

Studies at the CMS Experiment on

Silicon Microstrip Module Defects

and on

Photon Identification in Semileptonic $t\bar{t}$ -Events

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der
RWTH-Aachen University zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Physiker

Thomas Hermanns

aus Erkelenz

Berichter: Universitätsprofessor Dr. Joachim Mnich
Universitätsprofessor Dr. Achim Stahl

Tag der mündlichen Prüfung: 15. Dezember 2008

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar.



Abstract

The present thesis has been accomplished in the framework of the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). It is divided into two parts. The first one describes the studies that have been performed by a centralized Module Debug Centre to investigate the silicon microstrip detector modules for the CMS inner tracker endcaps that have not passed the qualification procedure during the mass production. A scheme of seven global defect classes has been introduced. The signatures of each class as well as the methods to sort a module into a certain class are illustrated. In total, 225 modules have been examined and 126 of them could be repaired. No systematic source of defects has been observed for the entire production phase.

The second part deals with the identification of semileptonic $t\bar{t}$ -events (electron and muon channel) containing photons that are radiated off top-quarks. A signature based Monte Carlo analysis reveals that the separation of the $t\bar{t}$ -events from the non- $t\bar{t}$ background (W + Jets, Z + Jets, di-boson, γ + Jets, and QCD-multijet events) is possible. The resulting photon energy spectra are dominated by photons radiated off top-quarks. The situation changes considerably if furthermore an inclusive $t\bar{t}$ + Jets data-set is included. The semileptonic fraction has the same event signature as the signal events. Due to the irreducible background from the decay $\pi^0 \rightarrow \gamma\gamma$ possibly being part of the hadronic structure of these events the energy spectra are now mainly formed by the latter photons. A signal-to-background ratio of 1/46 (electron channel) and 1/31 (muon channel) can be achieved.

If the $t\bar{t}$ + Jets events are not considered as background but replace the signal data-set because of a 10% to 15% admixture of photons that are radiated off top-quarks in these events, a first study on the distinction between a top-quark with an electromagnetic charge of $2/3e$ and $-4/3e$ has been implemented. It is shown that the potential exists to directly measure the value of this electromagnetic coupling parameter of the top-quark with a few 10 fb^{-1} of LHC collision data at the CMS experiment.

Kurzfassung

Die vorliegende Arbeit ist im Rahmen des Compact Muon Solenoid (CMS) Experiments angefertigt worden, das am Large Hadron Collider (LHC) installiert ist. Sie ist in zwei Teile aufgeteilt. Der erste Teil behandelt Studien, die in einem zentralisierten Module Debug Centre zur Untersuchung von Siliziumstreifendetektormodulen für die Endkappen der inneren Spurdetektors durchgeführt worden sind, die in der Massenproduktion als defekt aufgefallen sind. Eine Raster von sieben allgemeinen Defektklassen ist dazu eingeführt worden. Die speziellen Eigenschaften einer jeden Klasse werden beschrieben und die Kriterien, nach denen die Module in die Klassen einsortiert werden, dargelegt. Insgesamt sind 225 Module untersucht worden, von denen 126 repariert werden konnten. Systematische Fehlerquellen konnten während der gesamten Massenproduktionsphase nicht beobachtet werden.

Der zweite Teil der Arbeit beschäftigt sich mit der Identifikation von semileptonischen $t\bar{t}$ -Ereignissen (Elektron- und Myonkanal), in denen ein Photon vom top-Quark abgestrahlt wird. Eine schnittbasierte Monte Carlo Analyse zeigt, dass die Trennung der $t\bar{t}$ -Ereignisse vom nicht- $t\bar{t}$ -Untergrund (W + Jets, Z + Jets, Diboson, γ + Jets, und QCD-Multijet Ereignisse) gelingt. Das resultierende Photonspektrum wird durch Photonen bestimmt, die von top-Quarks abgestrahlt werden. Die Verhältnisse ändern sich beträchtlich, wenn zwingend ein inklusiver $t\bar{t}$ + Jets Datensatz als Untergrund hinzugenommen wird. Der semileptonische Anteil zeigt dieselbe Ereignissignatur wie die Signalereignisse. Aufgrund des irreduziblen Untergrundes durch den Zerfall $\pi^0 \rightarrow \gamma\gamma$, die in dem hadronischen Anteil dieser Ereignisse vorkommen können, werden die Spektren nun durch diese Photonen bestimmt. Ein Signal-zu-Untergrundverhältnis von 1/46 (Elektronkanal) und 1/31 (Myonkanal) kann nun erreicht werden.

Wenn nun die inklusiven $t\bar{t}$ + Jets Ereignisse nicht als Untergrund sondern als Signaldatensatz angesehen werden, kann eine erste Studie zur Unterscheidung eines top-Quarks mit einer elektromagnetischen Ladung von $2/3e$ und $-4/3e$ durchgeführt werden, da der $t\bar{t}$ -Datensatz eine 10 bis 15%ige Beimischung von Photonen enthält, die von top-Quarks abgestrahlt werden. Es kann gezeigt werden, dass es die Möglichkeit gibt, den Wert dieses elektromagnetischen Kopplungsparameters direkt zu messen, nachdem einige 10 fb^{-1} LHC-Daten mit dem CMS Experiment aufgezeichnet worden sind.

Table of Contents

Abstract	I
Kurzfassung	III
1 Introduction	1
2 The CMS Detector at the LHC	5
2.1 The Large Hadron Collider	5
2.2 The Main Detectors at the LHC	6
2.3 The Compact Muon Solenoid Detector	7
2.3.1 Key Parameters of the CMS Detector	8
2.3.2 CMS Coordinate System	9
3 The CMS Silicon Microstrip Detector	11
3.1 The CMS Silicon Tracker	11
3.2 The CMS Silicon Sensor Layout	12
3.3 The CMS Silicon Microstrip Detector Module Layout	13
3.4 The CMS Silicon Microstrip Detector Hybrid Layout	14
3.5 The Module Production and Qualification	15
3.5.1 ARC Test System	16
3.5.2 Classification Scheme of Defective Modules	17
3.6 The CMS Silicon Microstrip Module Quality	19
4 The TEC Silicon Microstrip Module Debug Centre	21
4.1 The TEC Module Debug Centre	21
4.2 The Classification of Defects and Failures	22
4.2.1 IV-Problem	22
4.2.2 Frontend Hybrid Problems	24
4.2.3 Damage of the Carbon Support Structure	25
4.2.4 Misaligned Modules	26
4.2.5 Number of Defective Channels	27
4.2.6 Inappropriate Handling	29

4.2.7	Other Defects	30
4.3	The Final Statistics and Defect Correlations	31
4.4	The Evaluation of the Procedure	32
5	Events with Top-Quarks and Photons	35
5.1	The Photon Radiation off Top-Quarks	35
5.2	The Signal Process	36
5.3	The Monte Carlo Event Generation	38
5.4	The Event Generation and Reconstruction	44
5.5	The High-Level Objects	46
5.6	The Analysis Flow	50
6	The Preselection of $t\bar{t}\gamma$-Events	53
6.1	The Motivation for an Event Preselection	53
6.2	The Event Preselection Criteria	54
6.3	The Results of the Event Preselection	54
7	The Selection of $t\bar{t}\gamma$-Events	59
7.1	The Top-Quark Analysis Framework	59
7.2	The Lepton Identification	60
7.2.1	Lepton Classes	60
7.2.2	Identification Criteria	61
7.2.3	Results of the Lepton Identification	66
7.3	The Photon Identification	67
7.3.1	Photon Classes	67
7.3.2	Generator Matching	70
7.3.3	The Photon Object	74
7.3.4	Identification Criteria	76
7.3.5	Results of the Photon Identification	82
7.4	The Particle Correlations	86
7.4.1	Correlations between Photons and Electrons	86
7.4.2	Correlations between Photons and Track Seeds	87
7.4.3	Correlations between Photons and Jets	88
7.4.4	Correlations between Electrons and Jets	88
7.4.5	Results of the Particle Correlation Investigations	88
8	The Event Reconstruction	93
8.1	The Identification of b-Jets	93
8.2	The Jet p_T -Distributions	95
8.3	The Missing Transverse Energy	95
8.4	The Kinematic Event Solutions	96

8.5	The Correlations between Photons and Reconstructed Top-Quarks	100
8.6	The Results of the Identification of $t\bar{t}\gamma$ -Events	101
9	The Statistical Results of the $t\bar{t}\gamma$-Analysis	103
9.1	The Method of Decomposing the Total Selection Efficiency	103
9.2	The Application of the Estimated Limits to the Photon Spectra	107
10	The Photon Energy Spectra	109
10.1	The Standard Model Photon Energy Spectra	109
10.2	The Non Standard Model Photon Energy Spectra	110
10.2.1	Signal Rescaling by the Baur-Generator	111
10.3	The Limits on the Electromagnetic Charge of the Top-Quark	113
11	Summary and Concluding Remarks	121
A	Background Statistics	125
A.1	The Data-Sets	125
A.2	Tables on the Event Selection of the Background Data-Sets	127
B	Example of a CMSSW Configuration File	133
C	Mathematical Definitions	137
	List of Figures	141
	List of Tables	144
	References	145
	Acknowledgment	153

Chapter 1

Introduction

THE picture of the micro cosmos formed by elementary particle physics is described by a few basic principles with only a small number of constituents. This *Standard Model of Particle Physics* (Standard Model, SM) groups the fundamental fermions, the quarks and leptons, by means of local gauge symmetries with respect to their quantum numbers. Hereby an electromagnetic charge allows for the electromagnetic interaction, the colour charge mediates the strong interaction, and due to the weak isospin particles interact weakly. While strong interactions are described by the Quantumchromodynamics (QCD) via an $SU(3)_C$ -gauge theory [1], the electromagnetic and weak interactions are unified in an $SU(2)_L \times U(1)_Y$ -group by the Glashow-Salam-Weinberg (GSW) theory [2]. The local gauge invariance of these theories implies the presence of four types of gauge bosons in total, mediating the forces between the fermions. These bosons complete the list of presently known fundamental constituents, which are summarized in Table 1.1 and Table 1.2 giving some of its key parameters as well.

Ref. [1]

Ref. [2]

Tab. 1.1

Tab. 1.2

Family	1	2	3	Interaction
$q = \frac{2}{3} e$	$u - quark$	$c - quark$	$t - quark$	EM, weak, strong
$m [MeV]$	1.5 – 3.3	1160 – 1340	171200 ± 2100	
$q = -\frac{1}{3} e$	$d - quark$	$s - quark$	$b - quark$	
$m [MeV]$	4.5 – 6.0	80 – 130	4130 – 4370	
$q = 0$	ν_e	ν_μ	ν_τ	weak
$m [MeV]$	$< 2 \cdot 10^{-6}$	$< 2 \cdot 10^{-6}$	$< 2 \cdot 10^{-6}$	
$q = -1 e$	e^-	μ^-	τ^-	EM, weak
$m [MeV]$	0.511	105.658	1776.84	

Table 1.1: The Standard Model of Particle Physics comprises three families of quarks and leptons. The table gives information about the mass m and the electromagnetic charge q (in multiples of the positron charge e) of the particles as well as about the kind of interaction, in which they participate. The upper limits on the neutrino masses are determined by the beta-decay of tritium for ν_e and for ν_μ and ν_τ by the mass differences Δm^2 which are maximally at the order of $10^{-3} eV^2$. All data are extracted from [3].

Although the Standard Model has been checked at numerous occasions (see for instance [4] and [5]), this ordering scheme of nature still gives rise to a variety of questions, like the number of fermions, the solution to the hierachy-problem or a grand unification of the forces. The last two problems are directly connected to searches for extensions of the Standard Model. A Minimal Supersymmetric Standard

Ref. [4]

Ref. [5]

Gauge Bosons	Charge [e]	Mass [MeV]	Interaction
photon	0	0	EM
gluon	0	0	strong
W-boson	± 1	80398 ± 25	EM, weak
Z-boson	0	91188 ± 2	weak

Table 1.2: The Standard Model of Particle Physics predicts four types of gauge bosons, mediating the interactions between the fundamental fermions. The mass, the electromagnetic charge and the kind of interaction, in which a particles participates, are extracted from [3].

Model (MSSM) extension would double the number of particles by assigning a supersymmetric partner to each SM-particle, for instance. Furthermore, the question must be answered, why the fermions of the same type only differ by their mass. This immediately leads to the question about the underlying mechanism that generates masses, because the fundamental local gauge symmetries are only valid in case of massless bosons.

At first, a hint to the solution of these problems or maybe even a full answer can be given by the discovery of new particles. A possible way to find new particles is to construct a new particle accelerator providing a centre-of-mass energy beyond the limit of the current existing machines. This implies new detectors to measure properties of particles, to more precisely cross-check theoretical predictions and to increase predictive power, or determine so far unmeasured properties of known particles or to detect predicted but very rare processes in addition. For these aims one also strives for the highest achievable luminosity to generate a sufficiently large number of events for all kinds of studies.

This task of the next generation of particle accelerators and experiments will be covered by a proton-proton collider, the Large Hadron Collider (LHC), and its four main detectors. They start operation in the second half of 2008 and are expected to provide data for the next decade. One of these detectors is the Compact Muon Solenoid (CMS), which constitutes the framework of the present thesis.

Chp. 2 Thus, chapter 2 gives an overview of some key facts of the LHC. In addition, the CMS detector will be described, briefly mentioning its design goals and some selected facts about the estimated physics performance. As the silicon microstrip detector is the key part of the CMS detector for track determination a precise knowledge about the defects especially on the main constituents, the silicon microstrip detector modules, is indispensable to better model the expected performance of that sub-detector. For that reason chapter 3 reviews the design and the functional principle of the silicon tracker focusing on individual detector modules. After that the qualification means and methods of the silicon microstrip tracker modules are treated. This section is concluded by an overview about the test system and the procedure that have been used for the qualification of the modules. Finally, the overall quality of the microstrip detector modules is quoted. As a mass production cannot be established without having a strategy how to handle defective components the chapter 4 is dedicated to a description of that strategy for the CMS silicon tracker endcaps. Defect classes are introduced and it is described how the defects can be traced, how often a certain type has been observed and, if possible, cured.

Chp. 4 As the construction of the CMS detector is very close to completion and the injection of the first proton beams to the LHC is envisaged for summer 2008 the analysis of data has to be prepared. In the second part of that thesis the identification of photons in $t\bar{t}$ -events ($t\bar{t}\gamma$ -events) is described. The longterm goal of that analysis is to measure the electromagnetic couplings between the top-quark and the photon which allows to investigate the electromagnetic charge or an anomalous magnetic moment of the top-quark, for instance. After an introduction to top-quark physics at the LHC and the analysis environment in chapter 5, the event preselection in the semileptonic (lepton + jets) $t\bar{t}$ -decay channel is described in

Chp. 5

chapter 6. As leptons and photons impose a very distinct signature in the signal events their identification Chp. 6
provide strong tools to reject physical background events. This task is discussed in chapter 7, which also Chp. 7
includes a section contemplating the correlations between the characteristic objects of a $t\bar{t}\gamma$ -event. The
next step of the analysis in chapter 8 comprises a tighter selection of the single reconstructed objects Chp. 8
in an event and afterwards the combination of the reconstructed objects to the initial $t\bar{t}$ -event signature.
The quantitative results of the analysis are statistically evaluated in chapter 9. Finally, the chapter 10 Chp. 9
contains the photon energy spectra which can be derived from the present analysis. As these spectra
are influenced by the coupling of the photon to the top-quark this thesis is closed by a first investigation Chp. 10
to distinguish between two coupling scenarios of the photon to the top-quark via its electromagnetic
charge. The summarizing and concluding chapter 11 reviews both parts of this thesis and discusses Chp. 11
possible improvements and extensions of the $t\bar{t}\gamma$ -analysis.

Chapter 2

The CMS Detector at the LHC

THIS chapter presents an overview about the next proton-proton collider, the Large Hadron Collider (LHC). Collision data are recorded at four interaction points each of them instrumented with a detector for particle detection. One of them is the Compact Muon Solenoid (CMS) experiment. As this thesis has been done in the framework of this experiment, the second part of this section deals with an introduction to this detector.

2.1 The Large Hadron Collider

In 1995, it was approved that the next collider at the European laboratory for particle physics (CERN) located near Geneva, Switzerland, will be the Large Hadron Collider (LHC) [6]. This new particle Ref. [6] accelerator replaced the Large Electron Positron Collider (LEP) because it is built in the same tunnel approximately 100 m below the surface of the earth. It is designed as a proton-proton collider with a centre-of-mass energy of 14 TeV (7 TeV per proton beam). The proton bunches are accelerated in two counter-rotating beams using superconducting radio frequency cavities and synchrotron technology. Besides magnets used for beam focusing, 8.4 T-superconducting dipole-magnets bend the proton beams to a ring with about 27 km circumference. At four points the beams collide at rate of 40 MHz corresponding to a time interval of 25 ns between two consecutive collisions. At these interaction points large detector systems are constructed to record the results of the proton-proton-collisions.

It is foreseen to periodically replace the proton beams by ionized elements, such as lead or calcium, to investigate heavy ion collisions at a rate of 8 kHz. At a centre-of-mass energy of 5.5 TeV per nucleon it is expected that a different state of matter called quark-gluon plasma can be studied.

Before reaching its final design parameters the LHC starts colliding protons in several phases. After a beam commissioning phase it is planned to run at a centre-of-mass energy of 10 TeV in the first physics phase, and after the first winter shutdown to go to the design energy of 14 TeV. At the same time not only the number of bunches per beam and the number of protons per bunch is increased but also the beam focusing is improved, so that the luminosity is evolving from about $3 \cdot 10^{30} \text{ cm}^{-2} \text{ s}^{-1}$ to $2 - 4 \cdot 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$, the target value for the first three years of operation. After that time an integrated luminosity of 30 fb^{-1} is expected to be collected. It shall be increased in the following seven years to about 400 fb^{-1} when running at the design luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. Then a single beam will consist of 2808 bunches, each filled with $1.5 \cdot 10^{11}$ protons.

The number of protons per bunch interacting with their counterparts of the other beam is not restricted to a single hard interaction¹. While at the lowest luminosities multi-interactions can be neglected, at $\mathcal{L} = \mathcal{O}(10^{33} \text{ cm}^{-2} \text{ s}^{-1})$ the reduced size of the beam spot accounts for an average of 3 interactions, while at the final luminosity about 22 hard collisions take place. These collisions are referred to as *in-time pile-up events*. But not only the products of these additional interactions, but also the remnants of the interacting protons, and debris from the last or even earlier bunch crossings, the *out-of-time pile-up events*, might enter a detector. Thus, the detector is designed to resolve as precisely as possible the traversing objects, so that the signature of the hard interaction that is of particular interest can be identified.

2.2 The Main Detectors at the LHC

As described above the two counter-rotating proton beams collide at four different points along the LHC ring. A global geographical sketch can be seen in Figure 2.1.

Fig. 2.1

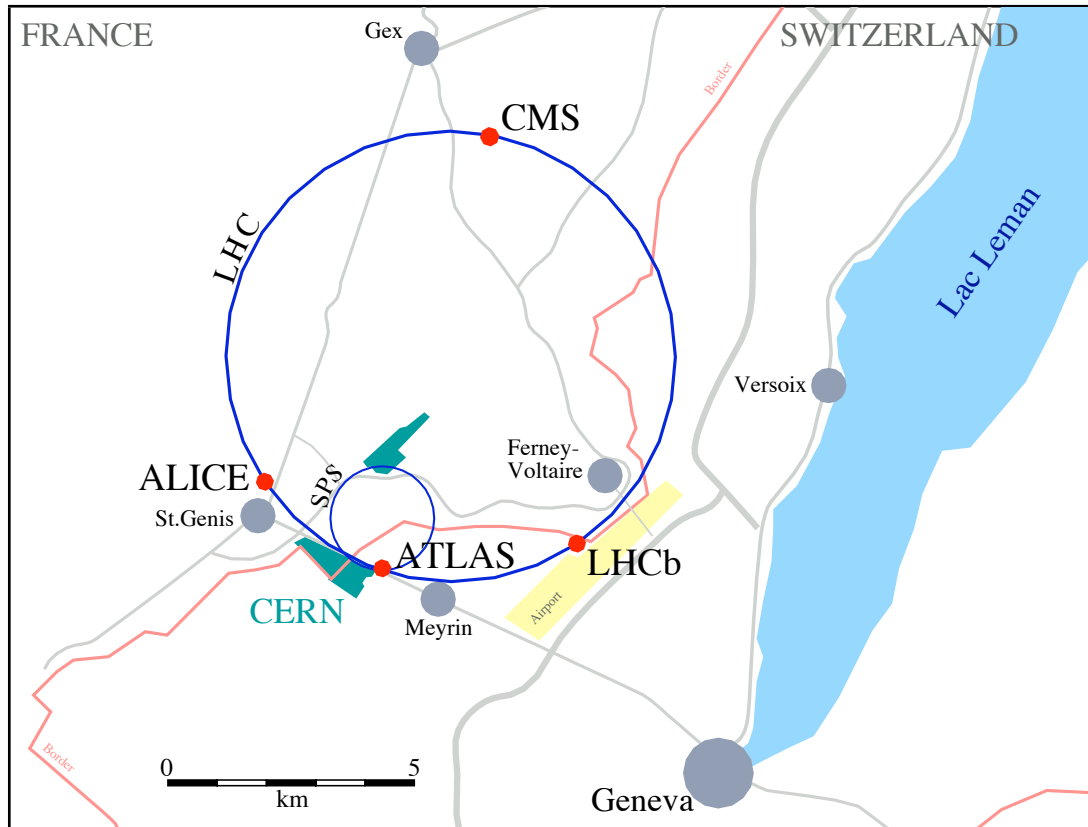


Figure 2.1: The map shows the surroundings of the European laboratory for particle physics (CERN) near Geneva. The large circle represents the positions of the LHC accelerator complex and the four small dots along the ring indicate the locations of the interaction points [7].

Each interaction point is equipped with the infrastructure for high energy experiments so that the LHC will provide beams for four main detectors: ATLAS² [9], CMS [10], LHCb [11], and ALICE³ [12]. ATLAS and CMS are designed as multi-purpose detectors focusing on the broad spectrum of physics at

¹A *hard interaction* is a collision when the momentum transfer is large enough that the two colliding protons break up and two partons interact so that new particles are created which do not occur in the initial state.

