving isolated high-energy final state photons performed at the LHC. The results were approved by the CMS collaboration for public release and a scientific communication was submitted for publication on the journal *Physical Review Letters*.

The sensitivity of the CMS detector to $\mathrm{W}(\mu\nu)\gamma$ final states in pp collisions was studied using simulated data. The study of this final state provides information on the structure of the SM gauge group and in particular on the coupling between the photon and W fields.
Deviations from the SM prediction may be found looking at observables sensitive to the high energy limit of the $\mathrm{W}\gamma$ scattering amplitudes. The most easily accessible of such observables is the photon E_{T} spectrum. In the presence of anomalous triple gauge coupling (aTGC), an enhancement in the production cross section with respect to the SM predictions is expected in the high E_{T} region of the spectrum.
Signal and background processes were simulated using shower MC. A full simulation of the detector response was performed, assuming start-up conditions for detector calibration and alignment. The event rate for signal and background was normalised to calculations with NLO precisions in pQCD.
Photon and muon candidates were selected with transverse momenta above 20 GeV/c. Muons were accepted in the region $|\eta| < 2.1$ and photons in the region $|\eta| < 1.45$. For a centre of mass energy of 10 TeV and an integrated luminosity of 50 pb^{-1}, the observation of approximately 100 $\mathrm{W}\gamma$ candidate events in the accepted kinematic region, with a signal-to-background ratio $S/B \sim 1$ was predicted. The dominant background is expected to come from neutral mesons produced in W+jet events.
A technique similar to the one used for the inclusive photon cross section measurement was used to estimate the dominant background in the simulated sample.
The main sources of systematic uncertainty for the measurement were studied. The largest effect was found to be due to the knowledge of the background.
Given these performances, corroborated by the results obtained on collision data, an integrated luminosity of 50 (60) pb^{-1} at a centre of mass energy of 10 (7) TeV was estimated to be sufficient to measure the $\mathrm{W}\gamma$ production cross section in pp collisions with a precision of $\sim 20\%$. Such precision is comparable to the ones presently obtained at other hadron collider experiments and could be improved upon taking into account the experience that meanwhile has been gained with data.
With the technique presented, the measurement of the photon E_{T} spectrum is also possible. This will allow the triple gauge W-photon coupling to be constrained and thus the gauge sector of the SM Lagrangian to be tested at the LHC.

This manuscript aimed at describing the original contribution of the author to the preparation of the CMS experiment and to analysis of the first collision data. All the elements needed to evaluate whether the work described here represents a significant contribution to the study of Physics at the TeV scale was presented.
Some of the results obtained will certainly be useful for laying solid basis to the search for new phenomena at the LHC. Others set the path to test some the SM predictions in pp collisions.
Judging whether the contribution can be qualified as "interesting" or "significant" is left to the reader. From his side, the author of the manuscript thinks that the contribution given is certainly of limited scope, compared to the vast research program of the LHC. On the other hand, the latter is a large collective effort made by several individual contributions, which only taken together will allow to gain a significant insight on the behaviour of matter at the newly probed energy frontier.

> Tapestries are made by many artisans working together. The contribution of separate workers cannot be discerned in the complete work and loose and false threads have been covered over.
>
> S.L.Glashow

Bibliography

[1] P. Musella et al., "The CMS Electromagnetic Calorimeter Data Acquisition System at the 2006 Test Beam",. Presented at 15th IEEE Real Time Conference 2007 (RT 07), Batavia, Illinois, 29 Apr - 4 May 2007.

[2] P. Musella et al., "The CMS ECAL database services for detector control and monitoring", *J. Phys. Conf. Ser.* **219** (2010) 022016. doi:10.1088/1742-6596/219/2/022016.

[3] P. Musella, "The ECAL online software in the commissioning of the CMS detector", *Nucl. Instr. Meth.* **A617** (2010), no. 1-3, 303 – 305. Proceedings of the 11th Pisa Meeting on Advanced Detectors. doi:10.1016/j.nima.2009.07.102.

[4] CMS Collaboration, "Performance of the CMS Level-1 Trigger during Commissioning with Cosmic Ray Muons", *JINST* **5** (2010) T03002, arXiv:0911.5422. doi:10.1088/1748-0221/5/03/T03002.