²A Toroidal LHC ApparatuS

³A Large Ion Collider Experiment

Ref. [9]

Ref. [10]

Ref. [11]

Ref. [12]

the TeV-energy scale, though they have been optimized for special benchmark processes (e.g. decay of a Higgs-boson into two photons). The most characteristic feature of this detector type is that they are as hermetic as possible to cover the entire solid angle. On the other hand the LHCb-detector aims at detailed studies of decays involving b-hadrons and the ALICE-experiment is designed to concentrate on the heavy ion collisions. Driven by these more special interests the two latter collaborations have built their detectors as a one-sided spectrometer (LHCb) and as a particle identification device, respectively (ALICE). Nevertheless, the physics programs of each experiment contributes to the special field of the others. For example, the CMS-collaboration has elaborated a program to analyze heavy ion collisions as well [13].

Ref. [13]

2.3 The Compact Muon Solenoid Detector

The size and the material of a modern multipurpose-detector are determined by the fact that all particles, except for muons and neutrinos, produced in a central collision are stopped within the detector volume. The goal is to measure the data of one collision, called an *event*. That means to determine not only the energy and momentum flow but also to gather and combine information to identify the type of particles and objects that have entered the detector. For that reason the CMS detector is built from several substructures (Figure 2.2).

Fig. 2.2

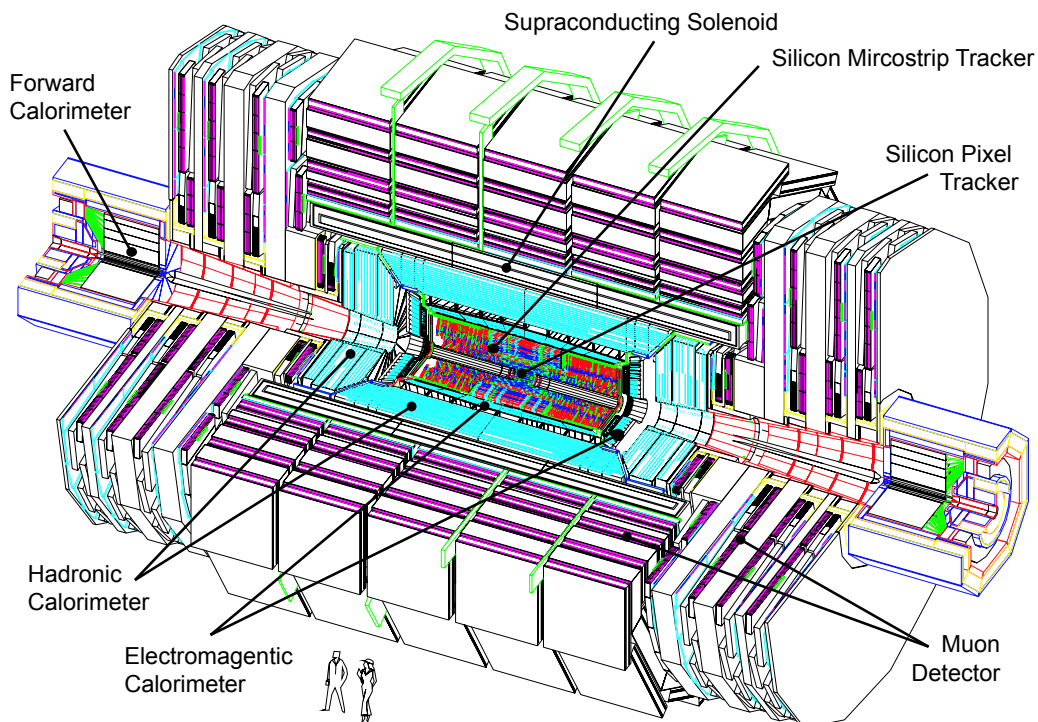


Figure 2.2: Schematic view of the CMS detector. The detector has a length of 21.5 m, a diameter of 14.4 m and weights 12500 t [8].

The interaction point is located at the centre of the detector inside the evacuated beam pipe. This is covered by the silicon pixel tracker and further outside followed by the silicon strip tracker. The data from this inner tracker, which is arranged in a magnetic field of 3.8 T, allow to determine the tracks of charged particles by fitting a track to the hits in the pixel and strip detectors. The curvature of such a

track provides the transverse momentum and the sign of the electric charge of the particle. On top of this, the combination and extrapolation of several tracks allow to reconstruct initial or secondary decay vertices.

Charged particles with a momentum of less than about $0.7 \text{ GeV}/c$ curl in the tracker. Therefore, the majority of particles of a collision leave the tracker and induce an electron-, a photon- or hadron-shower in the subsequent electromagnetic and hadronic calorimeter (ECAL and HCAL). The full cascade must be seen by the calorimeters in order to assign a proper energy to the particles. As these two detectors are segmented into cells they provide information about the spatial direction by means of the position of the cell that registers a signal. From the amplitude of the signal the energy can be derived that is deposited by the particles. Complementing the energy measurement the detector is equipped with a forward calorimeter to absorb particles escaping almost parallel to the beam line.

If summing over the entries of all cells a measure that is called *missing transverse energy* (*MET*) can be calculated, exploiting that the event kinematics of the proton-proton collision is approximately balanced in the plane transverse to the beam axis. Though this measure is smeared by detector inefficiencies and the detector resolution it still gives a hint how much energy is taken away by neutrinos or other, probably unknown particles, that do not interact with the detector material and thus escape undiscovered.

At this stage only muons remain from the collision products and traverse the solenoidal coil, which creates the magnetic field for the inner tracking detectors and calorimeters. The muons are registered by the muon system which is a detector determined to identify the tracks of muons. For that reason the iron return yoke of the solenoidal coil is placed between the layers of the muon detectors. It forms the return magnetic field of about 2 T so that the tracks of the muons are bent in the muon detector as well allowing for a second measurement of the transverse momentum and the sign of the electromagnetic charge.

The event that is read out from the complete CMS detector has a typical size of about 1 MB . As it is not possible to record each event it must be filtered which event is worthwhile to be stored and which can be discarded to reduce the collision rate from 40 MHz to an archivable value. The CMS filter scheme, the *trigger*, is based on two levels. The first one (level 1 trigger) reconstructs first primitive objects from the data of the muon detectors and the calorimeter, while in the second one (High Level Trigger, HLT) an event pre-reconstruction is processed. Only if some pre-defined signatures are recognized (e.g. a high energetic electron candidate or a certain minimum amount of hadronic energy) the event is accepted and the data are stored. This provides a reasonable HLT-rate of about 150 Hz . At the beginning the trigger menu is setup to the most interesting physical signatures (e.g. high energetic electrons, photons or muons, large amount of MET or hadronic energy etc.) but can be dynamically adjusted if needed.

Further details about the different techniques and technologies that are used for the sub-detectors and their readout electronics, as well as a more detailed description of the trigger scheme can be found in the Technical Design Reports [14–21] and in the Physics Technical Design Report which respects the latest upgrades and changes [22].

Ref. [14–21]

Ref. [22]

2.3.1 Key Parameters of the CMS Detector

The main design goal of the CMS detector have been

- a good muon identification and momentum resolution,
- a highly efficient charged particle track reconstruction and good momentum resolution,
- a good energy resolution for electromagnetic objects, and
- a large geometric coverage to get a good MET resolution.

Tab. 2.1

To meet these requirements Table 2.1 lists the technologies, materials and numbers of readout channels that have been chosen for the various sub-detectors. These key aspects are complemented by selected numbers on the resolution to estimate the expected performance of the CMS detector. They are not meant to represent a precise number but only an order of magnitude, because they depend in most cases on the event geometry. More precise descriptions on the technology and performance can be found in [22].

Ref. [22]

Sub-Detector	Technology	Channels	$ \eta $ -Coverage	Resolution	
Pixel Tracker	Silicon Solid State Detector	$66 \cdot 10^6$	0 – 2.5	spatial	10 – 20 μm
Strip Tracker	Silicon Solid State Detector	$9.6 \cdot 10^6$	0 – 2.5	spatial p_T	10 μm 1 – 7%
ECAL	PbWO ₄ -Crystals	$75.8 \cdot 10^3$	0 – 3.0	energy	0.4%
HCAL	Brass, Stainless Steel Plastic Scintillator	$4.6 \cdot 10^3$	0 – 3.0	p_T	10 – 14%
HCAL Forward	Stainless Steel Quartz Fiber	$1.8 \cdot 10^3$	3.0 – 5.0	p_T	10%
Muon Tracker	Drift Tube Chambers, Cathode Strip Chambers, Resistive Plate Chambers	$1 \cdot 10^6$	0 – 2.4	spatial	200 μm

Table 2.1: Summary of the key parameters of the CMS sub-detectors. The resolutions are calculated/extracted for characteristic objects with an energy/transverse momentum (p_T) of 100 GeV according to the information in [22]. The spatial resolution for the strip tracker is meant as the resolution of the transverse impact parameter.

2.3.2 CMS Coordinate System

Any geometrical measure is defined in the CMS coordinate system. The origin of that right-handed system is located at the nominal interaction vertex at the centre of the CMS detector. The x-axis is pointing to the midpoint of the LHC ring, the y-axis is pointing vertically upwards, and the z-axis is pointing tangential along the beam direction towards the ALICE-experiment (see Figure 2.1). The x- Fig. 2.1 y-plane is denominated as the *transverse plane*. Thus, each measure with a lower index "T", e.g. the transverse momentum p_T , denotes the components of a measure in that plane. In contrast the normal of that plane is referred to as the *longitudinal direction*, indexed with "L".

In this coordinate system two angles are defined. The azimuthal angle ϕ is defined as the angle in the transverse plane measured from the x-axis. The polar angle θ is running from the positive z-axis in the y-z-plane. The latter is transformed into the pseudorapidity $\eta = -\ln(\tan(\frac{\theta}{2}))$. From that a new distance ΔR between two points in the $(\eta-\phi)$ -plane is defined as

$$\Delta R = \sqrt{(\eta_2 - \eta_1)^2 + (\phi_2 - \phi_1)^2} = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \quad (2.1)$$

This measure is invariant with respect to Lorentz-transformations along the longitudinal axis.

Chapter 3

The CMS Silicon Microstrip Detector

THIS chapter gives an overview of the silicon tracking detector and its basic sub-structures. This knowledge is needed to understand from which components the tracking detector is built because the classification of defects of the microstrip detector modules¹ is addressed in that chapter. Therefore, the functional principle of a module as the key component of that silicon strip tracker is outlined. Then the qualification procedures are reviewed and finally the latest results on the quality of the silicon microstrip detector modules are quoted.

3.1 The CMS Silicon Tracker

In Figure 3.1 a schematic cross section of the CMS silicon tracker is depicted. As one can see this CMS-subdetector is divided into several sub-devices. These are

1. the Pixel Detector consisting of three barrel and two endcap layers of pixel modules,
2. the Tracker Inner Barrel (TIB) consisting of four layers of microstrip modules,
3. the Tracker Outer Barrel (TOB) consisting of six layers of microstrip modules,
4. the Tracker Inner Disk (TID) consisting of three disks of microstrip modules, and
5. the Tracker Endcaps (TEC) consisting of nine disks of microstrip modules.

Despite the fact that the modules are aligned parallel to the beam axis in the barrel region and perpendicular to the beam axis in the forward region, there is only a difference in the geometry but not in technology. There are two rectangular sensor types for the inner and outer barrel, and eleven wedge-shaped geometries² for the forward region. The thickness is either 320 μm or 500 μm and the surface is segmented by 512 or 768 aluminium readout strips with a pitch of 80 μm to 205 μm depending on the geometry. In total, 15148 modules cover a sensitive area of about 198 m^2 which is currently the largest silicon based tracker. The expected hit rate decreases from about 60 kHz to 3 kHz per mm^2 if the radius increases from 22 to 115 cm with respect to the interaction point.

All modules are assembled into larger components (sub-structures), called shells for the TIB, rods for the TOB, and disks/petals for the TID and TEC. These sub-structures represent the skeletal structure

¹A silicon microstrip detector module is only referred to as a *module* throughout the entire rest of this thesis.

²The modules of the same geometry have equal distances from the beam pipe in the forward regions. Hence in the final configuration they form a circle and thus they are also denoted as *ring-N* modules. For instance, a ring-1 module is located in the first circle counted from the beam pipe outwards.

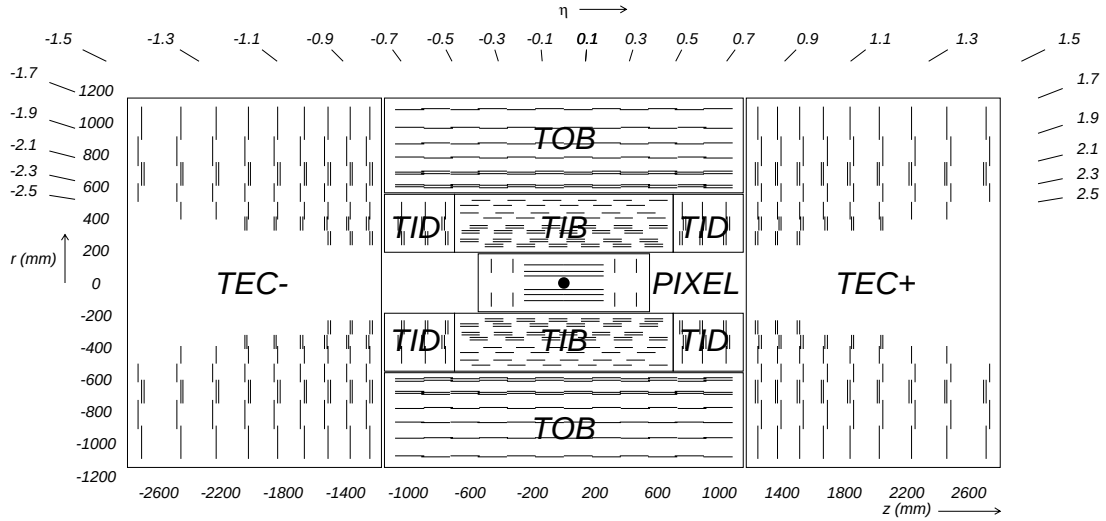


Figure 3.1: Schematic cross section through the CMS inner tracking detector in the $(\eta - r)$ -plane. The beam pipe, not shown in that picture, would run from the left to the right at $r = 0$. Larger sub-detector structures are enclosed by black boxes and labeled with the abbreviation of its name. The small horizontal and vertical lines represent the position of the silicon microstrip detector modules. Double lines indicate two modules that are mounted back-to-back. The surface of one of these two modules is rotated against the second one by a stereo angle of 100 mrad. [23].

of the tracker and provide the possibility to mount modules, as well as to house the infrastructure like electrical and optical cables, or cooling pipes to operate and read out the tracker.

3.2 The CMS Silicon Sensor Layout

If a charged particle passes through matter it continuously loses energy by interactions. The atoms along the path of the charged particle are ionized which is mathematically described by the Bethe-Bloch formula. These charge signals will be localized, and they form the starting point for the readout and reconstruction chain to determine on which track a charged particle has traversed the silicon tracker. In case of the CMS tracker these ionization charges are created and detected using the principle of a semiconducting pn-diode.

Fig. 3.2

In Figure 3.2 a three-dimensional cross section of the silicon sensor is depicted. It can be understood that a single silicon detector module is based on a structure of n-doped silicon with p^+ -doped implants below the readout strips (pn-junction).

If the sensor volume is reversely biased, i.e. applying a bias or depletion voltage of up to 500 V to the aluminium backplane with the counter pole being the bias ring connected to ground, the large n-bulk, which dominates the vertical dimension of a sensor, is depleted from free charge carriers despite of these which are generated by thermal excitation (*leakage current*). For that reason a silicon detector is operated at low temperatures to suppress not only the leakage current but also the clustering of damages of the silicon lattice induced by radiation [24]. For the CMS environment the sensors must stand an integrated energy deposition corresponding to $1.6 \cdot 10^{14}$ 1-MeV-neutron equivalents per square centimetre for a lifetime of ten years. This continuously destroys the silicon lattice and degrades the performance of the silicon tracker. That cannot be prevented but the defect clustering can be delayed when operating the tracker at -10°C .

Ref. [24]

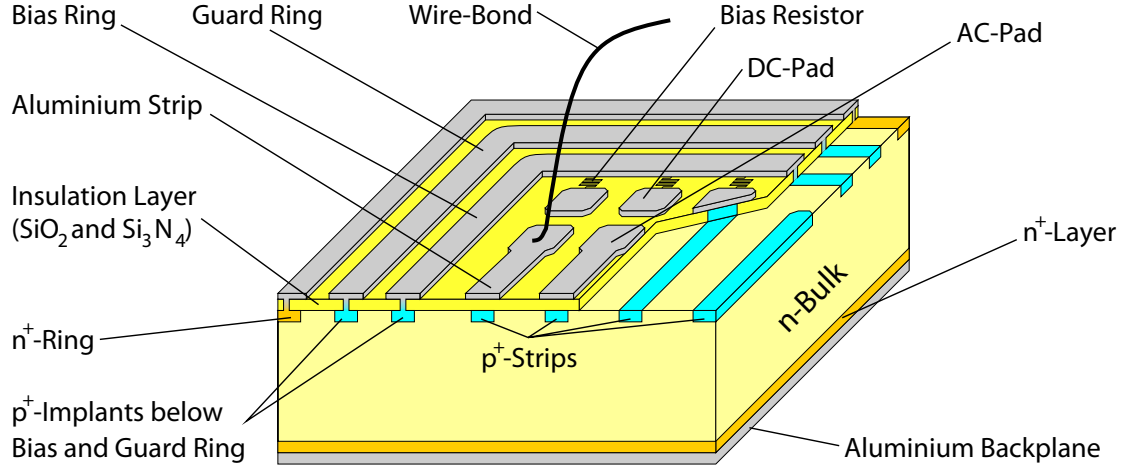


Figure 3.2: Schematic design of the CMS silicon sensors [25]. The sensitive volume is determined by the n-bulk in which p^+ -strips are implanted. They are separated from the aluminum readout strips by an insulation layer which capacitively couples the signal to the electronics. In addition to that the sensor houses structures to conduct the current to ground (DC-pad, bias resistor, and bias ring), and to degrade the internal electric field at the edges of the sensor to ensure the voltage stability.

If a charged particle from the proton-proton collision enters the depleted n-bulk it generates electron-hole pairs which start to drift towards the lower and upper side of the sensor due to the internal electric field which points from the p^+ -implants to the backplane. This signal is capacitively coupled to the aluminium readout strip because the p^+ -implants are separated from the strip via an insulation layer. It is fed into the readout electronics by means of a wire-bond connection on the AC-pad. The electric charges as well as the leakage current are extracted from the sensor by the DC-pad. This pad has a direct electrical connection to the p^+ -implant, and on the other side it is linked to the grounded bias ring via a bias-resistor. Finally, for field-forming reasons and to ensure the voltage stability the sensor is surrounded by two floating guard rings. This functional principle is equal for all geometries of microstrip modules. The mean signal for a minimum ionizing particle after traversing $500\text{ }\mu\text{m}$ of depleted silicon corresponds to about 50000 electrons.

More details on the concept of the CMS sensor design can be found in [26]. A mathematical description Ref. [26] of the physical processes in silicon solid detectors can be reviewed in textbooks like [27].

Ref. [27]

3.3 The CMS Silicon Microstrip Detector Module Layout

A silicon microstrip detector module represents the smallest independent unit for particle detection in the inner tracker. As depicted in Figure 3.3, it is composed of several components. The basis consists of a light but stiff carbon fiber structure serving as an efficient heat removal and to which the components are stuck by thermally conductive glue. These are the silicon sensor(s) and a kapton foil which houses the high voltage lines and is therefore in ohmic contact to the backplane of the sensor. On the right side the frontend electronics (FE-hybrid) can be recognized containing all readout and control chips (details will follow in chapter 3.4). The local separation of the sensor and the hybrid has the advantage that the heat dissipation is facilitated. Thus, the silicon sensor(s) can easier reach their envisaged operating temperature of $-10\text{ }^\circ\text{C}$. The connection between the readout channels on the sensor and the input to the readout chip is provided by the pitch-adapter. It serves as passive device to adapt the pitch of the readout strips on the sensor to the smaller pitch of $44\text{ }\mu\text{m}$ at the input of the readout chips. The electrical connections between the readout chips and the pitch-adapter, the pitch-adapter and the near sensor, and

Fig. 3.3

Chp. 3.4

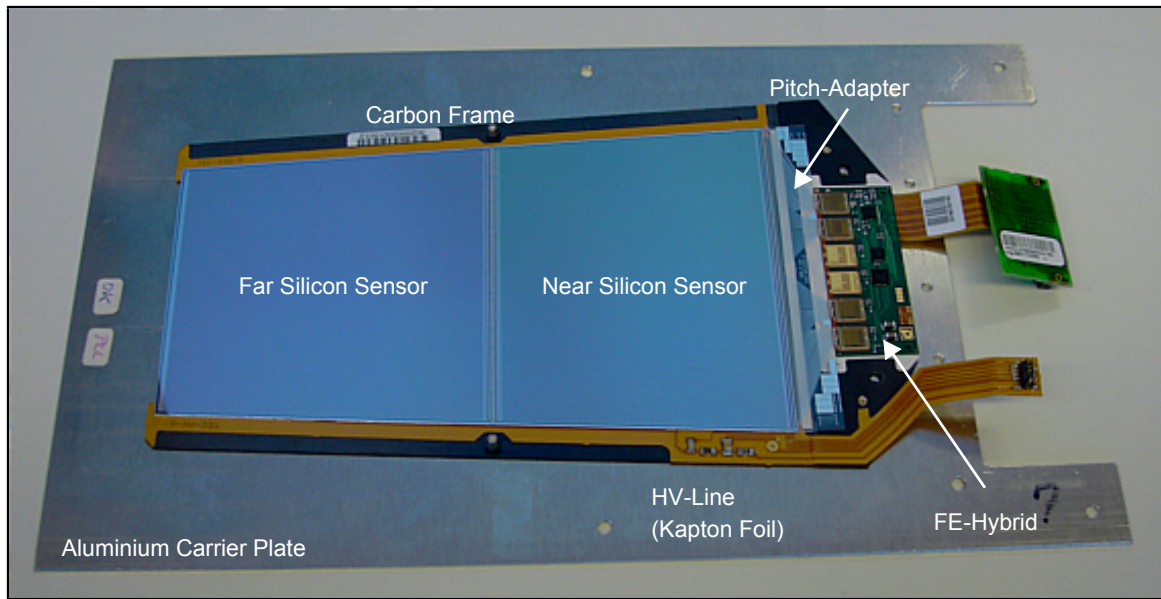


Figure 3.3: Photograph of a CMS silicon microstrip detector module with two sensors (TEC, ring-6 geometry). Modules employing only one near sensor equip TIB and TID, as well as TEC ring-1 to ring-4.

the near and the far sensor in case of two sensor modules are established by wire-bond connections. For the sake of a secure and facilitated handling and transporting the module is screwed to an aluminium carrier plate.

3.4 The CMS Silicon Microstrip Detector Hybrid Layout

A magnified photograph of the FE-hybrid with the pitch-adapter is shown in Figure 3.4. Stiffened by a ceramic plate the four layer polyimide copper multilayer flex circuit carries and interconnects the chips which are used to readout and control the module.

- | | |
|-----------|---|
| Ref. [28] | 1. The <i>Analogue Pipeline Voltage Mode (APV)-chip</i> [28] serves for the analogue readout of 128 channels. Each channels is equipped with a pre-amplifier and a CR-RC pulse shaping circuit. Furthermore, the APV-chip has a deconvolution logic to decrease the signal rise time from 50 ns to 25 ns which is essential for the operation of the readout in the LHC high luminosity phase. The data of any channel can be stored for up to 4.8 μ s by means of an internal pipeline storage matrix (192 cells times 25 ns per channel) which provides sufficient time until the level-1 trigger demands the data for further processing. Finally, the APV-chip houses a calibration signal logic which allows for the creation of well defined input signals. |
| Ref. [29] | 2. The <i>MUltipleXer (MUX)-chip</i> [29] reduces the number of readout lines by a factor of two, because it unifies the data output of two APV-chips to one single line. Additionally, it converts the differential current output of the APV-chips into differential voltages by means of a series of internal resistors. By that the signal height is adapted according to the requirements of the Analogue-to-Digital Converters in the frontend drivers [30] which represents the next step in the data readout. |
| Ref. [30] | |
| Ref. [31] | 3. The <i>Tracker Phase Lock Loop (PLL)-chip</i> [31] delivers the clock and, after decoding, also the trigger signals to the APV-chips. Furthermore, a time delay can be adjusted to compensate for signal |

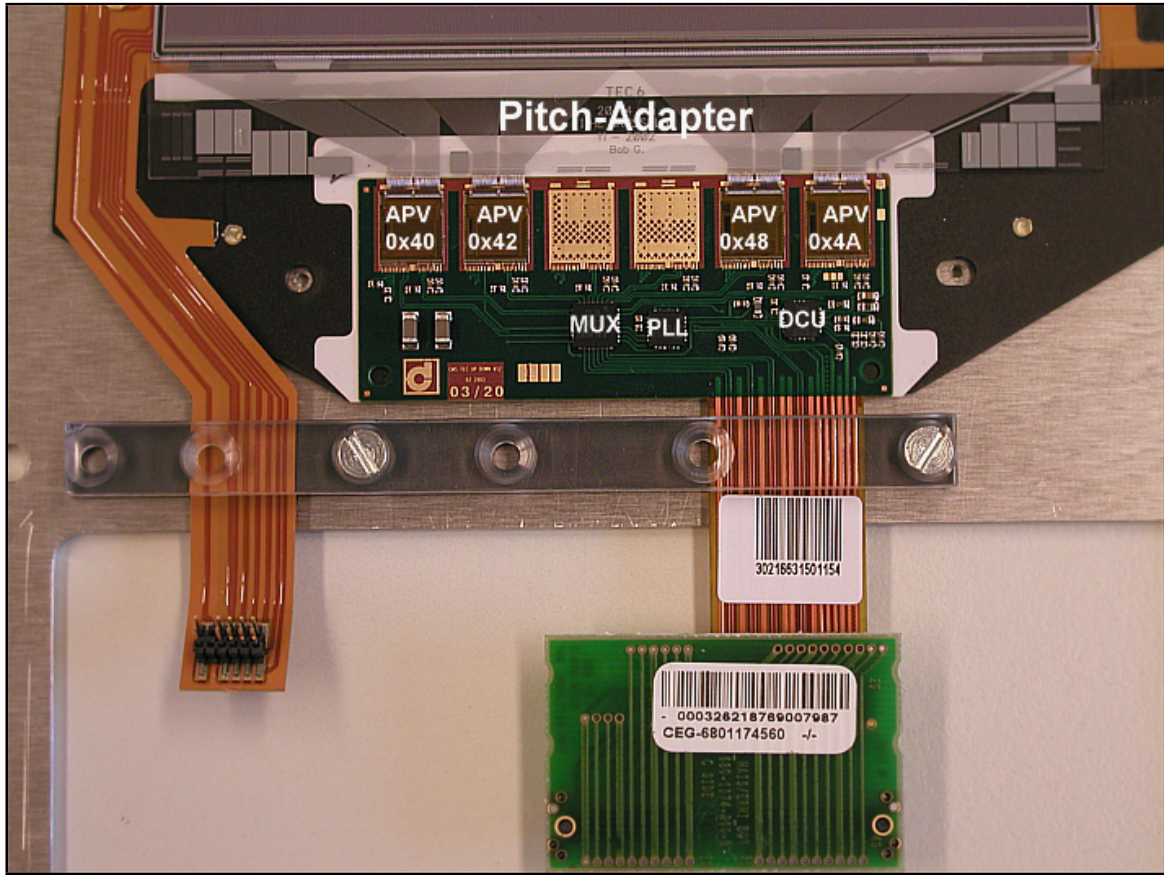


Figure 3.4: Frontend-Hybrid with pitch-adapter of a TEC ring-6 module. The FE-hybrid of that type houses four readout chips (APV-chips) as well as the common chips for the readout and control of a module (MUX, PLL and DCU).

propagations. This synchronizes the data output of all FE-hybrids to the same LHC-timestamp independent of the geometrical position of the FE-hybrid in the tracker.

4. The *Detector Control Unit (DCU)-chip* [32] monitors the operating voltages of the FE-hybrid and measures the leakage current of the sensor. Furthermore, it records its own and the temperatures of the silicon sensor and the FE-hybrid. Ref. [32]

The FE-hybrid holds either four or six APV-chips depending on the number of channels it must readout. Though there are twelve different types, due to the number of module geometries, the basic equipment and functional principle is the same for all types. The communication between the readout software and the FE-hybrid is realized via the I²C-protocol [33]. Ref. [33]

3.5 The Module Production and Qualification

In order to sort out defective components right from the beginning a very close mechanism of qualification, classification and recovery procedures had to be established for the production of the silicon microstrip modules. This did not only involve the physics institutes responsible for the assembly of modules but also the industrial manufactures of the components. Systematic sources of defects could be detected as early as possible, and the current status of each component could contemporarily be retrieved.

Chp. 4

The main aspects of the production and qualification are reviewed in the following while the recovery stage is explained in chapter 4.

At first, each active (e.g. the sensor) and passive (e.g. the carbon frame) component must continuously pass a visual prequalification to spot either production imperfections or mechanical damages happened during the production processes or on transports. Then for the silicon sensors the high voltage stability and the absolute value of the leakage current are checked. The FE-hybrid is tested to prove that all electrical connections are established and the communication to and between the chips is working properly. After that the FE-hybrid is linked with a pitch-adapter and the APV-chips are connected to the pitch-adapter strips by wire bonds. Again a connectivity and communication test is performed, and additionally, the new connections between the pitch-adapter and the APV-inputs are checked. This involves an electrical test and a mechanical pull-test as well, to grant the robustness of the connections. Control samples of FE-hybrids with pitch-adapters are additionally qualified in a cold-box test setup to verify the functionality of the electronics at the operating temperature of -10°C and even below.

At the next stage, the sensors and the FE-hybrid with pitch-adapter are glued by high precision robots (gantries) to the carbon support frame. In the last step before completing the module the wire-bonds between the pitch-adapter and the near sensor, the near and the far sensor (in case of two sensor modules) and the bias voltage connection to the backplane are done. The quality checks involve again mechanical pull-tests of the wire-bond connections as well as a high voltage test to verify that the bias bonds are fixed to the backplane in a sensible way.

In the end, the module has to pass one last test, where all components are again checked for their electrical functionality. Furthermore, there is a special interest to determine if and how many readout channels are defective and which type of defect is present. Apart from that cold tests are carried out to verify the electrical functionality again, and to survey its mechanical stability at operating temperature, too.

3.5.1 ARC Test System

Ref. [34]

Due to the variety of single production steps and the large number of modules that had to be produced the mass production could not take place at one central location. Therefore, the components had to be transferred from one institute to another at a high rate. As there are a lot of quality checks which were constantly repeated there was a need for a unique and automatized test system to accompany the entire production chain. Fast, automated and reliable routines were demanded, so that these tests could be performed at a high frequency during the mass production. These requirements were met by the *APV Readout Controller (ARC)-system* [34].

Chp. 4

It consists of a hardware part which contains all necessary devices for powering and reading out a module. On the software side it is an easily configurable program providing all test procedures for the entire qualification. These are an IV-test to monitor the leakage current as a function of the bias voltage, a functional test which extensively surveys all functionalities of the FE-hybrid, and a series of so-called deep tests. They run a multitude of redundant sub-tests and a highly evolved failure detection algorithm to determine any type of known defect of a readout channel. A detailed knowledge of the defect directly influences and facilitates a possible recovering investigation as it will be described in chapter 4.

Fig. 3.5

Ref. [35]

Two important investigations had to be accomplished before the algorithms were able to reliably determine the grade of a module. At first, the test setups (similar to these two located at Aachen, shown in Figure 3.5) of all contributing institutes had to be cross-calibrated to establish a common standard, especially for grounding and shielding the module, so that the test results did not depend on the test center [35].

In a second analysis, the test data of the institutes were combined and compared to identify the most convenient measures to not only separate the defective from the regular channels, but also to identify the

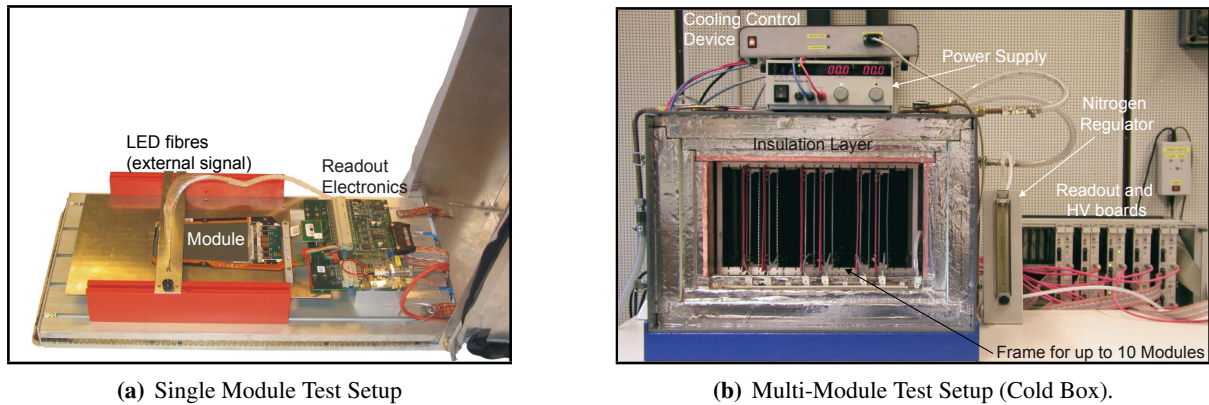


Figure 3.5: Test stations for single module tests (a) and multi-module tests (b). Both setups are constructed such that the preparation of the module for the test procedure requires only a minimum of time and plug connections. The multi-module setup allows in addition for tests at an adjustable temperature between -25°C and ambient temperature to verify the functionality of the module at operating temperature (cold test).

type of defect as efficient as possible and to keep the rate of misclassified, regular channels as low as possible. Beside the compulsive IV-test and functional test the measures listed in Table 3.1 were taken into account to determine the grade of a module. The dedicated distributions are well separated and

Tab. 3.1

Test Measure	Type of Defect
Amplitude of calibration signals	Short-circuits, open channels, defective inverters
Time when the maximum of a calibration signal is reached (peak time)	Open channels (important to localize open bonds)
Amplitude of calibrations signal versus leakage current	Pinholes and dead channels
Noise level	Noisy channels
Slope of amplitude of calibration signal response versus injected signal	Short-circuits, open channels (only redundant information)

Table 3.1: Summary of measures incorporated in the fault identification and classification algorithm of the ARC software. The type of defects are more closely explained in chapter 3.5.2.

have the lowest overlap not only for the distinction of regular and defective channels but also for the classification of defective channels (Figure 3.6). All details on that investigation and the distributions of the other measures are presented in [25, 36].

Fig. 3.6
Ref. [25, 36]

3.5.2 Classification Scheme of Defective Modules

The above described production procedure allows for a variety of defects that might occur during the production chain. The following list summarizes the most important classes of defects observed during the production which can be detected by the test system.

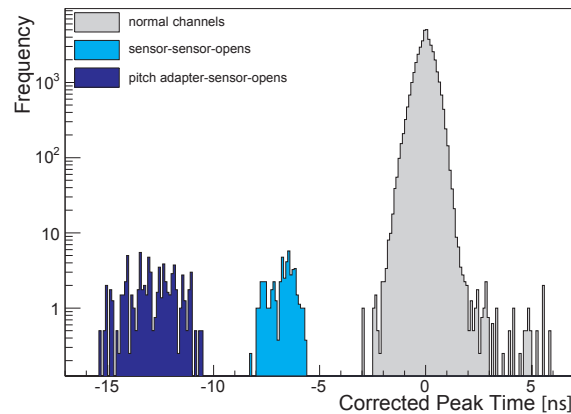


Figure 3.6: Example for the study how to not only separate defective from regular channels, but also to determine a measure to identify the type of defect (here pitch-adapter-sensor-opens and sensor-sensor-opens, see chapter 3.5.2) [36].

1. High-Voltage Problems

A sensor does not hold the high voltage. If this is detected at the sensor qualification stage the sensor is rejected and not used for module assembly.

2. FE-Hybrid Problems

The FE-hybrid does not pass the connectivity and communication test. If the connectivity failure can be traced down to a missing wire-bond connection on the hybrid surface this problem might be repairable. In all other cases as well as there is a problem in the communication the defect is most likely to be related to errors within the electrical circuit or the chips. If this is detected early enough the FE-hybrid is removed from the production chain. This also holds if the FE-hybrid is not passing the cold test.

If the hybrid is already connected to a pitch-adapter defects on single channels can occur. In most cases they are repairable, e.g. two or more channels are short-circuited by touching wire-bonds, or a single channel has a broken or missing wire-bond connection.

3. Mechanical Problems

If the module does not fulfill the mechanical specifications after the assembly on the gantry robots it is not rejected from the module production. If there is lack of a certain module geometry they are used accepting the following less precise track measurements instead of having an empty position in the tracker.

4. Quality Problems

After its final qualification test the module is rated as *Grade A* or *B* if the number of defective channels does not exceed 1% or 2%, respectively. In all other cases ($\geq 2\%$) it is graded as *C* and will not be used in the tracker. Possible channel defects are as follows.

- (a) A channel can be disconnected from the readout because of a missing or broken wire-bond connection. This defect is generally called an *open channel* and is usually specified by the position of the defect, e.g. sensor-sensor-open or pitch-adapter-sensor-open. A microscope picture of that kind of defect is shown in Figure 3.7(a). Readout strips with an interrupted aluminium metallization due to a very rare occurring dust pollution during the sensor fabrication are assigned to this category as well.

Fig. 3.7(a)

- (b) Two or more channels are denoted as *short-circuits* if they are electrically connected because of wire-bonds touching each other. In very rare cases metallization imperfections lead to the same kind of defect (Figure 3.7(b)).

Fig. 3.7(b)

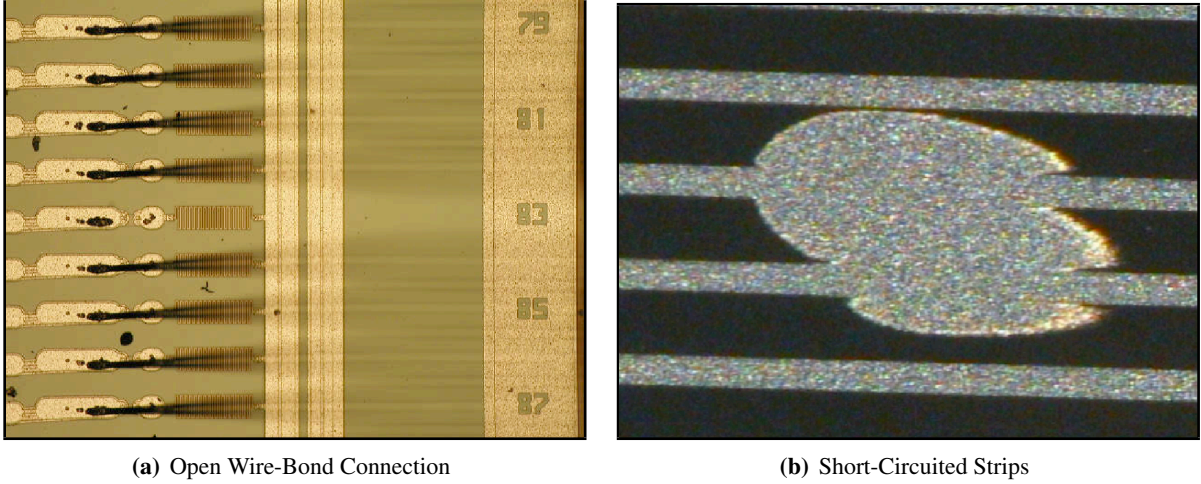


Figure 3.7: Microscope photograph of an open wire-bond connection (a). The wire-bonds are the black shadow-like objects pointing from the aluminium pads on the sensor surface to the right edge of the picture. The two short-circuited strips in (b) are created by a (very rare) metallization imperfection. The straight line from the left to the right represent the aluminium strips on the sensor surface.

- (c) A broken or weak insulation layer between the aluminium strip and the underlying p^+ -implant (*pinhole*) change the reference potential at the pre-amplifier input of the APV-chip. Therefore, the channel is not capable of registering any signal.
- (d) A *defective inverter* indicates a defect inside the electronic circuit of the APV-chip, when the polarity of a signal cannot be turned anymore. This becomes a problem, because the signals must be inverted before they are transmitted to the external data acquisition.
- (e) A *dead channel* shows no optical deviation from a good one, but the signal is not transmitted by the APV-chip. Most probably this type of error is caused inside the APV-chip by a defective pre-amplifier or a broken electrical contact to or from the pipeline storage matrix.
- (f) A *defective pipeline cell* is a minor problem because it affects only one time slot out of 192 of the whole pipeline depth. Thus, this is only an intermittent problem without losing an entire channel.
- (g) Channels that exceed a certain noise level are labeled as *noisy channels*. When a particle hits a single channel or a cluster of channels then the amplitude of the signal is compared to the noise of the channel or the cluster (signal-to-noise ratio). This ratio and therefore the requirement for a low noise is important because the signal height must exceed five times the mean standard deviation of the noise to consider the channel or cluster as being hit by a traversing particle.

3.6 The CMS Silicon Microstrip Module Quality

A detailed study of about 500 modules for the tracker endcap has proven that all known defects can be detected because they reveal a suspicious behaviour in any of the measures mentioned in Table 3.1. Tab. 3.1 It is necessary to introduce cut values for the different failures and to logically combine the different

Fig. 3.6

measures to attribute a type of defect to a faulty channel. Due to statistical fluctuation and as some distributions are not perfectly separated as these in Figure 3.6, the classification algorithm determines in 86% (dead channels) to 100% (pinholes) the correct defect class. Only the order of 0.01% of the regular channels are mistagged as faulty [36].

All test results are stored channel-wise in a central tracker database [37]. At the end of the mass production the overall quality of the CMS silicon microstrip detector modules has been extracted from there. A yield of 97.2% has been achieved. Out of these modules which can be integrated into the tracker the fraction of modules rated as Grade A is 97.7%, while the others are Grade B.

Ref. [38]

In order to get an idea how this figures can be interpreted for the tracker endcaps, the absolute number of defective channels has been determined to about 5500. This must be compared to the 6400 modules that are needed for the endcaps corresponding to approximately 4000000 readout channels. The failure rate is well below one defective channel per module or in total only of the order 0.1% on average. The impact of the defective channels on the track reconstruction efficiency is investigated in [38] and found to be negligible.

Chapter 4

The TEC Silicon Microstrip Module Debug Centre

THE large number of silicon microstrip detector modules that had to be produced for the CMS inner tracking detector required fast, and automated test setups. They had to provide a reliable and reproducible procedure for the quality assurance at all production steps in the laboratories participating in the production of the detector modules. This topic has been discussed in chapter 3. But additionally, ^{Chp. 3} a strategy how to deal with components that did not pass the quality criteria had to be established. They were excluded from being used for the tracker, but the source of defect should be determined to exclude the presence of systematic problems during the mass production. In addition, for economical reasons it should be attempted to recover as many of them as possible.

Only these modules were immediately debugged at the laboratory that observed a defect or failure if the problem was expected to be easily solved. All the other modules were collected and regularly sent to a *Module Debug Centre* where a thorough investigation and the bookkeeping of these modules was concentrated. By that procedure the mass production was not blocked by time-consuming debug processes in case that the observed failures occurred neither too frequent nor were expected to be too severe. Furthermore the identification of systematic sources of errors was facilitated.

This chapter deals with the debug procedure and results of these examinations being part of the mass production for the modules of the tracker endcaps (TEC).

4.1 The TEC Module Debug Centre

At the beginning it was expected that during the mass production of TEC modules mainly readout problems and defective strips led to a disqualification of modules. Thus, this centre was appointed to be the TEC Module Debug Centre which had the broadest expertise in the understanding of the functionality of TEC modules. The III. Physikalisches Institut B gathered this knowledge when designing and programming the official test-setup for the silicon microstrip detector modules. As it was furthermore equipped with the most convenient test facilities necessary for module qualification, the centre was capable of debugging nearly all kind of known defects. These test facilities are

- a microscope for precise optical examinations to uncover correlations between mechanical and electrical defects,
- a Little Frontend Hybrid Industrial Tester (LFHIT) for electrical tests of the FE-hybrid,
- two independent ARC setups for deep tests of entire modules, and

- a setup for cold tests to examine failures that might affect the mechanical or electrical long-term stability in an environment with low temperatures.

Ref. [39] A detailed explanation of the test-setups, their properties and functionalities can be found in [39] and [35].

Ref. [35] Furthermore, a part of the integration of modules to the larger TEC sub-structures (petals) was located at the same institute. The Debug Centre could interact with the involved staff to exclude failures that were related to problems only occurring in this integration phase or even later in the dedicated qualification tests for the sub-structures.