[5] CMS Collaboration, P. Musella et al., "Electromagnetic physics objects commissioning with first LHC data". CMS-PAS-EGM-10-001, 2010.

[6] CMS Collaboration, P. Musella et al., "Photon reconstruction and identification at $\sqrt{s} = 7$ TeV". CMS-PAS-EGM-10-005, 2010.

[7] P. Musella, "Performance of the CMS Electromagnetic Calorimeter and first results on electromagnetic physics objects". Presented at the Physics at the LHC Conference 2010, 7-12 June 2010, DESY, Hamburg (Germany).

[8] CMS Collaboration, "Measurement of the Isolated Prompt Photon Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV.", arXiv:1012.0799. CERN-PH-EP/2010-053.

[9] M. Peskin and S. D., "An Introduction to Quantum Field Theory". Westview, 1995.

[10] Particle Data Group Collaboration, "Review of particle physics", *Journal of Physics* **G37** (2010). doi:10.1016/j.physletb.2008.07.018.

[11] ALEPH, CDF, D0 et al., "Precision Electroweak Measurements and Constraints on the Standard Model", arXiv:0811.4682.

[12] B. Kayser, F. Perrier, and F. Gibrat-Debua, "The physics of massive neutrinos". World Scientific, 1989.

[13] D. H. Perkins, "Particle Astrophysics; 2nd ed.". Oxford Univ. Press, Oxford, 2009.

[14] T. Hambye and K. Riesselmann, "SM Higgs mass bounds from theory", arXiv:hep-ph/9708416.

[15] P. Langacker, "The standard model and beyond". High Energy Physics, Cosmology and Gravitation. Taylor and Francis, Boca Raton, FL, 2010.

[16] D. J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories. I", *Phys. Rev.* **D8** (1973) 3633–3652. doi:10.1103/PhysRevD.8.3633.

D. J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories. II", *Phys. Rev.* **D9** (1974) 980–993. doi:10.1103/PhysRevD.9.980.

[17] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

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

[19] T. Gleisberg et al., "Event generation with SHERPA 1.1", *JHEP* **02** (2009) 007, arXiv:0811.4622. doi:10.1088/1126-6708/2009/02/007.

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

P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, arXiv:hep-ph/0409146. doi:10.1088/1126-6708/2004/11/040.

[21] http://projects.hepforge.org/mstwpdf.

[22] L. N. Lipatov, "The parton model and perturbation theory", *Sov. J. Nucl. Phys.* **20** (1975) 94–102.

V. N. Gribov and L. N. Lipatov, "e+ e- pair annihilation and deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 675–684.

G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys.* **B126** (1977) 298. doi:10.1016/0550-3213(77)90384-4.

Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics", *Sov. Phys. JETP* **46** (1977) 641–653.

[23] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", *Phys. Rev.* **D78** (2008) 013004, arXiv:0802.0007. doi:10.1103/PhysRevD.78.013004.

[24] A. D. Martin, W. J. Stirling, R. S. Thorne et al., "Parton distributions for the LHC", *Eur. Phys. J.* **C63** (2009) 189–285, arXiv:0901.0002. doi:10.1140/epjc/s10052-009-1072-5.

[25] H1 and ZEUS Collaboration, "Determination of the proton parton density functions at HERA", *PoS* **EPS-HEP2009** (2009) 310.

[26] CMS Collaboration, "Measurement of the Underlying Event Activity at the LHC with $\sqrt{s}$ =7 TeV". CMS-PAS-QCD-10-010, 2010.

[27] S. Catani et al., "QCD", arXiv:hep-ph/0005025.

[28] S. Catani, M. Fontannaz, J. P. Guillet et al., "Cross-section of isolated prompt photons in hadron hadron collisions", *JHEP* **05** (2002) 028, arXiv:hep-ph/0204023.

[29] J. P. G. S. Catani, M. Fontannaz and E. Pilon, "Cross-section of isolated prompt photons in hadron hadron collision", *JHEP* **0205** (2002) 028.

[30] H.-L. Lai et al., "New parton distributions for collider physics", arXiv:1007.2241.