Ref. [40] Before a module was assigned to one of the common defect classes described in the next sections a standard reception procedure was applied to collect all available information to elaborate an individual strategy for the debug procedure. Having registered the module in an internal inventory table, it was checked for any optical deviation or damage. After that the central tracker database was queried to read the information provided to the Debug Centre. In order to facilitate this exchange a web interface [40] allowed to upload descriptive information as well as data files which might be useful for the Debug Centre. Afterwards the module was passed to the appropriate test-setup to finalize the evaluation.

Chp. 3.5.1 The APV-Readout Controller (ARC) System served as a common starting point for all the above mentioned test setups. The means and methods of that system to readout a module, to analyze and determine probable defects played an essential role for the work of the Debug Centre as it had been for the qualification of modules (see chapter 3.5.1). In the end, the results of all single steps were combined, uploaded to the central database, and the module were assigned to a global defect class. Depending on the kind of defect it was decided how to proceed with the module according to the grade of severity of the problem.

1. The module was excluded from the integration into the tracker.
2. The module was excluded from the integration into the tracker, but some components could be recycled after disassembling the module.
3. The module could be integrated into the tracker, if there was a lack of this module type.
4. The module was repairable and could be used for the tracker without restrictions after repeating the standard qualification procedure.

Ref. [36] There is a fifth category which contains all modules where no problem could be reproduced or observed. These modules were assigned to one of the defect classes only due to the information that were provided via the central database. After a very careful investigation and a final qualification following the standard rules [36] they were fed back to the production line.

4.2 The Classification of Defects and Failures

The subsections of this chapter summarize a list of classes of defects which were actually observed by the Debug Centre. It does not include all errors that can theoretically affect a module. Each subsection contains a statistics to show at which frequency the various defects occurred and how many modules could be repaired.

4.2.1 IV-Problem

The IV-problem assesses the behaviour of a module with respect to its leakage current as a function of the bias voltage. The value of the leakage current at 450 V must not exceed 10 μA or 20 μA for 1-sensor

and 2-sensor-modules, respectively. Furthermore a relative cut of less than five times the leakage current of the measurement from the sensor test must be passed.

As the leakage current strongly depends on environment variables like humidity and ambient light the procedure to verify an IV-problem consists of storing the module for about 60 hours in an enclosed box that is flushed with dry air. Periodically the depletion voltage is ramped up to 450 V and the current is determined in steps of 10 V. For modules with regular IV-curves the leakage current approaches an almost constant value between $0.3\ \mu\text{A}$ and $1\ \mu\text{A}$ depending on the module geometry, i.e. the size of the silicon area, when reaching the target voltage. Whereas when facing an IV-problem the current suddenly increases exponentially (IV-breakdown). Two examples for a regular IV-curve and an IV-breakdown are depicted in Figure 4.1.

Fig. 4.1

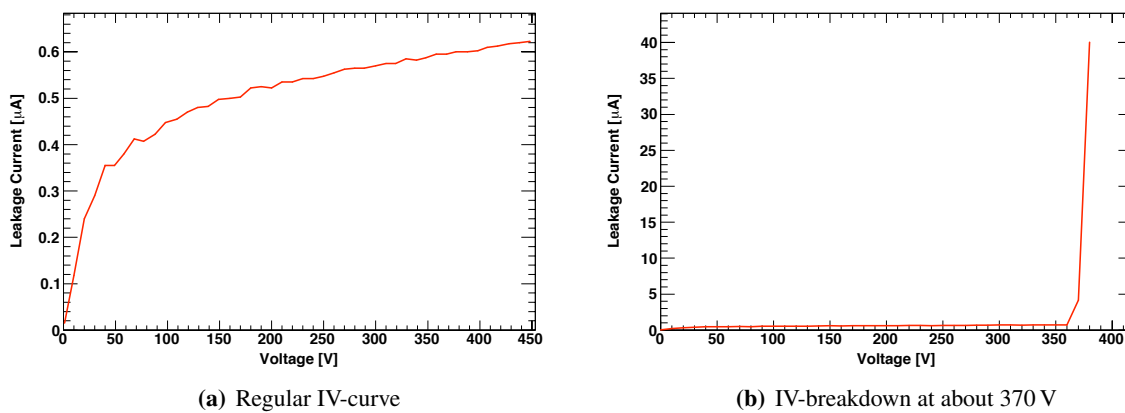


Figure 4.1: Typical shapes of IV-curves of a module (a) without and (b) with an IV-breakdown

For many modules already the initial measurement after a few minutes in a dry and dark environment revealed the expected IV-curve remaining long-term stable for the entire sequence of measurements. Most probable the environment was not properly adjusted to standard conditions when the module was tested with the potential IV-problem. In contrast, really affected modules had only slightly relaxed, when the voltage of the breakdown point increased with time. But they never arrived at the shape of the IV-curve of a regular module. The final statistics on the IV-problem is given in Table 4.1.

Tab. 4.1

The reason why modules, which were assembled with originally good sensors, were not passing the IV-test is not entirely clear. It is assumed that the production process unintentionally affected the quality of the sensors. This is supported because damages like scratches on the sensor surface were observed. In a few cases the IV-problem could be related to the *micro-discharge effect (MDC)*, an electrical charge avalanche inside the sensor due to local peaks of the electric field. Finally, modules that could not be repaired were sent to the sensor qualification centre in Vienna¹ for an even deeper investigation of that problem.

In addition to the described failure all deviations from the regular shape of an IV-curve were assigned to this category. The most prominent one was a weak contact of the high voltage lines to the sensor backplane. In this case the HV-circuit can be interrupted and the current drops down or is still at zero at a significant bias voltage. This problem can be traced by optical investigations and measuring the ohmic resistance of the HV-connection to be extremely high or infinity. To systematically exclude this error the conductive glue connection between the HV-line and the backplane was enforced and later wire-bonds were added by default.

¹Institut für Hochenergiephysik, Vienna, Austria.

Ring	1n	1s	2n	2s	3	4	5n	5s	6	7	total
Number of Modules	1	-	1	-	10	8	3	7	22	7	59
Optically or MDC suspicious	1	-	0	-	4	2	1	1	10	1	20
Repaired	0	-	1	-	7	2	0	1	7	4	22

Table 4.1: Number of modules sorted by the ring geometry revealing an IV-problem. The third row gives the number of module showing any suspicious reason for a probable IV-problem, while the fourth row contains the number of repaired modules.

The supplements "n" and "s" at the geometry indicate that there exist modules without and with a stereo angle for ring 1, 2 and 5. The ring geometries 3, 4, 6 and 7 have a non-stereo version only.

4.2.2 Frontend Hybrid Problems

The FE-hybrid is the decisive device for the entire data transfer from the module to the global readout and storage. Any failure either in the I²C communication, or in the functionality and synchrony of components on the FE-hybrid, or any deviation in the power consumption are assigned to the defect class of FE-hybrid problems. Only in a few cases the reason of the error can be spotted optically when an interrupted electrical line or any other damage to the readout circuit, mostly mechanical, can be observed. In most cases the origin of the failure is founded inside the electronic devices and thus can only be located on the level of determining the defective chip on the hybrid. Although its interior is not accessible by the means of the Debug Centre each quoted failure shows an unambiguous behaviour in the tests.

A module was typically assigned to this category if it failed in the functionality test during an ARC test routine. The first check was to reproduce the problem using two independent test setups. Afterwards the LFHIT-test proved the functionality of the FE-hybrid at a deeper level (i.e. for instance a dedicated connectivity test of all FE-hybrid components). Table 4.2 numerically summarizes the results of the defective modules in that category.

Tab. 4.2

Ring	1n	1s	2n	2s	3	4	5n	5s	6	7	total
APV synchronization	-	1	1	-	-	-	1	-	1	-	4
Open or short-circuit on hybrid	-	1	1	-	-	-	1	-	1	-	4
I ² C communication	-	-	1	-	1	5	-	-	-	4	12
PLL or DCU problem	-	-	-	-	1	-	-	-	-	3	4
Power consumption	-	1	1	-	-	-	1	-	1	-	4
Bad hybrid	1	1	-	-	-	1	2	1	-	-	6
Problem not reproducible	-	-	1	-	1	-	-	1	-	-	3

Table 4.2: Number of modules sorted by the ring geometry revealing FE-hybrid problems.

The supplements "n" and "s" at the geometry indicate that there exist modules without and with a stereo angle for ring 1, 2 and 5. The ring geometries 3, 4, 6 and 7 have a non-stereo version only.

In one third of all cases a reason for the problem could be determined. For instance, half of the modules failing with an I²C-error suffered from an accident during the assembly to larger sub-structures. The FE-hybrid supply voltage line was connected to the sensor bias voltage and the I²C-registers mainly of the PLL-chips were destroyed by an over-voltage. The DCU-failures could be traced back to a bad soldering of the thermistors to the sensor so that the sensor temperature could not be determined. In addition, a missing metallization on the surface of the hybrid was detected once.

Modules which were considered to be fed back to the production chain had to pass a final long-term test.

Over several days there were permanently read out and tested using the ARC-setup. If the initial error could be reproduced and not a single other error occurred during that time they were put back to the production chain with a hint to still carefully check the behaviour of that module. Modules that finally revealed a FE-hybrid problem are sent to the sensor recycling centre in Vienna² where the module was disassembled and the sensor re-assembled to a new module, provided it still passed the sensor qualification criteria.

Bad hybrids were assigned to this category as well. These are readout circuits which have already been integrated to a module when the entire hybrid batch to which they belong has been rejected. This was mainly due to a via-problem where the quality of the vertical electrical contact between the four metal layers of the FE-hybrid was below the specifications.

4.2.3 Damage of the Carbon Support Structure

The carbon support structure does not only hold the hybrid and the silicon sensor(s) but the module is screwed to the sub-structures by means of precision holes in that frame allowing for a fixation of the module with a precision of 20 μm [23]. If a too large torque is used to fix them on the sub-structure Ref. [23] the carbon frame might break so that it cannot be guaranteed that small pieces might fall off when the module is exposed to thermal stress. As even a very small free flowing piece of carbon might affect the functionality of a module, because of the electrical conductivity of carbon, this must be excluded.

This problem was solved in two ways. On the one hand the torque was determined to safely screw the module to the sub-structure without breaking the carbon frame. In order to not exceed the maximum allowed torque special screw drivers had to be used which can be adjusted to a maximum torque. On the other hand those modules which already suffered from this kind of damage were repaired by thin ceramic rings glued around the holes to enforce the connection of the partially broken piece to the rest of the carbon frame. Figure 4.2 reveals a microscope photograph of a repaired crack in the carbon frame. Fig. 4.2

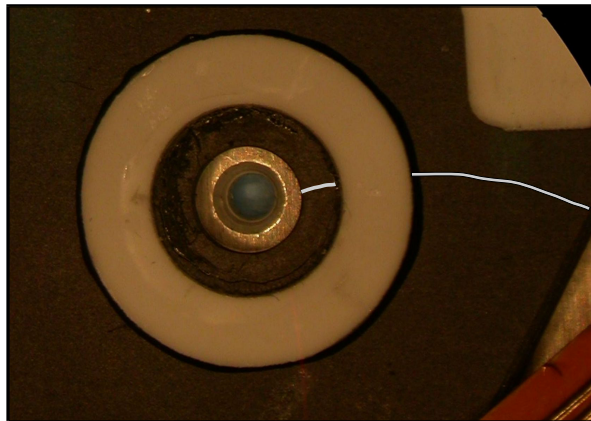


Figure 4.2: Module with a crack in the carbon frame which was repaired by a ceramic ring. The course of the crack has been accentuated by the white line starting at nearly outermost right point of the hole running to the right. The white circular object is the ceramic ring used for the enforcement of the carbon frame [41].

As listed in Table 4.3 16 modules were affected by that failure in total. All but one could be repaired. Tab. 4.3 There was no obvious reason for the large number of ring-3 modules affected. This was just an accidental coincidence at the beginning of the production when the first exercises of integrating the modules to sub-structures took place.

²Institut für Hochenergiephysik, Vienna, Austria.

Ring	1n	1s	2n	2s	3	4	5n	5s	6	7
Frequency	-	1	1	-	11	-	-	-	2	1

Table 4.3: Number of modules separated by the ring geometry where a break of the carbon support structure occurred.

The supplements "n" and "s" at the geometry indicate that there exist modules without and with a stereo angle for ring 1, 2 and 5. The ring geometries 3, 4, 6 and 7 have a non-stereo version only.

4.2.4 Misaligned Modules

Chp. 3.3 As described in chapter 3.3 a module is composed of several components which are glued together. High-precision gantry robots guarantee the precision to facilitate the later alignment and to achieve the envisaged spatial resolution of the silicon tracker. The deviation of the components from global reference points on the module (translations) are determined as well as the angles between the frame and the FE-hybrid/sensor(s) and among the sensors (rotations) after the glue has cured. The Gaussian distribution of a sub-sample of the measured values for all silicon tracker modules (not only TEC) is depicted in Figure 4.3 including the projection to the various assembly centres. As the systematic shifts between the distributions for the integration centres are smaller than the RMS no severe correlation has been observed. If the sensors on a module are misaligned by more than $39\mu\text{m}$ (x-translation), $65\mu\text{m}$

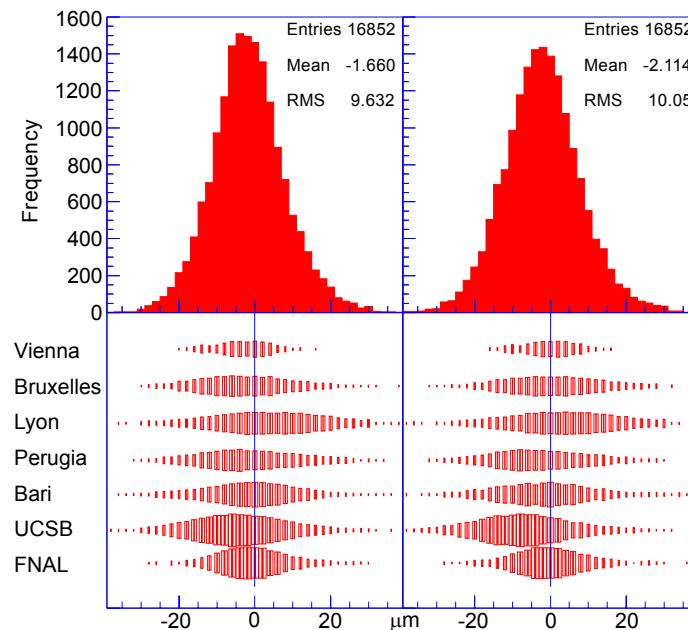


Figure 4.3: Deviation of the near sensor from the nominal position of two reference points in the coordinate perpendicular to the strip direction (x-axis). The lower half reveals the distribution broken down to the seven gantry centres [23] involved in the module production.

Ref. [42] (y-translation) or 0.013° (rotation) [42] the module has only been considered to be integrated into the tracker in case that there is a lack of a certain type of module because the grade in terms of defective channels is actually very good (see Table 4.4).

Tab. 4.4 After a larger number of TEC modules had been produced the cuts on the grade of misalignment were reviewed basing on a larger statistical sample. Driven by the unequal treatment of the translational and rotational cuts, i.e. 0.013° rotational deviation corresponds to approximately $11\mu\text{m}$ deviation while at the same time a translation of $39\mu\text{m}$ is allowed, and by the need for good modules it was decided that the

Ring	1n	1s	2n	2s	3	4	5n	5s	6	7
Frequency	-	1	-	-	3	-	1	3	4	3
Grade A	-	1	-	-	3	-	1	3	4	3
Grade B	-	-	-	-	-	-	-	-	-	-
Number of Repaired Modules	-	1	-	-	2	-	1	2	4	3

Table 4.4: Number of misaligned modules separated by the ring geometries. Besides the total number of modules the quality of the modules is given as well as the number of modules that were repaired after adjusting of the rotational cut from 0.013° to 0.020° (see text).

The supplements "n" and "s" at the geometry indicate that there exist modules without and with a stereo angle for ring 1, 2 and 5. The ring geometries 3, 4, 6 and 7 have a non-stereo version only.

initially chosen cuts on the rotational deviations were too tight, so that some of the originally excluded modules could now be returned to the integration line without a loss of performance [43]. The number of repaired modules is shown in Table 4.4 as well. Ref. [43]

Tab. 4.4

4.2.5 Number of Defective Channels

This category contains all modules which initially do not pass the qualification procedure due to a too large number of defective channels (Grade C) or reveal a suspicious channel which test data deviates from any known standard behaviour. The grading, reviewed in Table 4.5, is applicable only to modules passing the IV- and the FE-hybrid functional test. Except for this general grading scheme there is Ref. [43]

Total Number of Channels	Ring	Grade A	Grade B	Grade C
512	3, 4, 6, 7	0 - 5 ch.	6 - 10 ch.	≥ 11 ch.
768	1, 2, 5	0 - 7 ch.	8 - 15 ch.	≥ 16 ch.
Relative Number for both types		$< 1\%$	$< 2\%$	$\geq 2\%$

Table 4.5: Grading conventions for the two types of TEC modules. The absolute and relative maximum number of defective channels that are allowed for a certain grade are presented. Clustered defective channels are not treated separately.

one exceptional treatment for channels affected by the pinhole defect. As these channels might cause an entire readout chip or a module to fail after a few years of operation when the leakage current has increased due to radiation damage [44], they must be disconnected from the readout for security reasons. Ref. [44]

In order to reliably distinguish defective from regular channels the cut parameters were determined by comparing and averaging the test data of seven centres [25]. Nevertheless, a fluctuation of the test data might have caused the channels to be flagged. In order to evaluate a module of the present category, its test result was verified by an independent test setup and compared to the initial data due to which the module was sent to the Debug Centre. If it was considered to be necessary the test data of the bad strips were contemplated channel-wise to check if the deviation could be reproduced. Table 4.6 summarizes the results of the Debug Centre examinations of 48 modules in that category. Ref. [25]

The optical investigation before each module test provided helpful information to find reasons for the bad grading. In about one fifth of the cases either an inappropriate electrical contact for the bias voltage was observed, or a striking mechanical damage of the module (scratch on the surface, no wire-bonds connection due to damaged bond pads) could be seen generating a significant number of bad channels (mainly short-circuited -, open channels, or pinholes). Another 15% of the defects could be related Tab. 4.6

Ring	1n	1s	2n	2s	3	4	5n	5s	6	7
Grade A	2	-	-	-	3	1	1	3	4	3
Grade B	-	1	-	-	1	-	3	3	3	1
Grade C	1	-	1	-	1	1	5	4	4	2
Total Number of Modules	3	1	1	-	5	2	9	10	11	6

Table 4.6: Grading of modules which arrived at the Debug Centre either due to a too large number of defective channels or because of problems in the interpretation of test data for suspicious channels.

The supplements "n" and "s" at the geometry indicate that there exist modules without and with a stereo angle for ring 1, 2 and 5. The ring geometries 3, 4, 6 and 7 have a non-stereo version only.

to a significant fluctuation of the calibration signal test from its usual behaviour (e.g. bad fit results or fluctuations in the signal injection). After a detailed investigation contemplating more test data than used in the standard qualification algorithm it could be concluded that no defective channel was observed. The same statement holds for another 12% of the module. Here the sub-structure or the assigned test-setup were most likely responsible for the bad result. This explanation was confirmed because the module passed the qualification test in its production centre, only failed on the sub-structure, and passed again the test in the Debug Centre.

In a few cases the failure could be related to defective pipeline cells especially if a certain failure mode occurs only intermittent. In that case the readout trigger just randomly hits a noisy or dead storage cell of the data pipeline so that the signal and the test result are significantly influenced. In Figure 4.4 an example of the influence of a dead pipeline cell on the calibration signal amplitude is depicted.

Fig. 4.4

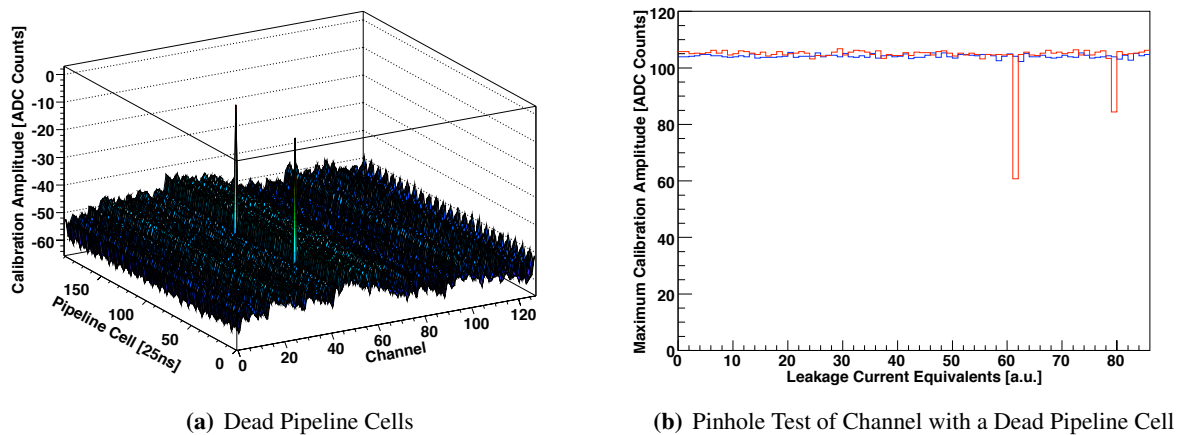


Figure 4.4: Calibration pulse amplitude distribution (negative amplitude) of an entire pipeline matrix showing two dead cells (a) and the resulting maximum pulse amplitude (b) as a function of the leakage current (pinhole test) for a regular (flat curve) and the affected (curve with spikes) channel. If a dead cell is read out the pulse amplitude is zero so that the kinks in the maximum pulse amplitude occur. Though there is more than one measurement per data point the number is still limited so that the contribution of the dead cell is not completely vanishing even if the average value is calculated.

For the rest of the modules either the number of bad channels could not be confirmed, so that there were set back to the production chain with a note to watch the test data. But of course, there have been modules which were disqualified because of a too large number of defective channels. They revealed many noisy channels probably due to the micro-discharge effect (chapter 4.2.1) or various other channels defects (chapter 3.5.2).

4.2.6 Inappropriate Handling

All modules which suffered from mechanical accidents were assigned to this category. This could be an irreparable damage of the module like the destruction of the silicon sensor which happened three times. But the overwhelming majority consisted of modules on which the very fragile electrical wire-bond connections were bent or broken (colloquially called pushed bonds). These accidents happened mainly during the assembly of already bonded components to larger sub-structures when the operator was not cautious enough when either picking up or mounting the component, i.e. either a FE-hybrid becoming part of a module or a module being mounted on its position on a sub-structure. The latter was the most frequent situation because space and sight were limited during this precision work.

Pushed bond are detected by means of a optical investigation and classified by the severity of the damage. A single affected bond can either be ripped off from its mechanical fixation point so that the electrical connection is interrupted. In that case the broken wire is removed and replaced by a new one. The wire can also have a partial crack only or it can just be bend so that the connection might only be weakened. In the latter case the wire is moved back to its original position and non-destructive pull-tests are performed to check if the connection is still strong enough. The wire must stand a force equivalent to more than six grams to be considered as good [45], otherwise it is re-bonded.

Ref. [45]

If more wires are affected it must further be guaranteed that they do not touch each other. This would short-circuit the underlying readout channel, and the bonds must be separated [46]. Figure 4.5(a) shows the various types of bonds on a module. Especially the wire-bonds close to the edges of the FE-hybrid or

Ref. [46]

Fig. 4.5(a)

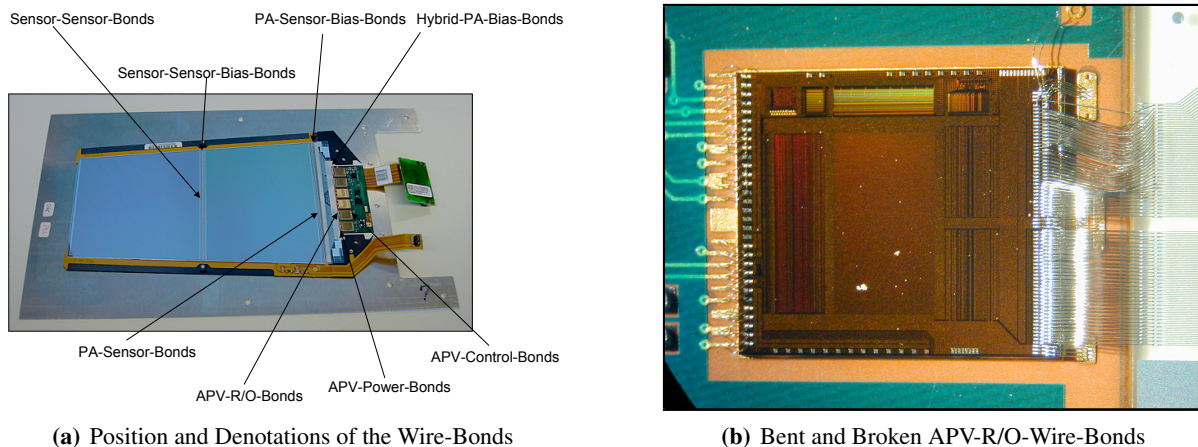


Figure 4.5: (a) Position and denominations of the different types of bonds. The high-voltage backplane-bonds (not visible on that picture) are placed on the backside of the module. (b) Example of an accident with about 40 bent or broken APV-R/O-bonds visible on the upper right side.

sensor(s) were affected because here the danger of touching them is much higher than for wire-bonds in the centre of the FE-hybrid or module. Accidents when a screw or a washer fell on the bonds occurred very rarely.

The statistics in Table 4.7 reveals the number of modules that were affected by that problem. This kind of defect proved to be the most critical one because of many reasons. At first, the number of modules suffered from various accidents was very high (more than 30%) compared to the total number of defective modules. Then the repairing was not possible in one centre let alone the Debug Centre. Due to the TEC module production strategy the hybrid-bonds were done at the CERN bonding centre while all other bond connections were done in other centres diversified by the module geometry. Modules had to be sent to the dedicated centre which was equipped with the machines and tools to repair the damage. Finally if a module arrived at the right place to be repaired it could not be guaranteed that a

Tab. 4.7

repairing is possible. In order to replace a bond a position on the bond pad must be found where no wire has previously been fixed because otherwise it would not be possible to re-establish the connection. Therefore every bond repair has to be done manually. This is a very time consuming procedure especially for the APV-R/O- and APV-Control Bonds because not only the bonds pads are very tiny but also the clearance between adjacent bonds is very small. The repairing of a damage of about 40 wire-bonds as shown in Figure 4.5(b) takes several hours.

Fig. 4.5(b)

Ring	1n	1s	2n	2s	3	4	5n	5s	6	7	total
Number of modules	3	4	1	1	9	3	6	15	23	19	84
Repaired	2	3	1	-	8	1	5	10	17	14	61
Recycled	1	-	-	1	-	2	-	3	3	3	13

Table 4.7: Number of modules with damaged wire-bonds. The second and third row quotes the number of repaired modules as well as the number of recycled module, where the damage was so severe that it has been too time-consuming or it could not be guaranteed that the quality of the repairing would be good enough. The increase for the rings 5, 6, and 7 can be explained by the fact that these module have two sensors and thus an additional row of wire-bonds located very close to the fixation points for the module on the sub-structure. Thus there is a higher risk to damage wire-bonds. The supplements "n" and "s" at the geometry indicate that there exist modules without and with a stereo angle for ring 1, 2 and 5. The ring geometries 3, 4, 6 and 7 have a non-stereo version only.

All modules which could not be repaired were considered for the sensor recycling in the recycling centre at Vienna. It can be seen from Table 4.7 that the majority of all modules were actually re-used. The remaining ten modules were not recycled. The majority of them suffered from another failure (mainly IV-problems) so that the sensor became useless for the module production.

Tab. 4.7

4.2.7 Other Defects

Besides the defects described above there were a few module which did not fit into that scheme.

- The ring 1, 2, and 5 modules were mounted pairwise back-to-back on the sub-structure. Once, there was an accident that destroyed the sensor of one module of such a pair. Therefore, the other was checked if it was affected, too. No damage was found.
- Three modules were sent to the Debug Centre because they could not be bonded. In all cases it was found that there has been a lack of glue to properly fix the sensor to the carbon frame. These modules were transferred to the production centre and the recycling centre, respectively, to rework or recycle the components.
- For another two modules there was absolutely no reason apparent why they were sent to the Debug Centre. These module were re-qualified and given back to the production chain.

In addition to the production modules there was an investigation of 29 module arriving from the petal reception test-setup [47]. This test-station of the TEC integration centre was used to prove the quality of the TEC sub-structures (petal) before these were integrated into the endcap structure. In order to build and to qualify this station a special petal was assembled. As it could not be guaranteed that these module were treated properly like the production modules mainly with respect to the supply voltages, and as this petal was extensively used for construction exercises, these modules were not used for the final tracker but rather to check if the sensors could be recycled. After a detailed optical investigation and a module qualification test which mainly focused on the quality of the sensors all but one were sent to the recycling centre.

Ref. [47]

4.3 The Final Statistics and Defect Correlations

The Figure 4.6 summarizes the key numbers of the investigation of production modules in the Debug Centre. The modules which were repaired (56%) are either redeclared as regular by the Debug Centre or another production centre was ordered to execute a certain repairing. About 15% of the modules were sent to the recycling centre. Having finished the mass production most of the modules which could not be repaired (26%) were sent to CERN for a final and centralized storage, or are used for demonstrations, exhibitions, and laboratory exercises.

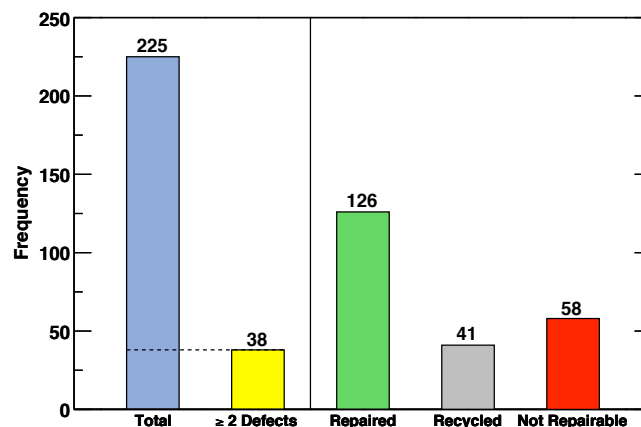


Figure 4.6: Final statistics about investigated, repaired and recycled module in the TEC Module Debug Centre.

Out of the total 225 modules there were 38 modules examined due to two or more problems (17%). The majority of these (26 modules) contained a damage from inappropriate handling in addition to one of the other failures. In some cases correlations between the different type of failure could be observed.

1. Touching APV-control bonds can influence the I²C-communication of the FE-hybrid if they create a short-circuit.
2. Any other mechanical damage to the wire-bonds can immediately affect the grading of a module. If wire-bonds are touching each other they can short-circuit adjacent readout channels, or if the wire-bonds connection is broken they result in a number of open channels.
3. The same statement holds for a bad high voltage contact at the backplane. Not only the IV-test reveals a suspicious IV-curve but also virtually all channels are flagged as noisy, because the silicon sensor is not completely depleted.
4. The micro-discharge effect can generate an IV-problem due to a strange shape of the leakage current. At the same time the readout channels which are positioned above this field deforming region can have a larger noise and thus the number of defective channels increases.

The last aspect to mention about repairing and recycling is that one should not neglect their economical benefit. Each repaired or recycled sensor saves about 1000 CHF, while the cost per FE-hybrid are estimated to be 100 CHF, let alone the human work that must be invested to assemble and qualify a module.

4.4 The Evaluation of the Procedure

The applied procedure was not perfect for several reasons. But before discussing possible improvements the main results of the TEC Debug Centre are summarized.

- No correlation between a certain stage of the mass production or an institute involved and a certain defect type was found.
- The investigations of the Debug Centre contributed a lot to the success of the mass production. More than one half of the modules could be repaired and used for the tracker integration.
- The problem of the pushed bonds which occurred on about one third of all investigated modules was constantly published until people were sensitized to that problem so that they started to handle the modules more carefully.

For a future mass production of silicon microstrip modules which involves the installation of a Debug Centre some ideas are presented in the following list to increase the efficiency of the investigations.

1. As there are many people and centres involved in the production chain of the modules it is very useful to nominate one responsible person per centre who should coordinate one centre and be capable of providing information to the collaborators. This person should carefully check if the agreed standard procedure is respected. This ensures that all available data are uploaded to the central database and the available forms are filled in to provide useful information to the people to whom the module is transferred. For instance, in a few cases the information about pushed bonds on the FE-hybrid were lost, that were already investigated at an earlier stage in the production chain. Hence, the same problem was inspected twice.

This statement does not only hold if the module is damaged. During a mass production the contemporary flow of information is absolutely essential.

2. The centre nominated to be Module Debug Centre should have more capabilities to repair modules. In principle, to send modules to one laboratory is a good idea from the point of view of book-keeping of all defects to spot systematic sources as early as possible. But even if the problem is identified the procedure is inefficient if people cannot start the repair actions due to the lack of equipment and detailed knowledge. Thus many transfers had to be done just in order to get the module to the right location.

In the second half of the mass production phase this procedure was altered to become more efficient and to reduce the time until the module returns to the production chain if it turned out to be repairable. Especially for the pushed bond problem, the collaborating institutes were asked to send the modules directly to the site where a repairing was feasible, with the drawback that a detailed global book-keeping was abandoned.

3. The agreed procedure in the second half of the production lead to another problem. Due to no better knowledge modules were often transferred to the wrong institution. Essentially the procedure to have several problems and about as many sites for repairing lead to some confusion. First and foremost when the difference between some repair actions were marginal (e.g. replacing wire-bonds of ring-6 and ring-7 modules) the procedure must be even more clearly communicated to the collaboration.

Therefore, to achieve an even more efficient Debug Centre the establishment of a centralized working group consisting of experts for all kind of module defects, fully equipped with the knowledge and tools

to take all measures to repair a module, would be worthwhile to be discussed in future if the problem of funding such an institution can be solved. This would be a transparent solution for the outside people whom to ask for the repairing, the exchange of information would have been facilitated, the global book-keeping would be regained, and time and costs would have been saved because of reduced transfer rates.

The above discussed TEC Module Debug Centre would have fitted excellently into such a project. The basic experience of this centre would have been able to cover at least all aspects that were related to module qualification tests with the ARC system.

Chapter 5

Events with Top-Quarks and Photons

AT the time when these lines are written the last parts of the CMS detector have been lowered to the underground cavern, the cabling of the detector is approaching its completion, the final and global commissioning of the detector has already begun, and the first injection of the proton beams is only a few days ahead. In preparation of the data taking period extensive studies have been and are still done using event generators to simulate as precisely as possible the proton-proton collisions, and the response of the CMS-detector to these physical reactions. The second part of this thesis contributes to these analyses in the realm of top-quark physics, investigating these type of events where a photon is radiated off the top-quark¹. Starting with a short physical introduction to that topic, the chapter continues describing the data-sets, framework, and tools that have been used to perform that analysis.

5.1 The Photon Radiation off Top-Quarks

In 1994, the discovery of the missing sixth quark was published to experimentally complete the three generation structure of the Standard Model in the quark sector [48]. Since then the experimental precision has been improved so that the latest mass measurement of the top-quark basing on up to 2.8 fb^{-1} of data from the Tevatron $p\bar{p}$ -collider gives a current world average of [49]

$$m_{top} = 172.4 \pm 0.7(stat.) \pm 1.0(syst.) \text{ GeV}/c^2. \quad (5.1)$$

While the relative precision of the mass measurement has reached the percent-level the properties of the top-quark like the electromagnetic charge q_{top} are still poorly investigated to prove that this particle is really the up-type partner of the b-quark with a value of $2/3$ for the electromagnetic charge quantum number². Although there is a strong belief that the particle discovered by the Tevatron collaborations is really the top-quark of the Standard Model there exists an alternative model in accordance with the electroweak precision data which published the first evidence for a heavy particle at a mass of about $170 \text{ GeV}/c^2$. These data are ambiguous because they fit to a heavy quark having a mass of more than $230 \text{ GeV}/c^2$, too. In that case the top-quark has not been discovered yet and the particle with a mass of $172.4 \text{ GeV}/c^2$ could be an exotic quark with an electrical charge quantum number of $-4/3$ belonging to a new fourth generation [50].

¹If not explicitly distinguished the charge conjugated particle, the antitop-quark, is always included.

²Throughout this thesis the value of the electromagnetic charge of the top-quark is referred to as $2/3$, for instance. This term is used as an abbreviation but it is not entirely correct because it represents the quantum number of the electromagnetic charge rather than the value of the electromagnetic charge itself. In order to obtain the latter the quantum number must be multiplied by the electromagnetic constant $e = 1.6 \cdot 10^{-19} C$.

The fundamental problem behind this question is that the mass analysis is based on events where pairs of top-quarks are generated. Each of them subsequently decays almost exclusively into a W-boson and b-quark, because the value of the element of the quark transition matrix (CKM-matrix) is almost equal to unity ($|V_{tb}| \approx 1$). As the studies do not distinguish between a b-quark and a $\bar{b}$ -quark, assigned to a t-quark and a $\bar{t}$ -quark if $q_{top} = 2/3$, it is possible to connect the b-quarks to the other top-quarks in each case, so that q_{top} could be $-4/3$. Meanwhile there has been a study trying to solve this ambiguity by measuring the charge of the b-jets, the streams of hadrons which are induced by the b-quarks. The exotic scenario is excluded at a confidence level of 92% by this indirect measurement [51] but no direct measurement of the electromagnetic charge of the top-quark has been performed so far.

Ref. [51]

An alternative to directly measure the electromagnetic charge of the top-quark is to determine a photon spectrum, which is enriched by a significant number of photons that are radiated off top-quarks. In that case the height and the shape of the spectrum depend on the coupling between the photon and the top-quarks so that it contains information not only on the electromagnetic charge but potentially on properties of the top-quark like the anomalous magnetic moment. Figure 5.1 depicts an example of two different coupling scenarios yet neglecting any influence of a detector.

Fig. 5.1

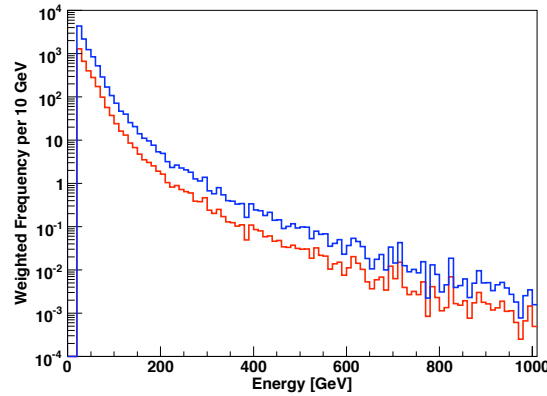


Figure 5.1: The photon spectra for two different coupling scenarios of the photon to a top-quark with $q_{top} = 2/3$ (lower curve) and with $q_{top} = -4/3$ (upper curve). The spectra are theoretical predictions not yet respecting any influence of a detector. A detailed discussion of this matter will be presented in chapter 10.2.

But as this radiation process is suppressed by the magnitude of the QED-coupling constant, and the radiation off lighter particles is more favoured than that off top-quarks, one needs a high statistics to identify a spectrum which is dominated by photons radiated off top-quarks. At a centre-of-mass energy of 14 TeV and the initially expected luminosity of $2 \cdot 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ during the first physics run about one $t\bar{t}$ -pair per second is expected to be created ($\sigma = 830 \text{ pb}$, NNLO). This high event rate should principally generate sufficient events. Thus, it is worthwhile to start an analysis and to search for photons that are radiated off top-quarks.

5.2 The Signal Process

Fig. 5.2(a)

The signal production process can be described by two types of Feynman diagrams (Figure 5.2(a) and Figure 5.2(b)). Each diagram represents a class of eight graphs, assuming that the photon is radiated off the top- or the antitop-quark, and the quark-pair in the initial state of the annihilation process are either u- or d-quarks. The gluons and quarks from the initial states stem from the colliding protons according to the parton density functions of the proton.

Fig. 5.2(b)

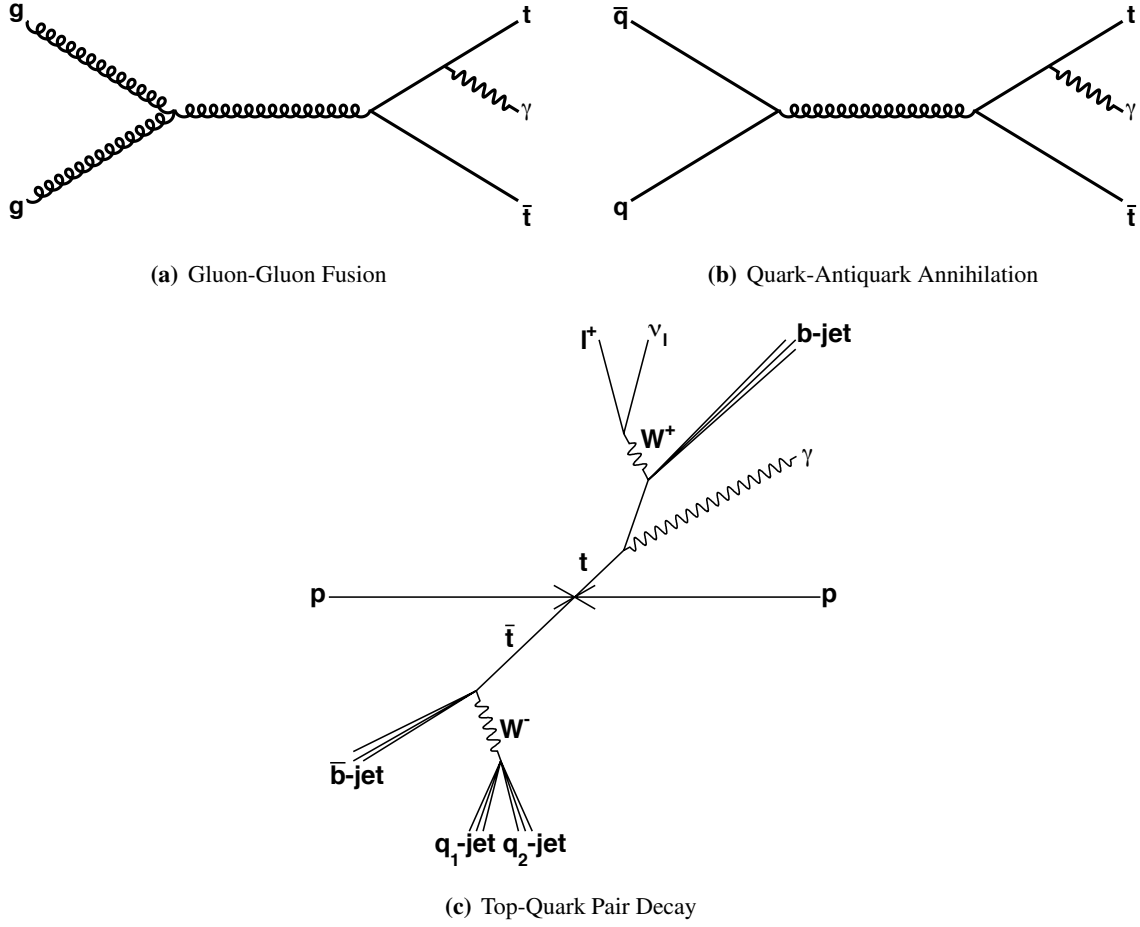


Figure 5.2: Production of $t\bar{t}\gamma$ events in the (a) gluon-gluon fusion and (b) quark-antiquark annihilation process. (c) Decay chain of the top-quark pair sketching which objects enter the detector. The proton remnants are not shown.

The top-quarks of the final states are not stable and each decays after only $4.23 \cdot 10^{-25}$ s into a W-boson and a b-quark, while the W-bosons decay immediately ($\mathcal{O}(10^{-25}$ s)) into two fermions ($f f'$ -pair). The present analysis of $t\bar{t}$ -pairs considers only these decay chains where one W-boson decays into an electron or muon and its partner neutrino, and the other one decays into a pair of light quarks, the *semileptonic* (lepton + jets) channel depicted Figure 5.2(c). The branching ratio of this process is approximately 29.6% Fig. 5.2(c)

$$pp \rightarrow t\bar{t}\gamma \rightarrow W^+ b W^- \bar{b} \gamma \rightarrow l \nu_l q_1 q_2 \bar{b} \bar{b} \gamma. \quad (5.2)$$

When the quarks of the final state have fragmentated and hadronized into collimated streams of particles, i.e. *jets*, the following signal signature is to be observed by the CMS detector:

1. one electron or muon,
2. four jets (two from b- quarks and two from light quarks),
3. one photon, and
4. missing transverse energy due to the neutrino which escapes undetected.