[31] L. Bourhis, M. Fontannaz, and J. P. Guillet *Eur. Phys. J.* **C2** (1998) 529.

[32] ALEPH Collaboration, "Inclusive production of neutral vector mesons in hadronic Z decays", *Z. Phys.* **C69** (1996) 379–392. doi:10.1007/s002880050038.

L. Bourhis, M. Fontannaz, and J. P. Guillet, "Quark and gluon fragmentation functions into photons", *Eur. Phys. J.* **C2** (1998) 529–537, arXiv:hep-ph/9704447. doi:10.1007/s100520050158.

[33] P. Aurenche et al., "A critical phenomenological study of inclusive photon production in hadronic collisions", *Eur. Phys. J.* **C9** (1999) 107–119, arXiv:hep-ph/9811382. doi:10.1007/s100529900018.

[34] E704 Collaboration, "Measurement of single spin asymmetry for direct photon production in p p collisions at 200-GeV/c", *Phys. Lett.* **B345** (1995) 569–575. doi:10.1016/0370-2693(94)01695-9.

[35] P. Aurenche, M. Fontannaz, J.-P. Guillet et al., "A New critical study of photon production in hadronic collisions", *Phys. Rev.* **D73** (2006) 094007, arXiv:hep-ph/0602133. doi:10.1103/PhysRevD.73.094007.

[36] D0 Collaboration, "Measurement of the differential cross-section for the production of an isolated photon with associated jet in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$-TeV", *Phys. Lett.* **B666** (2008) 435–445, arXiv:0804.1107. doi:10.1016/j.physletb.2008.06.076.

CDF Collaboration, "Measurement of the Inclusive Isolated Prompt Photon Cross Section in ppbar Collisions at sqrts = 1.96 TeV using the CDF Detector", *Phys. Rev.* **D80** (2009) 111106, arXiv:0910.3623. doi:10.1103/PhysRevD.80.111106.

[37] R. Ichou and D. d'Enterria, "Sensitivity of isolated photon production at TeV hadron colliders to the gluon distribution in the proton", *Phys. Rev.* **D82** (2010) 014015, arXiv:1005.4529. doi:10.1103/PhysRevD.82.014015.

[38] DELPHI Collaboration, "Study of W boson polarisations and Triple Gauge boson Couplings in the reaction $e^+e^- \to W^+W^-$ at LEP 2", *Eur. Phys. J.* **C54** (2008) 345–364, arXiv:0801.1235. doi:10.1140/epjc/s10052-008-0528-3.

ALEPH Collaboration, "Improved measurement of the triple gauge-boson couplings gamma W W and Z W W in e+ e- collisions", *Phys. Lett.* **B614** (2005) 7–26. doi:10.1016/j.physletb.2005.03.058.

OPAL Collaboration, "Measurement of charged current triple gauge boson couplings using W pairs at LEP", *Eur. Phys. J.* **C33** (2004) 463–476, arXiv:hep-ex/0308067. doi:10.1140/epjc/s2003-01524-6.

L3 Collaboration, "Measurement of triple gauge boson couplings of the W boson at LEP", *Phys. Lett.* **B586** (2004) 151–166, arXiv:hep-ex/0402036. doi:10.1016/j.physletb.2004.02.045.

[39] K. Hagiwara, R. D. Peccei, D. Zeppenfeld et al., "Probing the Weak Boson Sector in e+ e- → W+ W-", *Nucl. Phys.* **B282** (1987) 253. doi:10.1016/0550-3213(87)90685-7.

[40] U. Baur and E. L. Berger, "Probing the $WW\gamma$ Vertex at the Tevatron Collider", *Phys. Rev.* **D41** (1990) 1476. doi:10.1103/PhysRevD.41.1476.

[41] U. Baur, T. Han, and J. Ohnemus, "QCD corrections to hadronic $W\gamma$ production with nonstandard $WW\gamma$ couplings", *Phys. Rev.* **D48** (1993) 5140–5161, arXiv:hep-ph/9305314. doi:10.1103/PhysRevD.48.5140.

[42] U. Baur and D. Zeppenfeld, "Unitarity Constraints on the Electroweak Three Vector Boson Vertices", *Phys. Lett.* **B201** (1988) 383. doi:10.1016/0370-2693(88)91160-4.