The advantage of this $t\bar{t}$ -pair decay channel is the trade-off between the purity of the signature, the lack of kinematic information due to the neutrino which cannot be detected and the event rate. Compared to the all-hadronic channel, here both W-bosons decay into light quarks, kinematic information is lost because of the undetected neutrino and the event rate is lower. But on the other hand the number of jets is lower as well, so that the signature is better defined and therefore expected to be more easily separated from background events. The dilepton channel, where both W-bosons decay into lepton pairs, is even cleaner than the semileptonic channel, but here two neutrinos are present in the final state making the kinematical event reconstruction more difficult. In addition, the event rate is lower by a factor of two to four depending on the types of leptons which are taken into account.

5.3 The Monte Carlo Event Generation

As the LHC has not produced collisions so far and thus the CMS experiment has not yet recorded data the analysis is based on computer simulated event data. In order to address the quantum mechanical nature of particle collisions and their succeeding decays a set of random number is created to describe probability functions representing physical variables or distributions (e.g. parton density functions or decay probability functions). These values, starting from the extraction of the parton inside the proton, via the calculation of the matrix element of a chosen hard process, down to the decay of the last short-lived particle, are merged together to form what is called a *simulated event*.

It is obvious that there is a discrepancy between the true behaviour of nature which contains all available physical processes to its correct precision and the implementation of these processes in a computer program. The latter is founded on quantum mechanical perturbation calculations up to a certain order and phenomenological models as well, thus predicting an approximation of the real process only. But on the other hand the weakness of these Monte Carlo (MC) generators is their strength as well. As any physical model or process can be implemented in these packages and studied, the predicted results can be compared to experimental data to a certain accuracy and it can be tried to uncover which physical configuration nature has chosen.

The most commonly used Monte Carlo generator providing a large variety of physical processes in leading order is called *Pythia* [52] which has been used for the signal generation of this analysis. Furthermore some of the background processes have been generated with the *Alpgen* package [53]. The choice of two generators has been made because the description of processes which include additional jets in the final state is more accurate concerning the transverse momentum of these jets in Alpgen compared to Pythia³.

Ref. [52]

Ref. [53]

The Signal Events

As mentioned above the $t\bar{t}\gamma$ signal sample has been created with the Pythia generator. The basic process has been a proton-proton interaction at a centre-of-mass energy of 14 TeV creating a $t\bar{t}$ -pair. The *top photon*, i.e. a photon that is radiated off the top-quark, is produced by final state radiation. Soft and collinear divergences present in any photon radiation process are handled by the algorithm suppressing the radiation if the momentum scale reaches a lower limit of 1 GeV/c.

In preview of the actual analysis, the signal events must fulfill a minimum set of requirements already at generator level. In order to find robust criteria to distinguish top photons from any other photon or photon-like object events enriched with these top photons are needed. As the process of final state

³This is due to the fact that the Alpgen generator includes the partons from which the jets are originating in the matrix element of the hard process, whereas the Pythia generator treats them as final state radiation which is factorized from the hard process.

radiation is also subject to a statistical process, especially the radiation off heavy particles is less favoured than that off lighter ones. Therefore, events are accepted only if they contain a photon that is radiated off a top-quark, of which the energy exceeds a minimum of 10 GeV, and if it is generated within the η -acceptance of the electromagnetic calorimeter, i.e. $|\eta| < 2.5$. These cuts are chosen because the photons must be reconstructed in the instrumented region of the CMS detector and separated from the majority of low energetic photons and detector noise fluctuations. A dedicated data filter checks every event for these criteria. Approximately 6.8 millions events must be generated to get 10000 signal events. Those consist of about 5.8 millions of the gluon-gluon fusion type (Figure 5.2(a)) and only 1 million of the $q\bar{q}$ -annihilation type (Figure 5.2(b)). Demanding only semileptonic (electron and muon) events the cross-section for this signal process is

Fig. 5.2(a)

Fig. 5.2(b)

$$\sigma_{Signal} \approx 218 \text{ fb.} \quad (5.3)$$

In total a signal data-set of 9912 events where the lepton is an electron, *signal electron events*, and 10038 events where the lepton is a muon, *signal muon events*, has been generated. Among these there are 37 out of 19950 events with more than one photon radiated off one of the top-quarks. But this multiple radiation can be neglected because all of the additional photons have an energy which is lower than 10 GeV.

In order to validate the generator output several kinematic distributions have been checked. Figure 5.3 reveals that the generated top-mass apart from statistical fluctuations peaks at its input value of 175 GeV/ c^2 and is smeared by a Breit-Wigner-distribution having an RMS of 1.7 GeV/ c^2 . The same has been studied for the W-mass and the distribution of the energy of the top-photon. The latter proves that the filter has been correctly implemented because no photon is generated below 10 GeV and the shape of the distributions is parametrized by an expected E^{-1} -power law of the final state radiation.

Fig. 5.3

The Estimation of Interference Effects in Signal Events

This subsection is devoted to the question if the Pythia signal process is close enough to reality. The final state radiation in the Monte Carlo generator does not respect any interference effects in the generator between the signal process and events where a photon is radiated off one of the charged decay products of the top-quarks. The following calculation estimates the size of the interference effect and shows that it can be neglected if the phase space of the events is properly adjusted by kinematic cuts. The graphical representation of the processes can be seen in Figure 5.4. The outgoing particles are described by their Lorentz-vectors⁴

Fig. 5.4

$$\begin{aligned} \text{t-quark} \quad p_t &= (p_t^0, \vec{p}_t) \\ \text{b-quark} \quad p_b &= (p_b^0, \vec{p}_b) \\ \text{fermion } f_j \quad p_j &= (p_j^0, \vec{p}_j) \quad j = 1, 2 \\ \text{photon} \quad k &= (k^0, \vec{k}). \end{aligned} \quad (5.4)$$

For each of the diagrams the final state particles can be summed up and the energy-momentum conservation at the top-quark decay vertex gives

⁴The velocity of light is temporarily set to 1. To retain SI-units each 3-vector $\vec{p}$ must be replaced by $c\vec{p}$.

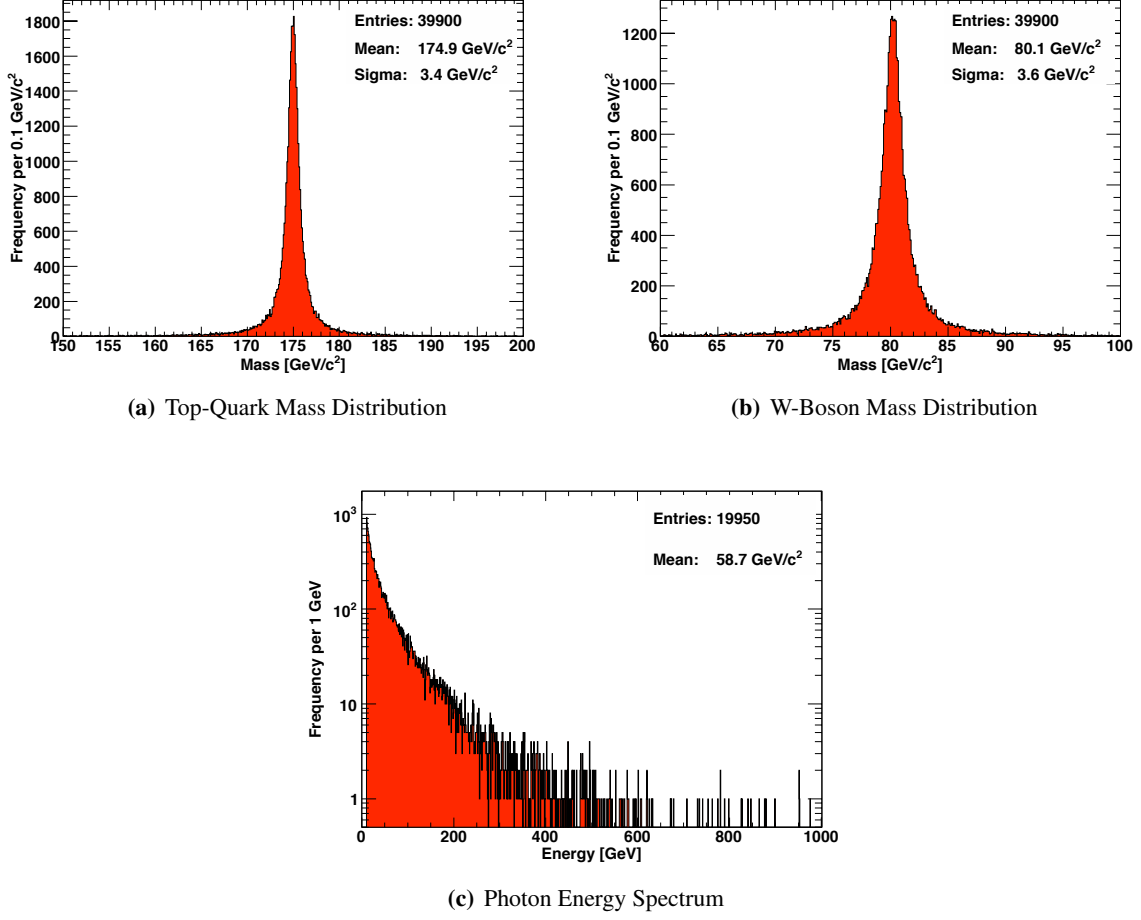


Figure 5.3: Mass distributions of the top-quarks (a) and the W-bosons (b) peaking close to their input values of $175.0 \text{ GeV}/c^2$ and $80.2 \text{ GeV}/c^2$. The distributions of the top- and anti-top-quarks as well as that of the W^+ - and W^- -boson are merged in one histogram. The energy spectrum of the top-photons is shown in (c) proving the E^{-1} -power law of the final state radiation.

$$\begin{aligned}
 m_t^2 &= \underbrace{(p_b + p_1 + p_2 + k)^2}_{=p \text{ with } p^2=M^2} \\
 &= M^2 + 2pk \\
 &= M^2 + 2p^0 k^0 - 2\vec{p}\vec{k} \\
 &= M^2 + 2p^0 E_\gamma - 2|\vec{p}|E_\gamma \cos \varphi \quad \text{with } \varphi = \angle(\vec{p}, \vec{k}) \\
 &= M^2 + 2E_\gamma \left(p^0 - |\vec{p}| \cos \varphi \right)
 \end{aligned} \tag{5.5}$$

If the photon is radiated off one of the top-quark decay products the mass of the top-quark can be calculated from the invariant mass of the three-particle-system plus a correction which depends on the energy of the photon and the angle between the three-particle-system and the photon. This correction is smallest if assuming a non-vanishing but fixed photon energy and $\varphi \rightarrow 0$. In that case one gets

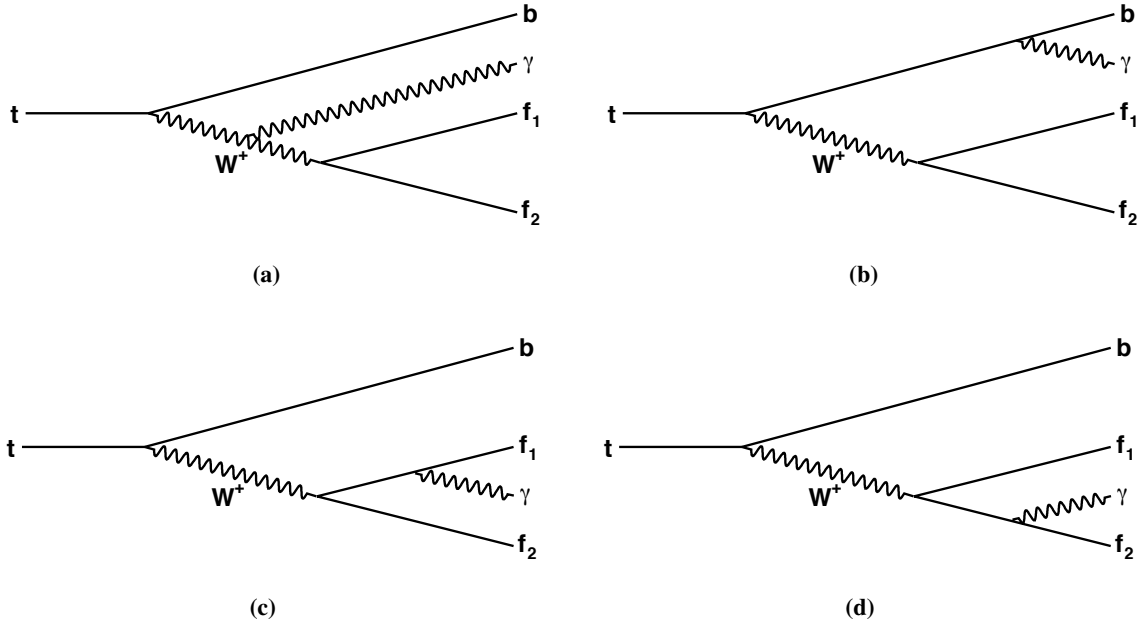


Figure 5.4: Feynman diagrams of a decaying top-quark where a photon is radiated off one of the final state particles. They diagram (d) does only exist if the second fermion (f_2) is a charged particle rather than a neutrino of a leptonic W-boson decay.

$$\begin{aligned}
 m_t^2 &= M^2 + 2E_\gamma(p_0 - |\vec{p}|) \\
 &= M^2 + 2E_\gamma \frac{M^2}{p_0 + |\vec{p}|} \\
 &= M^2 \left(1 + \frac{2E_\gamma}{p_0 + |\vec{p}|} \right)
 \end{aligned} \tag{5.6}$$

If the invariant mass of the three-particle-system approaches the mass peak of the top-quark one must take care of interference effects. This only happens if one of these two kinematic relations is true.

$$\Rightarrow M \rightarrow m_t \quad \text{if} \quad \begin{aligned} (1) & \ E_\gamma \rightarrow 0 \\ (2) & \ p_0 + |\vec{p}| \gg E_\gamma \end{aligned} \tag{5.7}$$

Before discussing the influence of these two cases on the physics process a numerical example is presented for $E_\gamma = 20 \text{ GeV}$ and $p_0 + |\vec{p}| = 1000 \text{ GeV}$.

$$\frac{M}{m_t} \approx 0.98 < \frac{m_t - \Gamma_t}{m_t} \approx \frac{175 \text{ GeV} - 1.5 \text{ GeV}}{175 \text{ GeV}} \approx 0.99 \tag{5.8}$$

In Figure 5.5 three different cases for the inference effects are qualitatively discussed in the $(p^0 + |\vec{p}|) - E_\gamma$ -space. If the energy of the photon is low (lower band, first case of equation 5.7) or the three-particle-system is very high-energetic compared to the photon (upper right area, second case of equation 5.7) the interference effects cannot be neglected. In order to reduce the impact of the first one a minimum energy of the photon is required. The second one is less critical because if demanding a minimum energy of the photon the three-particle-system must have an even higher energy. But for that region of phase space the event rate decreases and so the admixture to the remaining events is marginal. This has also been pointed out by the numerical example in equation 5.8. Even for that example with a photon at the

lower edge of the spectrum and a relatively high-energetic three-particle-system the overlap between the invariant mass M and the top-quark mass only starts to become relevant. All other configurations with a higher-energetic photon or a lower-energetic three-particle-system shift M further away from m_t and the interference effects can be neglected (upper left area). But this is the region of phase space which is populated by the signal events.

This estimation can become invalid if the top-quark has drifted away from its mass shell after the radiation process. But this case it not relevant for the present analysis because the Pythia generator has only included top-quarks that remain on-shell after the photon has been radiated off. This lack of physical precision can be overcome by choosing a more appropriate Monte Carlo generator for the signal events in future which can be integrated into the entire simulation chain.

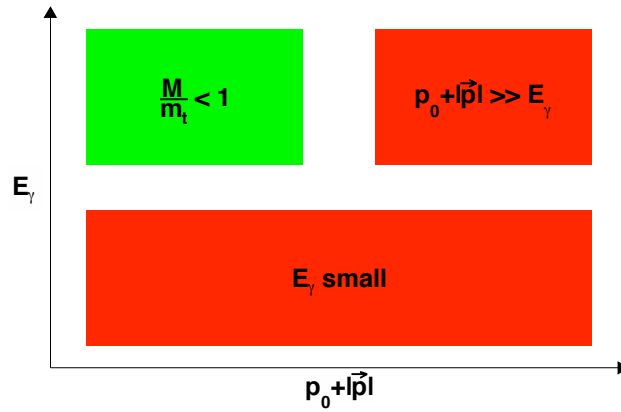


Figure 5.5: The $(p^0 + |\vec{p}|) - E_\gamma$ -space can be separated into three regions. These parts which would contribute to the interference can either be cut out (lower band) by a minimum energy of the photon, or contribute only marginally because of a low event rate (upper right). For the remaining area (upper left) the interference effects can be neglected.

Background Events

At first, a background event is an event which mimics the signature of a signal event. But then, these events must be taken into account as well, which have a sufficiently large cross-section so that a non negligible number of events can potentially degrade the final results because a few events might pass the event selection. Analyzing the signal process in equation 5.2 the following aspects must be considered.

1. The jets can be easily generated in a hadron environment either by initial or final state radiation, by the proton remnants that did not take part in the hard collisions, or by peripheral collisions. Thus, no special reaction or decay is needed producing these constituents of the process.
2. The photon can in principal be radiated off any charged particle in the detector. Thus, there is a priori no necessity that it is created in the hard collision. Furthermore, a photon can be faked by an electron or jet, because all three object are calorimetric objects.
3. The neutrino and its missing transverse energy can be simulated by each object that is not fully registered in the detector because it is close to the edges of the acceptance regions, due to any other detector insufficiency or detector resolution effect.
4. The lepton is the most crucial ingredient of the signal process produced in a hadronic environment. It must be created in a hard interaction or originate from a leptonic decay of a b-meson. In addition,

it can be rarely faked by a jet or photon, so that pure hadronic processes must be considered having a very high cross-section.

According to these prescriptions the background samples listed in Table 5.1 have been analyzed. In Figure 5.6 the cross-sections are compared to each other normalized to the signal cross-section. It can be seen that the background suppression must succeed to more than ten orders of magnitude for the worst case, but the background ($t\bar{t}$ + Jets) resembling the signal to the highest extent is larger by more than three orders of magnitude as well.

Sample	Cross-Section [pb]	MC Generator
$t\bar{t}$ + Jets	$5.610 \cdot 10^2$	Alpgen
W + Jets	$4.147 \cdot 10^4$	Alpgen
Z + Jets	$3.426 \cdot 10^3$	Alpgen
WW	$1.143 \cdot 10^2$	Pythia
WZ	$4.990 \cdot 10^1$	Pythia
ZZ	$1.610 \cdot 10^1$	Pythia
$W\gamma$	$4.470 \cdot 10^0$	Pythia
γ + Jets	$1.060 \cdot 10^5$	Pythia
QCD-multijet	$5.748 \cdot 10^{10}$	Pythia

Table 5.1: List of background data-sets with their cross-sections and the MC-generator used for the event generation. They must be compared to the signal cross-section of 0.218 pb (see equation 5.3).

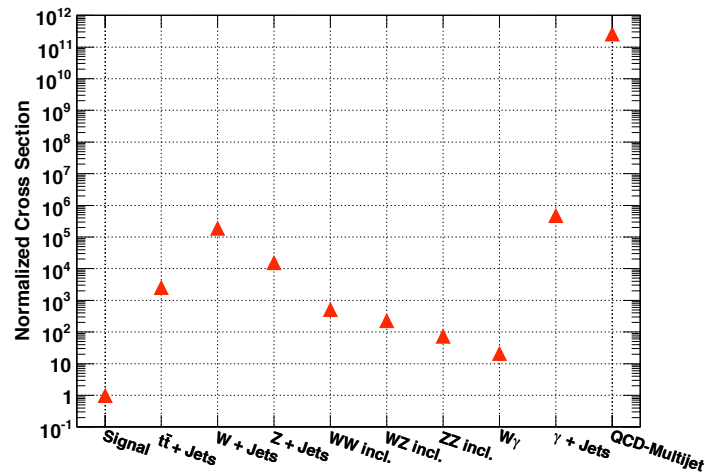


Figure 5.6: Comparison of the cross-sections normalized to the signal cross-section of 0.218 pb.

The data-sets which have been taken from an official CMS production cycle called *Spring 07* can be separated into four classes. The first consists of the $t\bar{t}$ -events which incorporate not only the all-hadronic and dileptonic decay channels but also the semileptonic events which do not explicitly have a top photon. Except for the photon the signature of the latter decay channel matches perfectly to the signal events. The second class comprises the single boson processes. At least one high-energetic lepton might come from the decay of the W- or Z-boson. The number of jets also meets the requirements especially if the

number due to additional radiation processes becomes at least four. The statement with respect to the lepton holds again for the di-vector-boson class. But here the jets can also come from the decay of the second vector-boson (except for the $W\gamma$ case). The γ + Jets and the QCD-multijet classes have mainly to be considered because the cross-section exceeds the signal process by several orders of magnitude. Even if the fake rates for leptons and photons are potentially low these samples can become relevant for the background contributions.

All signal and background data-sets are optimistic in the sense that they miss the multiple-proton interaction per bunch crossing (in-time pile-up). This is due to the fact that the necessary events were not available at the time when the data were generated. In addition, event trigger information are not available. But the latter has been overcome by kinematic cuts that are even tighter than the trigger thresholds. As it will be shown in chapter 7.2.2 each event is required to contain a high-energetic electron or muon with $p_T > 30 \text{ GeV}/c$. In contrast, the threshold for triggering events with an isolated electron is $p_T > 29 \text{ GeV}/c$ and for an isolated muon $p_T > 19 \text{ GeV}/c$. Hence, trigger thresholds will intrinsically be respected by the selection cuts, so that the final event selection will not be reduced further by trigger efficiencies.

Chp. 7.2.2

5.4 The Event Generation and Reconstruction

Chp. 5.3

As mentioned in chapter 5.3 the analysis of collision data can be prepared by generating events with a Monte Carlo program. But when neglecting the uncertainties imposed by the underlying physics model and the computational restrictions, this represents a perfect world in the sense that all particles (kinematics and properties) are very well known and any measurement of a physical observable can be performed to an infinite precision. To compensate for that especially at a time when the real detector is not running not only the physical process but also the propagation of an event through the detector has to be simulated by computers. A flow chart indicating the functional principle of this simulation chain is shown in Figure 5.7.

Fig. 5.7

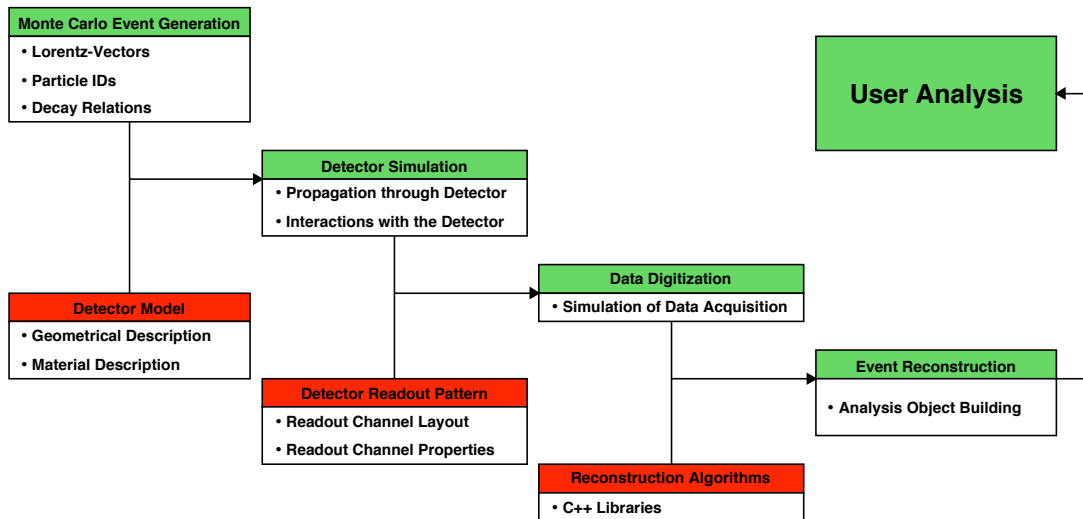


Figure 5.7: Flow chart of the main steps of the event generation and reconstruction chain. A trigger stage to preselect an event due to a certain characteristic signature which is not included in the present analysis would be inserted right before the event reconstruction.

At the beginning of the event simulation sequence which is divided into four main stages a Monte Carlo generator is let run. The output is a list of particles containing information about the kinematics of the particles by means of Lorentz-vectors, their identities, and all decay relations. The second step is to process that list into the detector simulation based on the Geometry and Tracking package (Geant 4) [54]. This is configured with a geometrical and physical model of the detector. The interactions of the particles with the detector material are simulated to derive in which regions of space and how a particle would be detected. It describes how a particle changes its energy and momentum if it is traversing through the detector including the effect of the magnetic field. During that procedure some additional particles can be created like photons via bremsstrahlung which are appended to the list of particles that pass the detector to give a realistic picture of the detector response. By that simulated propagation a picture is formed at which point of the detector the energy of a particle is released due to the interaction with the detector material. Ref. [54]

These points are then overlaid by the detector readout pattern, to get information in which channels a signal is observed. In the following procedure these signals are treated as if they would be real data. That means that they are digitized and the properties of the readout-electronics are respected. For example, certain thresholds for a minimum signal must be exceeded to suppress the noise of the readout channels. This considerably reduces the amount of data that must be stored. Now the reconstruction algorithms start producing objects suitable for the final and individual physics analysis. For the data output the ROOT-framework [55] is interfaced, such that all required event data are stored in ROOT-trees, the modern n-tuple format. The physical results of these detector-simulated data are now resembling to a large extent the real data which will be produced by actual proton-proton collisions. The entire chain for a $t\bar{t}\gamma$ -event takes in the order of minutes to run a single event on a modern computer (AMD Athlon 3500 MHz, 1 GB RAM). The disk space usage is about 1.5 MB per event. Ref. [55]

The CMS Software

All the above described different routines accessing external programs and libraries, running the detector simulation, the digitization and reconstruction are developed within a general software framework, to which the user analysis can be added. In case of the CMS experiment it is called *CMS SoftWare* (CMSSW). It is written completely object-orientated in the C⁺⁺-programming language. The version CMSSW 1.3.6 has been used for this analysis including a package for the electron re-reconstruction to fix a programming bug for electrons in the endcaps of the electromagnetic calorimeter.

The software provides general functionalities like data input and output, but in general the reconstruction and analysis steps are structured in modules. As it will be seen in chapter 5.6 for the $t\bar{t}\gamma$ -analysis, a module represents a logical unit of the program chain. Though some modules depend on others, i.e. for instance, the object building is only possible after the hits in the detector have been reconstructed, this structure allows to insert or append modules that are specifically programmed for a certain analysis. A configuration file for each module offers the opportunity to pass external parameters into the program chain. These configuration files can be nested such that there is a global configuration file which defines if and in which time order the modules are executed. An example of such a configuration file is attached in appendix B. Chp. 5.6

As many of the CMSSW-routines are integrated in a steadily evolving process, the documentation concerning the software or the analysis objects are either extracted from a TWiki-webpage-project⁵ ⁶ or from the source code itself⁷. App. B

⁵The CMS Offline WorkBook: <https://twiki.cern.ch/twiki/bin/view/CMS/WorkBook>

⁶The CMS Offline Guide: <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuide>

⁷CMSSW Software Cross-Reference: <http://cmslrx.fnal.gov/lxr/>

Additional Aspects of the Generator Studies

There are some additional aspects why there is a special interest to prepare the analyses with generator studies. The entire computational infrastructure, i.e. data storage, data transfer, and data analysis in institutes world-wide must be tested under conditions as realistic as possible [56]. Furthermore, the reconstruction algorithms can be validated by comparing the results to the information given by the Monte Carlo events. Then, the influence of the detector on a physical observable can be predicted. As an example the material budget of the silicon tracker derived from its geometrical description is shown in Figure 5.8. This is an important distribution if one wants to predict the conversion rate of photons into electron-positron pairs. This type of information is relevant when designing the detector to pre-estimate to which extent an observable is altered and thereby limited by the detector. Finally, if later the real collision data are compared to the prediction from the generator studies it becomes evident that the geometric model has been correctly implemented or that the latter needs to be modified to more precisely predict the data output of the experiment. If two new physical models should be distinguished it is important to understand their sole difference with respect to the detector influence.

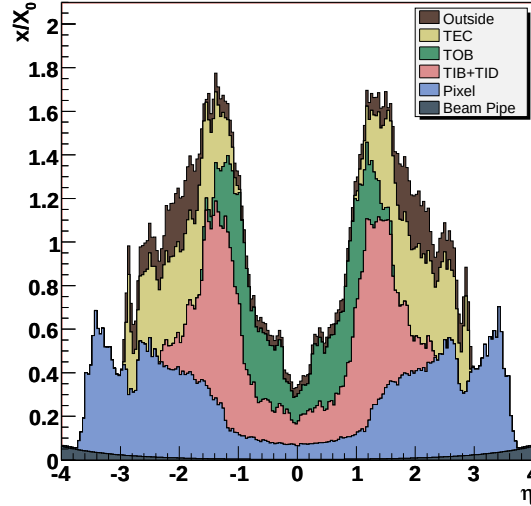


Figure 5.8: Material budget of the silicon tracking detector. The histogram reveals the contributions of the different sub-detectors as a function of the η -acceptance in terms of the radiation length X_0 . This is an equivalent after which distance a photon has converted into an electron-positron-pair with a probability of approximately 63% [23].

5.5 The High-Level Objects

A *high-level object* is a reconstructed object that has been built by the software algorithms. Data of several sub-detectors are combined to create a list of candidates having left similar traces in the detector. They serve as the input to select the objects more tightly according to an expected underlying physical process. This emphasizes the importance of a Monte Carlo analysis because selection criteria cannot be derived without knowing the origin of a reconstructed object. In the following information on the reconstruction algorithms of the most important standard physics object candidates incorporated in the $t\bar{t}\gamma$ -analysis are summarized to understand at which point the user analysis.

Reconstructed Tracks

Any charged particle that is traversing the silicon pixel and silicon strip tracker creates a trace of hits along its path. Starting from a seed in the pixel or strip tracker the tracks are fitted iteratively from layer to layer by the combinatorial track finder. In order to find a valid track some quality criteria must be met. The most important one is the minimum number of reconstructed hits per track which must be five or greater. Every hit that fits to a track must not deviate too large ($\chi^2 < 30$), while the number of missing hits, i.e. the track is crossing a sensitive area without seeing an expected hit, must not exceed one. Finally to keep the track collection at a manageable size the transverse momentum must fulfill $p_T > 0.9 \text{ GeV}/c$ for each track that is stored in the data file. Further details on the track-finding procedure can be found in [57].

Ref. [57]

Though a track is not a high-level object in its sense it provides an essential information with respect to all other objects, because these are either assigned to tracks (electrons, jets and muons) or explicitly not (photons). Thus, the most important information on a track are the direction in which it is pointing, and the momentum of the particle that has created that track.

The tracks which are used for this analysis are called *CTFwithMaterialTracks*.

Reconstructed Photons and Electrons

The reconstruction of photons and electrons starts with clustering the cells of the electromagnetic calorimeter to basic clusters and super-clusters, and then the energy of the super-clusters is rescaled [58].

Ref. [58]

Two types of algorithms are used for building the basic clusters. For the *island algorithm* (see Figure 5.9) every cell, that exceeds a transverse energy of $500 \text{ MeV}/180 \text{ MeV}$ in the barrel/endcaps of the ECAL, serves as a seed for the cluster algorithm. It integrates the energy of all adjacent cells until the next cell is either a hole, i.e. a cell with no energy deposition after noise suppression, or if the energy increases compared to the last one added to the cluster. Inspired by recollecting the energy that is taken away from electrons by bremsstrahlung and spread along the ϕ -direction these basic clusters are combined to super-clusters starting with a seed of 1 GeV . All basic clusters in a rectangular but fixed $(\eta-\phi)$ -segment around the main cluster seed are integrated which do not necessarily have to possess adjacent cells.

Fig. 5.9

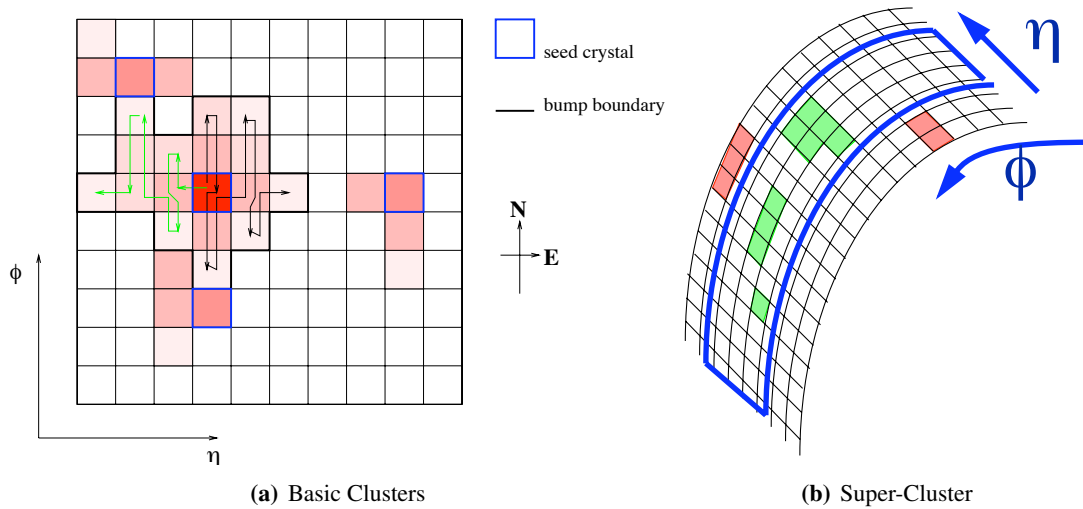


Figure 5.9: Island Algorithm: Combining the cells of the electromagnetic calorimeter to basic clusters (a). All marked basic clusters within a fixed rectangular in the $(\eta-\phi)$ -plane are integrated to a super-cluster (b) [58].

Fig. 5.10

The second algorithm, the *hybrid algorithm* (see Figure 5.10), is looking for seed cells in the barrel region with a transverse energy greater than 1 GeV. Then either the next or the next two neighboring cells, called 1×3 - or 1×5 -dominoes, along the η -direction are added to that seed cluster. In the following the algorithm is scanning a fixed additional number of these dominoes aligned in both ϕ -directions. If these stripes exceed $E > 100$ MeV the domino is kept. This might give a number of again probably separated sub-clusters which are the constituents of the super-cluster, if each sub-cluster has a central seed with $E > 350$ MeV. Thus, the hybrid algorithm generates by definition super-clusters.

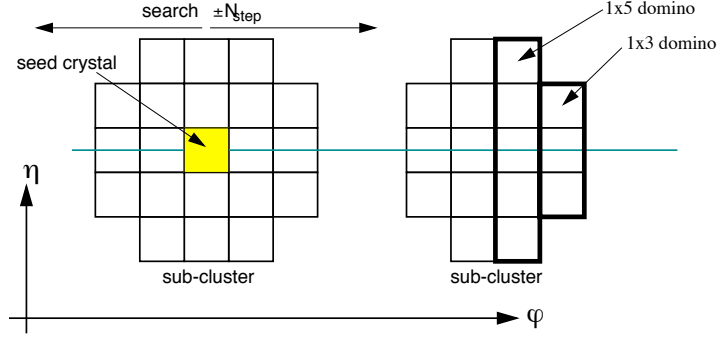


Figure 5.10: Hybrid Algorithm: The search path is extended to the right side, along a fixed number of strips in the ϕ -direction adding 1×3 - or 1×5 -dominoes to the sub-clusters [58].

After applying an energy and position correction all super-cluster become photons candidates giving a collection named *corrected photons* that has been used for the photon studies. Besides the usual data on kinematics, information about the shower shape, i.e. the energy deposited in squares or rectangles of electromagnetic calorimeter cells, and the potential matching to a track seed in the pixel detector are stored for any photon object.

This pixel seed is important for the upgrade of super-clusters to electron candidates because it represents the starting points for a reconstructed track. If a track is found that fits to the super-cluster the photon candidate is nominated as a new electron candidate when a few loose identification criteria are fulfilled. These are

- a minimum transverse momentum of 5 GeV/c,
- a track match within $\Delta\eta < 0.05$ and $\Delta\phi < 0.1$ and $E_{\text{Super-Cluster}}/p_{\text{track}} < 3.0$, and
- the hadronic energy behind the super-cluster is not allowed exceed 20% of the electromagnetic energy in a cone of $\Delta R=0.2$. Ideally it should be zero.

Ref. [59] Precise descriptions for the electron reconstruction can be found in [59]. The electron candidates that have been used for the present analysis are stored in a collection called *PixelMatchGSFElectron*.

Finally, it is worthwhile to be repeated and explicitly emphasized that every super-cluster becomes a photon candidate. It is not removed from that list if it is promoted to an electron candidate.

Reconstructed Jets

The geometry of the hadronic calorimeter is adapted to the electromagnetic one such that one hadronic tower is precisely positioned above a square of 5×5 electromagnetic calorimeter cells. These compositions are called *calorimeter towers* and enter the algorithms for the jet reconstruction [60]. For the CMSSW-project three algorithms are implemented, the iterative cone algorithm [61], the midpoint-cone algorithm [62] and the k_T -algorithm [63]. The first two sum up the energy of all towers iteratively

Ref. [60]

Ref. [61]

Ref. [62]

Ref. [63]

within a predefined cone in the $(\eta-\phi)$ -plane around seed towers until the jet middle axis and the energy converge to stable values. While the iterative cone algorithm does not produce overlapping jets, because a calorimeter tower is always assigned to one single jet only, the midpoint-cone algorithm incorporates a routine to split and merge jets which are overlapping in the $(\eta-\phi)$ -plane. In contrast, the k_T -algorithm is based on a procedure merging these calorimeter towers to a jet which are close in the E_T -weighted $(\eta-\phi)$ -plane. Thus, the cut-off is defined by a weighted distance in that plane. By definition there does not exist an overlap between the jets, and the edges of the jets in the $(\eta-\phi)$ -plane are not circular-like but can be aligned along any line. In the end the energy of the reconstructed jets is calibrated by means of an energy- and η -dependent correction function which is derived from $\gamma + \text{Jets}$ -events [64].

Ref. [64]

In the present analysis the jets reconstructed with the iterative cone algorithm and a cone size of $\Delta R = 0.5$ are used. In order to be considered as a seed for a jet the calorimeter tower must exceed a transverse energy of 1 GeV and all contributing towers must have $E_T > 0.5$ GeV. Apart from the basic kinematic information, further data especially on the fraction of the electromagnetic and hadronic energy are saved for a jet object.

Again the difficulty to distinguish between photon and jet candidates becomes evident. As the electromagnetic cells are parts of calorimeter towers it is possible that an object is stored not only in the list of photon candidates but also in the jet collection.

Muons

Muons are reconstructed by their tracks in the muon system and in the silicon tracker. A three-stage approach is followed which starts with the reconstruction of track segments in the drift tube chambers of the muon barrel and the cathode strips chambers of the muon endcap region. As the single detector cells of the muon system are combined to larger sub-structures (layers, super-layers, and chambers) a combinatorial pattern recognition algorithm fits the first parts of a muon track within such a super-layer or a chamber. These track segments, if they are constructed from at least three points, do not share more than one point with another segment and point to the primary vertex, and the reconstructed hits from the resistive plate chambers serve as the input for the following standalone-muon reconstruction. The result of that second step are tracks using the full information of the muon system. Finally, to achieve the best resolution the standalone-muons are matched to a track which has been recorded by the silicon tracker. This gives the *global muons* which have been used in this analysis. The information stored for the global muons are the same as for standard tracks, apart from the fact that the particle has already been identified as a muon candidate.

A description on the muon reconstruction can be found in [22] and more details on the local reconstruction in [65].

Ref. [65]

Multiplicity of Reconstructed Object Candidates in $t\bar{t}\gamma$ -events

The Figure 5.11 reveals the number of reconstructed object candidates to understand how many of the high-level objects described in the previous chapters occur in a $t\bar{t}\gamma$ -signal event. The information on the signal electron and signal muon events are merged together in one histogram per high-level object. Fig. 5.11

There are somewhat less than 100 tracks per event on average. The number of reconstructed photon candidates (about 15 on average), not only induced by the ambiguities with respect to electron and jet candidates, is much larger than the one photon that is expected to be radiated off the top-quark. The jet candidate multiplicity is much larger than the four jets from the semileptonic top-quark pair decay as well. Gluons can be radiated of any quark in the process (initial or final state radiation) resulting in an additional jet. In addition, even if the proton remnants in most cases disappear in the not instrumented forward directions, they can leave a jet signal in the detector as well, if they get an adequate transverse

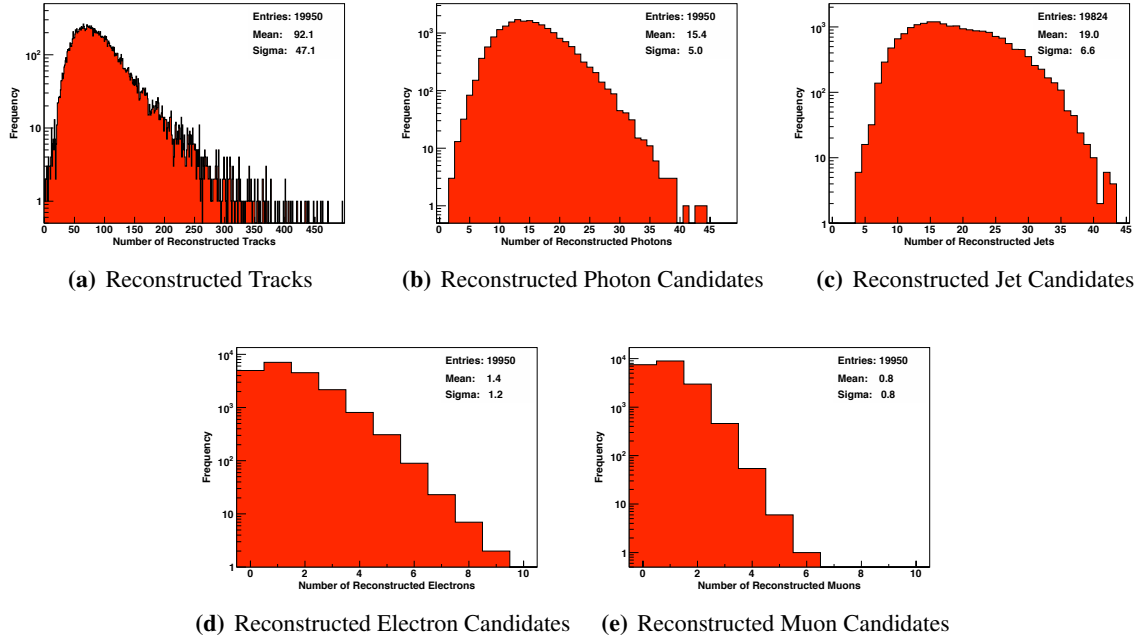


Figure 5.11: Multiplicity of reconstructed objects for semileptonic $t\bar{t}\gamma$ -events (signal electron and signal muon events combined).

momentum so that they can hit the detector. For the two types of lepton candidates the situation less complex. Even if electrons can be faked by jets the number of reconstructed electron candidates is much lower than the jet candidate rate because some preliminary identification criteria are already applied so that a loose separation between electron and photon/jet candidates is already included. The number of muon candidates is even lower because the reconstruction algorithm requires a clear signature in the muon detector combined with the silicon tracker so that is less probable to fake a muon candidate by another object.

5.6 The Analysis Flow

Fig. 5.12

In Figure 5.12 the analysis flow is depicted as it will be described in the next chapters. It can be seen that the analysis is designed in a chained but modular way. On the one hand the usage of different external software packages and libraries demands a rough splitting. On the other hand the new modules introduce an even finer granularity to avoid a reprocessing of the entire chain if the analysis code or parameters are changed at a certain point. Each depicted module might include a number of sub-modules, especially the event reconstruction. But finally, one C⁺⁺-main class exists per module.

Although a variety of predefined packages could be included in the analysis chain, this does not prevent a careful configuration of these modules because in most cases they provide only standard settings which is not specially adapted to the $t\bar{t}\gamma$ -analysis.

Each box in the diagram represents a different step of the analysis after which a new data-file has been stored on a hard-disk. The generator module at the beginning is the Monte Carlo event producer while most of the intermediate modules are event filters, keeping only these events that pass the dedicated analysis modules. Besides this filtering effect each module creates an individual ROOT-file summarizing information and results on each analysis step.