[43] D0 Collaboration, "Combined measurements of anomalous charged trilinear gauge-boson couplings from diboson production in p-pbar collisions at sqrt(s)=1.96 TeV", arXiv:0907.4952.

CDF Collaboration, "Limits on Anomalous Triple Gauge Couplings in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.96-TeV", *Phys. Rev.* **D76** (2007) 111103, arXiv:0705.2247. doi:10.1103/PhysRevD.76.111103.

[44] CLEO Collaboration, "First measurement of the rate for the inclusive radiative penguin decay b → s gamma", *Phys. Rev. Lett.* **74** (1995) 2885–2889. doi:10.1103/PhysRevLett.74.2885.

[45] UA2 Collaboration, "Direct measurement of the W - gamma coupling at the CERN anti-p p collider", *Phys. Lett.* **B277** (1992) 194–202. doi:10.1016/0370-2693(92)90979-E.

[46] http://lepewwg.web.cern.ch/LEPEWWG/lepww/tgc/.

[47] B. J. Huckvale, "Studying anomalous WWgamma couplings and developing the Global Calorimeter Trigger control system for the CMS experiment". PhD thesis, University of Bristol, 2008. CMS TS-2010/030.

[48] http://accel-general.web.cern.ch/accel-general.

[49] L. Evans, (ed.) and P. Bryant, (ed.), "LHC Machine", *JINST* **3** (2008) S08001. doi:10.1088/1748-0221/3/08/S08001.

[50] E. J. N. Wilson et al., "An introduction to particle accelerators". Oxford University Press, 2001.

[51] K. V., "Commissioning of the LHC with beam", in *Proceedings of HB2010, Morschach, Switzerland*. 2010.

[52] CMS Collaboration, "Measurement of CMS Luminosity". CMS-PAS-EWK-10-004, 2010.

[53] https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults2010.

[54] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[55] CMS Collaboration, "Performance of CMS Muon Reconstruction in Cosmic-Ray Events", *J. Instrum.* **5** (Nov, 2009) T03022 . 47 p.

[56] CMS Collaboration, "Measurement of the W and Z inclusive production cross sections at sqrts=7 TeV with the CMS experiment at the LHC". CMS-PAS-EWK-10-002, 2010.

[57] CMS Collaboration, "Electromagnetic calorimeter calibration with 7 TeV data". CMS-PAS-EGM-10-003, 2010.

[58] CMS Collaboration, "Time Reconstruction and Performance of the CMS Electromagnetic Calorimeter", *J. Instrum.* **5** (Nov, 2009) T03011. 27 p.

[59] CMS Collaboration, "Electromagnetic calorimeter commissioning and rst results with 7 TeV data". CMS NOTE-2010/012, 2010.

[60] R. Wigmans, "Calorimetry. Energy Measurement in Particle Physics". Oxford Science Publications, Oxford, 2000.

[61] S. Abdullin et al., "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c", *Eur. Phys. J.* **C60** (2009) 359–373. doi:10.1140/epjc/s10052-009-0959-5.

[62] CMS Collaboration, "Jet Performance in pp Collisions at 7 TeV". CMS-PAS-JME-10-003, 2010.

[63] CMS Collaboration, "Studies of Tracker Material". CMS-PAS-TRK-10-003.

[64] CMS Collaboration, "Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays", *J. Instrum.* **5** (Oct, 2009) T03009. 41 p.

[65] CMS Collaboration, "Tracking and Vertexing Results from First Collisions". CMS-PAS-TRK-10-001, 2010.

[66] CMS Collaboration, "CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems",. CERN-LHCC-2000-038.

[67] CMS Trigger and Data Acquisition Group Collaboration, "The CMS high level trigger", *Eur. Phys. J.* **C46** (2006) 605–667, arXiv:hep-ex/0512077. doi:10.1140/epjc/s2006-02495-8.

[68] http://hsi.web.cern.ch/HSI/s-link/.

[69] CMS Collaboration, "CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger",. CERN-LHCC-2002-026.

[70] C. Eck et al., "LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)". Technical Design Report LCG. CERN, Geneva, 2005.

[71] R. Alemany et al., "Overview of the ECAL off-detector electronics of the CMS experiment", *IEEE Trans. Nucl. Sci.* **52** (2005) 1918–1924. doi:10.1109/TNS.2005.856596.

[72] N. Almeida et al., "Data filtering in the readout of the CMS electromagnetic calorimeter", *JINST* **3** (2008) P02011. doi:10.1088/1748-0221/3/02/P02011.

[73] M. Raymond, J. Crooks, M. French et al., "The MGPA electromagnetic calorimeter readout chip for CMS",. CERN Document Server record 712053.

[74] G. Minderico, C. Fachada, I. L. Chan et al., "A CMOS low power, quad channel, 12 bit, 40MS/s pipelined ADC for applications in particle physics calorimetry",. CERN Document Server record 712054.

[75] M. Hansen, "The new readout architecture for the CMS ECAL",. CERN Document Server record 712052.

[76] IBM, "Token-ring network: Architecture reference". Research reports // IBM. 1987.

[77] J. E. Varela, "CMS L1 Trigger Control System",. CMS NOTE-2002/033.

[78] N. Almeida et al., "Calorimeter trigger synchronization in the CMS experiment", *Nucl. Instrum. Meth.* **A568** (2006) 634–641. doi:10.1016/j.nima.2006.08.032.

A. David, N. Almeida, J. da Silva et al., "Test beam operation of the CMS calorimeter trigger synchronization boards", *JINST* **3** (2008) P05004. doi:10.1088/1748-0221/3/05/P05004.

[79] J. C. Silva et al., "Design of a data concentrator card for the CMS electromagnetic calorimeter readout",. Prepared for 7th Workshop on Electronics for LHC Experiments, Stockholm, Sweden, 10-14 Sep 2001.

[80] J. Gutleber, S. Murray, and L. Orsini, "Towards a homogeneous architecture for high-energy physics data acquisition systems", *Comput. Phys. Commun.* **153** (2003) 155–163. doi:10.1016/S0010-4655(03)00161-9.

[81] G. Bauer et al., "The run control and monitoring system of the CMS experiment", *PoS* **ACAT** (2007) 026. doi:10.1088/1742-6596/119/2/022010.

[82] I20 Special Interest Group Collaboration, "Intelligent I/O (I2O) architecture specification V2.0", 1999.

[83] H. Sakulin, "DAQ Operations and Plans". Run Coordination Workshop, 2-3 November 2010, U. Zurich.

[84] CMS Collaboration, "Commissioning of the CMS Experiment and the Cosmic Run at Four Tesla", *J. Instrum.* **5** (Nov, 2009) T03001. 37 p.

[85] CMS Collaboration, "Performance and Operation of the CMS Electromagnetic Calorimeter", *JINST* **5** (2010) T03010, arXiv:0910.3423. doi:10.1088/1748-0221/5/03/T03010.

[86] CMS Collaboration, "Time Reconstruction and Performance of the CMS Electromagnetic Calorimeter", *JINST* **5** (2010) T03011, arXiv:0911.4044. doi:10.1088/1748-0221/5/03/T03011.

[87] C. Broutin, "Level-1 Electron and Photon Trigger Commissioning and Performance on 7 TeV data",.

[88] CERN, "GEANT, detector description and simulation tool", 1994. Program library long writeup W5013.

[89] CMS Collaboration, "Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 7$ TeV. oai:cds.cern.ch:1266262", *Phys. Rev. Lett.* **105** (May, 2010) 022002. 26 p.

[90] M. Anderson et al., "Review of clustering algorithms and energy corrections in ECAL",. CMS IN-2010/008.

E. Meschi et al., "Electron Reconstruction in the CMS Electromagnetic Calorimeter",. CMS IN-2001/034.

[91] P. Meridiani and C. Rovelli, "Clusters containment corrections using ECAL position measurements at the 2006 H4 testbeam",. CMS DN-2008/007.

M. Takahashi, "Rapidity Dependence of the Shower Containment in the CMS ECAL",. CMS IN-2004/039.

[92] R. Field, "Studying the Underlying Event at CDF and the LHC", arXiv:1003.4220. In the Proceedings of the First International Workshop on Multiple Partonic Interactions at the LHC (MPI08).

[93] CMS Collaboration, "First Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} =$ 0.9 TeV. oai:cds.cern.ch:1271104",. Submitted to the European Physical Journal C.

[94] J. Huston et al., "A Global QCD study of direct photon production", *Phys. Rev.* **D51** (1995) 6139–6145, arXiv:hep-ph/9501230. doi:10.1103/PhysRevD.51.6139.

A. D. Martin, R. G. Roberts, W. J. Stirling et al., "Parton distributions: A New global analysis", *Eur. Phys. J.* **C4** (1998) 463–496, arXiv:hep-ph/9803445. doi:10.1007/s100520050220.

[95] CDF Collaboration, "Measurement of the Inclusive Isolated Prompt Photon Cross Section in ppbar Collisions at sqrts = 1.96 TeV using the CDF Detector", *Phys. Rev.* **D80** (2009) 111106, arXiv:0910.3623. doi:10.1103/PhysRevD.80.111106.

D0 Collaboration, "Measurement of the isolated photon cross section in $p\bar{p}$ collisions at $\sqrt{s} =$ 1.96-TeV", *Phys. Lett.* **B639** (2006) 151–158, arXiv:hep-ex/0511054. doi:10.1016/j.physletb.2006.04.048.

[96] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions". CMS-PAS-TRK-10-005, 2010.

[97] CMS Collaboration, "Test of the energy scale of the CMS ECAL with $Z \to ee$ decays",. Document in preparation.

[98] T. Kolberg et al., "Identification of isolated converted photons in CMS 7 TeV data". CMS AN-2010/329.

[99] CMS Collaboration, "Studies of Tracker Material",. CMS-PAS-TRK-10-003.

[100] CMS Collaboration, "Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at sqrt(s) = 7 TeV". CMS-PAS-QCD-10-006.

CMS Collaboration, "Single-Punpublished Response in the CMS Calorimeters". CMS-PAS-JME-10-008, 2010.

[101] A. D. Martin, W. J. Stirling, R. S. Thorne et al., "Parton distributions for the LHC", *Eur. Phys. J.* **C63** (2009) 189–285, arXiv:0901.0002. doi:10.1140/epjc/s10052-009-1072-5.

[102] R. D. Ball et al., "A first unbiased global NLO determination of parton distributions and their uncertainties", *Nucl. Phys.* **B838** (2010) 136–206, arXiv:1002.4407. doi:10.1016/j.nuclphysb.2010.05.008.

[103] PDF4LHC Working Group Collaboration, "The PDF4LHC Working Group Interim Report". The PDF4LHC recommendation, NLO prescription.

[104] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **07** (2002) 012, arXiv:hep-ph/0201195.

[105] V. Chetluru et al., "Prediction of Isolated Photon Cross Section at sqrt(s) = 7 TeV from JETPHOX and Correction of Non-perturbative Effects". CMS AN-2010/272.

[106] P. Z. Skands, "The Perugia Tunes", arXiv:0905.3418.

[107] J. Campbell and E. R.K., "Calculation of the Zbb and other backgrounds to a ZH signal at the Tevatron", *Phys. Rev. D* **62** (2000) 114012.

[108] J. Pumplin et al., "Uncertainties of predictions from parton distribution functions. 2. The Hessian method", *Phys. Rev.* **D65** (2001) 014013, arXiv:hep-ph/0101032. doi:10.1103/PhysRevD.65.014013.

[109] G. Abbiendi et al., "Muon Reconstruction in CMS". CMS AN-2008/097, 2008.

[110] G. Landsberg et al., "Missing ET performance". CMS AN-2007/041, 2007.

[111] G. Landsberg et al., "MET Reconstruction, Performance, and Validation". CMS AN-2008/089, 2008.

[112] CMS Collaboration, "Missing Transverse Energy Performance in Minimum-Bias and Jet Events from Proton-Proton Collisions at $\sqrt{s} = 7$ TeV". CMS-PAS-JME-10-004, 2010.

[113] CMS Collaboration, "Towards a measurement of W and Z cross sections into muons in pp collisions at rts=10 TeV". CMS-PAS-EWK-09-001.