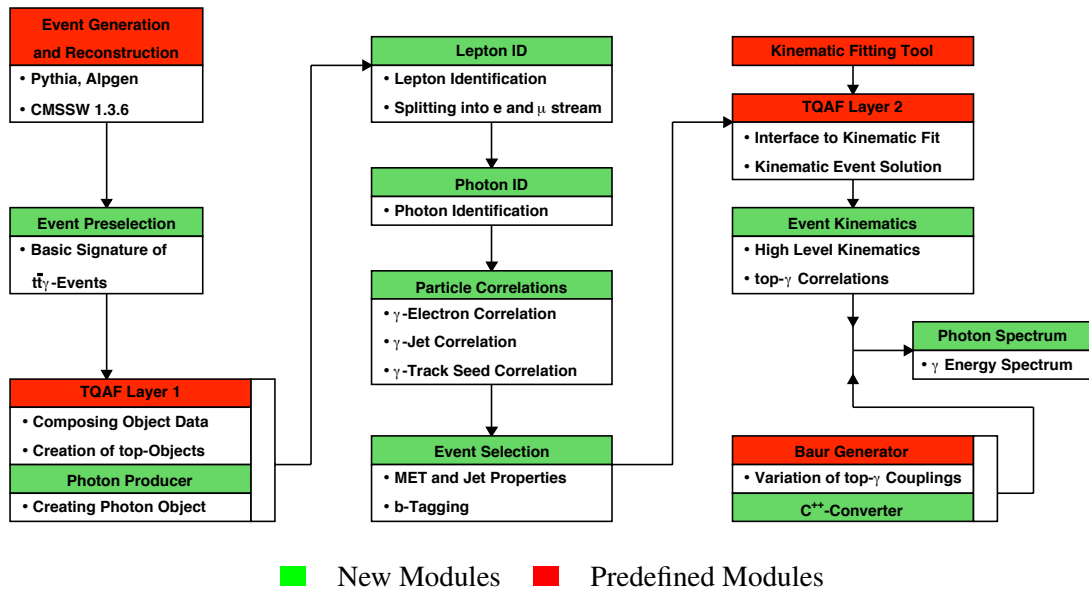


Figure 5.12: Flow chart of the different steps of the analysis of $t\bar{t}\gamma$ -events. The pre-defined modules are provided by global software packages, while the newly implemented modules were developed for this analysis. Per module a short outline is given summarizing its essential features. Details on each step of the analysis follow in the next chapters.

The following chapters explain in detail the physical relevance and results of the modules which have been added to the analysis chain. They constitute the technical implementation to select $t\bar{t}\gamma$ -events, to reject background events, and to create photon spectra from the remaining data.

Chapter 6

The Preselection of $t\bar{t}\gamma$ -Events

6.1 The Motivation for an Event Preselection

An event preselection is a sensible first step of an analysis due to physical and technical reasons. At first, more than eight million events must be analyzed in the present analysis and only these events which fulfill some minimum criteria with respect to the event signature shall be retained. If the aim is to measure a photon spectrum with a minimum energy of 20 GeV then it is not necessary to keep events which does not provide a candidate of this type.

From the technical side it is the amount of time and disk space that can be saved by requiring a minimum event signature. Reprocessing millions of events from the very beginning is not necessary if the first modules of the analysis code are finalized. Thus dropping events that do not pass the initial criteria increases the speed of an analysis significantly. Finally, especially the background events are generated for a wider community and provide information that suit to many more studies than a $t\bar{t}\gamma$ -analysis. By dropping these parts of the data which need not be used the analysis is again fastened and the necessary disk space for storing the events decreases as well.

Furthermore, a positive side effect of the reduction of the number of events is that after only two steps the present analysis can run on a local computer, rather than on computing clusters. As the initial data size reaches almost 20 TB, the preselection must use grid tools to run on the world-wide spread data-sets. For that reason an auxiliary program called *CMS Remote Analysis Builder (CRAB)* [66] has been used which serves for the creation and submission of grid jobs. Though this procedure is highly automated, there is an overhead with respect to time of several minutes up to hours depending on the work load of the grid infrastructure. The jobs are not directly transmitted to the computing sites, where the job shall run, but are distributed to free resources via a central computing facility, the resource broker. Especially in the end, when the processing of an analysis step itself lasts only a few minutes, this time can be saved when the data are stored on local disks.

6.2 The Event Preselection Criteria

The event preselection assumes the most general structure of a semileptonic $t\bar{t}\gamma$ -event. This includes a minimum number of high-energetic reconstructed objects respecting the acceptance regions of the CMS detector. Thus according to the reaction of equation 5.2 it is required that an event contains at least

- one photon candidate (not necessarily a top photon) having $E > 20 \text{ GeV}$ within $|\eta| < 2.5$,
- four jet candidates, each with $p_T > 20 \text{ GeV}$ and $|\eta| < 2.5$, and
- one lepton candidate
 - (a) for electron candidates: $p_T > 20 \text{ GeV}$ and $|\eta| < 2.5$,
 - (b) for muon candidates: $p_T > 20 \text{ GeV}$ and $|\eta| < 2.4$.

For jet, electron and photon candidates the whole η -acceptance of the calorimeters is not exploited because correlations among the particles and to tracks will be considered in later phases of the analysis. Thus, it is reduced to the intersection of all contributing sub-detectors, which is dominated by the acceptance region of the silicon tracker.

In a first attempt an inclusive $t\bar{t}$ -data-set of $3.9 \cdot 10^6$ events (corresponding to an integrated luminosity of $\mathcal{L} = 4.7 \text{ fb}^{-1}$) was filtered at generator level for events, where a photon is radiated off a top quark. This gave only approximately 13000 events, and only about 1000 of them remained after the preselection with energies of a photon candidate which is radiated off a top-quark as low as 1 GeV. This provides another justification for the generation of the dedicated signal data-sample. The events passing the preselection are the input for the studies, elaborating robust criteria for a proper event selection which are enriched with photons that are radiated off top-quarks.

6.3 The Results of the Event Preselection

When counting the number of object candidates that have passed the kinematic and acceptance cuts of the event preselection some striking properties can be found. The histograms are depicted in Figure 6.1. The distribution for each data-set is normalized to unity and then scaled according to the relative fraction of the total cross-section of a data-set in case that it consists of more than one sub-data-set¹. Already the event preselection considerably suppresses the event rates for the vector-boson and especially the QCD-multijet and $\gamma + \text{Jets}$ samples by demanding one high-energetic photon and a larger number of jets. Entering more into detail it becomes obvious, that these events only start to contribute noticeably if the jet multiplicity is generated by the hard interaction (compare the number of preselected events in the tables of appendix A). In any case the number of photons and jets for the background events is dominated by the number of photons and jets, respectively, that are created due to hard initial and final state radiation. Only if they appear at a high momentum scale these events might have passed the event preselection.

For the multiplicity of the electron or muon candidates the situation looks similar. Again, the QCD-multijet and $\gamma + \text{Jets}$ background data-sets are lacking of a high-energetic lepton candidate within the acceptance region. In case of the W- and Z-production real electron or muon candidates can stem from their leptonic decay channels. This can clearly be observed by the maxima in the counting rates in the one-lepton-bin (W + Jets and Z + Jets) and in the two-leptons-bin (Z + Jets). These lepton candidates are similar to a lepton candidate from the signal process because they are decay products of a massive boson

¹The entries for W/Z + Jets or QCD-multijet/ $\gamma + \text{Jets}$ can exceed one for single bins because for the sake of a clearer presentation these data-sets are combined as they are physically very similar.

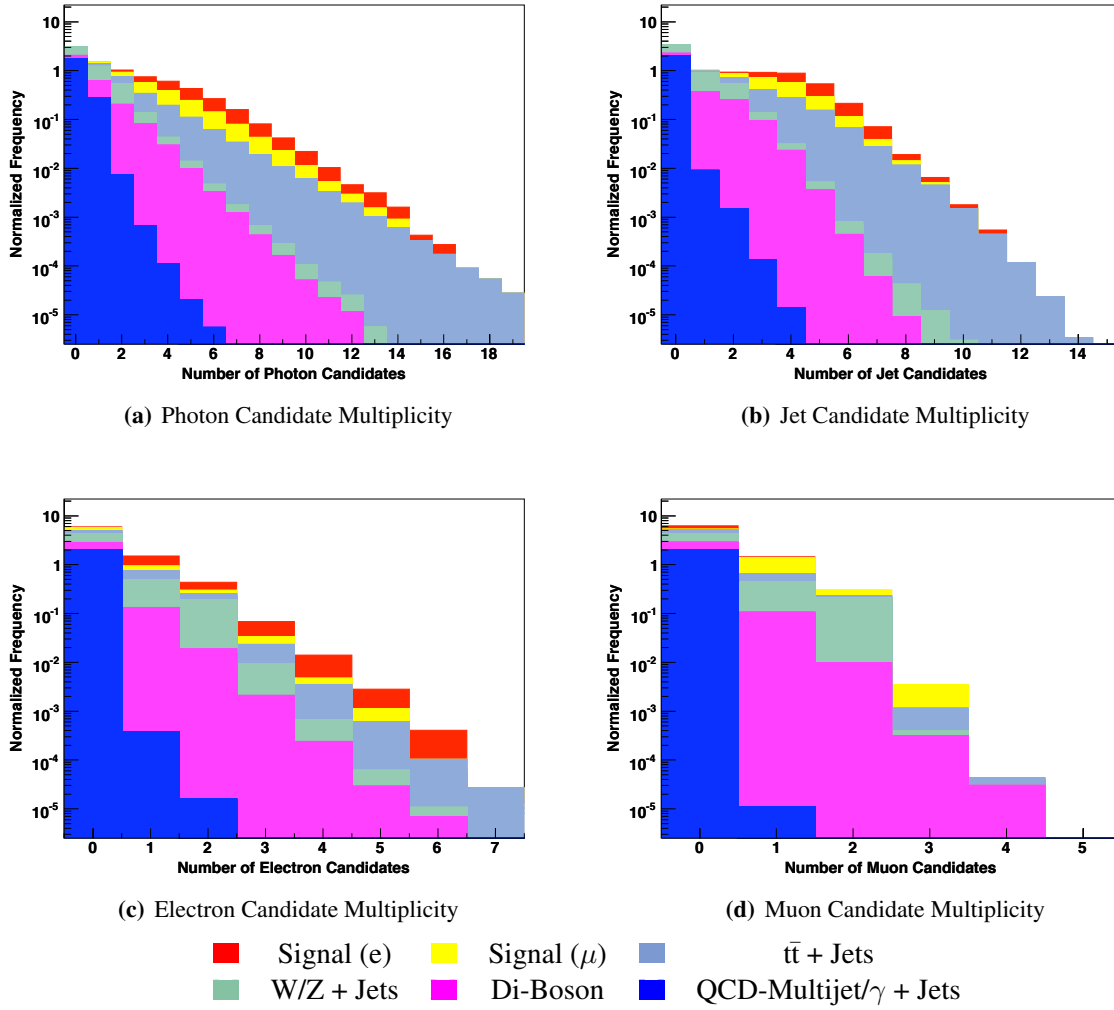


Figure 6.1: Multiplicity of photon (a), jet (b), electron (c) and muon (d) candidates after applying the preselection cuts. The histograms are stacked. The distribution for each data-set is normalized to unity and then scaled according to the relative fraction of the total cross-section of a data-set in case that it consists of more than one sub-data-set.

as well. Nevertheless, even if they pass the lepton identification criteria the multiplicity distributions allude to the strategy how to reject these events. If the number of leptons is restricted to exactly one, at least the Z + Jets background events should be rejected.

An aspect of the reconstruction algorithm can be confirmed when comparing the multiplicity of the photon and lepton candidates for signal electron and muon events. The number of photon candidates in signal electron events is basically shifted by one unit to higher counting rates, and the number of electrons is systematically higher than the number of muons. This is mainly because electron and photon candidates are very similar objects with respect to their reconstruction. By that stage no criteria have been worked out to separate electron and photon candidates so that the one might be counted as the other and vice versa. The last point to mention is that there is no obvious distinction between the signal process and the $t\bar{t}$ + Jets data-set. This is expected and gives a first hint on what the dominating background is going to be.

The energy and transverse momenta distributions of the preselected object candidates (Figure 6.2) show no striking difference between the various data-samples except for the QCD-multijet and γ + Jets events. Fig. 6.2

They reveal more likely low energetic distributions because in these processes no massive decaying particle is involved. In some bins of the QCD-multijet and γ + Jets distributions the steady shape is disturbed by single events mimicking high probabilities to select an event of one of the two data-sets. This can be explained by an statistical effect because the scaling factors for the events can differ by about fifteen orders of magnitude but only a limited number of events is available so that this effect is not compensated.

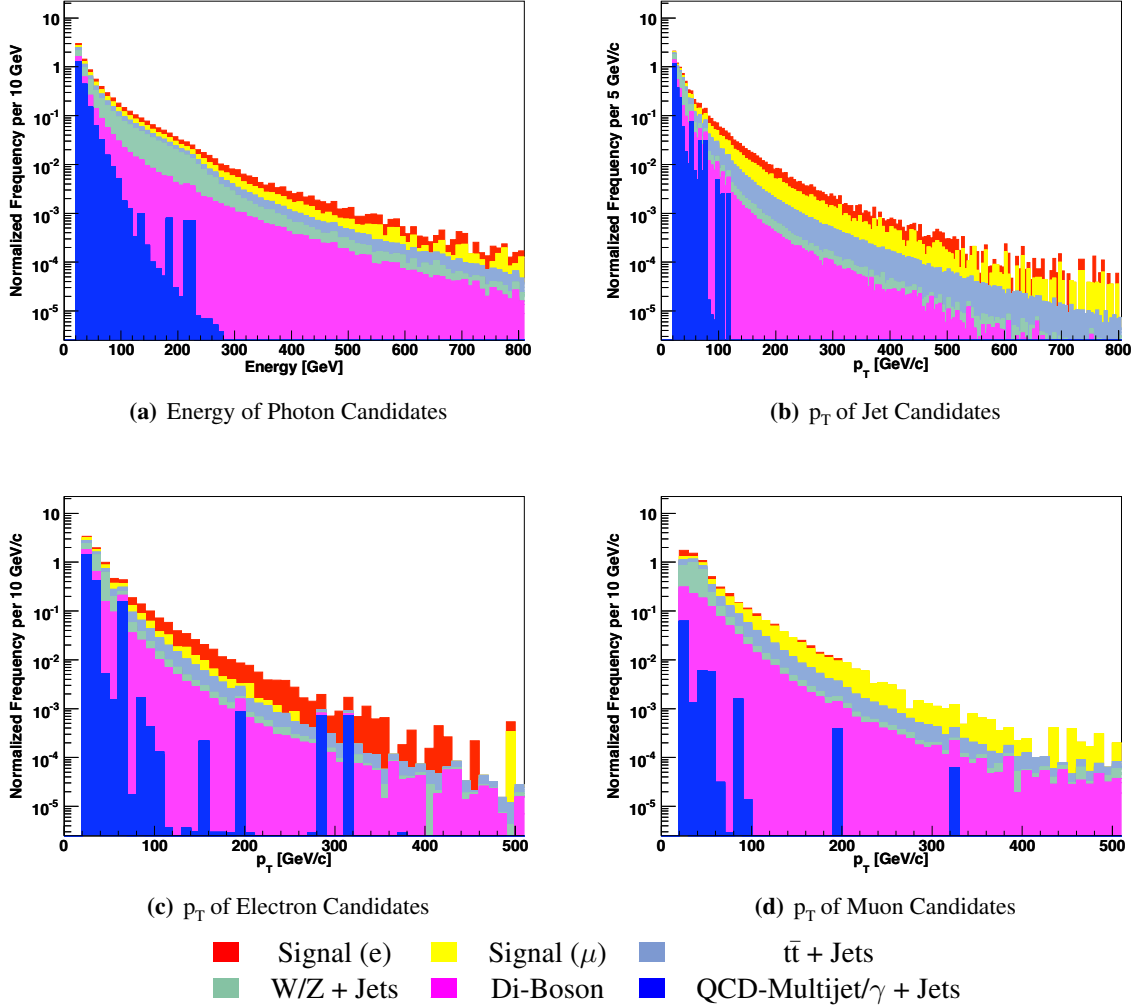


Figure 6.2: Distributions of energy of photon candidates (b), and transverse momenta of jet (b), electron (c) and muon (d) candidates after applying the preselection cuts. The histograms are stacked. The distribution for each data-set is normalized to unity and then scaled according to the relative fraction of the total cross-section of a data-set in case that it consists of more than one sub-data-set.

Tab. 6.1 The statistical result of the event preselection is presented in Table 6.1. For the signal events the selection efficiency in the electron channel is higher than in the muon channel due to the ambiguity of the electron and photon candidates but in the same order of magnitude. The $t\bar{t}$ + Jets events are even more suppressed but not yet by an order of magnitude. This gives a first hint that the reduction of this background data-set will become the most important challenge in the $t\bar{t}\gamma$ -analysis. In contrast, it is observable all non- $t\bar{t}$ -events are suppressed by several orders of magnitude compared to the $t\bar{t}$ -events. The reduction is more successful if massive particles like the vector-bosons are not involved in the physics process. Thus, the γ + Jets and QCD-multijet events are suppressed to the highest extent.

Data-Set	Number of Events		Efficiency
	Initial	After Preselection	
Signal Electron	$1.09 \cdot 10^3$	$6.38 \cdot 10^2$	$5.86 \cdot 10^{-1}$
Signal Muon	$1.09 \cdot 10^3$	$4.59 \cdot 10^2$	$4.21 \cdot 10^{-1}$
$t\bar{t}$ + Jets	$5.61 \cdot 10^6$	$1.43 \cdot 10^6$	$2.55 \cdot 10^{-1}$
W + Jets	$4.15 \cdot 10^8$	$5.44 \cdot 10^5$	$1.31 \cdot 10^{-3}$
Z + Jets	$3.43 \cdot 10^7$	$2.29 \cdot 10^5$	$6.68 \cdot 10^{-3}$
WW	$6.97 \cdot 10^5$	$8.34 \cdot 10^3$	$1.20 \cdot 10^{-2}$
WZ	$2.67 \cdot 10^5$	$4.56 \cdot 10^3$	$1.71 \cdot 10^{-2}$
ZZ	$1.11 \cdot 10^5$	$2.47 \cdot 10^3$	$2.23 \cdot 10^{-2}$
$W\gamma$	$5.62 \cdot 10^5$	$2.36 \cdot 10^2$	$4.20 \cdot 10^{-4}$
γ + Jets	$1.06 \cdot 10^9$	$2.32 \cdot 10^3$	$4.19 \cdot 10^{-6}$
QCD-multijet	$5.75 \cdot 10^{14}$	$2.56 \cdot 10^8$	$4.45 \cdot 10^{-7}$

Table 6.1: Final statistics and selection efficiency on the event preselection scaled to $\mathcal{L} = 10 \text{ fb}^{-1}$.

Chapter 7

The Selection of $t\bar{t}\gamma$ -Events

THIS chapter represents the main part of the study for the identification of photons in $t\bar{t}$ -events. By applying cuts on different measures the identification of leptons is studied at first. It has proven to be one of the most efficient tools to reduce the background events. Having split these events where a lepton could be identified into an electron and a muon stream, the second part is dedicated to the photon identification. Various variables relying on all sub-detectors of the CMS experiment are tested for their separating power to distinguish between photon candidates that are radiated off top-quarks and other photon candidates. As a consequence, these two steps of the analysis are upgrading the leptons and photons from mere object candidates to objects that are believed to be created by real leptons and photons. Based on these results the third section considers correlations between the particles also including the reconstructed jet candidates to prepare the objects for the kinematic event reconstruction. But before starting a brief overview on the Top-Quark Analysis Framework (TQAF) is given, which provided auxiliary program tools for the analysis.

7.1 The Top-Quark Analysis Framework

The Top-Quark Analysis Framework (TQAF) [67] is a common effort of the CMS top-quark group to share analysis code among the users. The major goals behind this software package which can be used in any analysis in the realm of top-quarks physics is to avoid multiple implementations of the same code for similar studies. On the one hand this is saving a lot of time, because the code duplication can be prevented, on the other hand the same code is reviewed several times by different users so that possible programming errors are detected more probably. It is implemented in a very modular way such that any functionality can easily be switched on or off, or more specific code can be integrated. Ref. [67]

The data structure of the TQAF is based on the event data model of the CMS software. The package does not interfere with the CMSSW reconstruction algorithm but rather expects the input files to be of the official format. Furthermore the output is of the same structure like the official one. The TQAF itself is structured in three layers.

1. The first layer creates what here is generally called a *TQAF-object* though the objects (particle candidates) need not necessarily be derived from a top-quark. It inherits all properties of a standard physics object candidate and adds further information to them. These can be isolation criteria for lepton candidates or additional b-tag information for jet candidates, for instance. It is important to mention that no special signal signature is required to create these objects.
2. The main functionality of the second layer is to build event solutions from the objects of the first layer. It provides tools for kinematic fits, event combinatorics as well as measures for evaluating

the probability of a certain event solution. In that layer a special event signature must be defined, e.g. semileptonic $t\bar{t}$ -events.

3. The third layer represents the special analysis code of the user. Strictly spoken, it is not a part of the TQAF but it is based on the output of the second layer.

The present analysis has used the TQAF as it was developed for the CMSSW 1.3-serie. Neither the TQAF nor the CMSSW version contain the latest improvements of the code development. But when the software version for the $t\bar{t}\gamma$ -analysis had to be fixed, only data-sets for CMSSW 1.3 were available, and the improved software versions have proven to be incompatible with these data-samples.

The following variables and methods have been adopted from the TQAF for the present analysis.

- Calculation of isolation criteria for the leptons (silicon tracker and calorimeters).
- A module to generate additional b-tag information.
- Modules which access CMSSW-internal methods to calibrate jets and to create event solutions by means of a kinematic fit to combine the physical objects to the semileptonic $t\bar{t}$ -initial state.
- A jet likelihood method to evaluate which of the combinatorial solutions matches best the semileptonic $t\bar{t}$ -event solution.

Ref. [22] The TQAF has included a complete electron identification as well, but this method is not taken into account. This algorithm is based on a likelihood analysis for an old CMS software package [22] and has not been validated for the version of the CMS software that is used for the current analysis.

Chp. 7.3 Finally, the first layer of the TQAF creates TQAF-objects from all physical object candidates but photons. Therefore an TQAF-object-like class for photons has been developed compatible with that framework. Details on that are described in chapter 7.3. This code is so far only private but it can be considered to be implemented into the TQAF or a similar software framework.

7.2 The Lepton Identification

The lepton identification is one of the most decisive steps of the semileptonic $t\bar{t}\gamma$ -analysis because the requirement for a high energetic and isolated lepton provides a powerful discriminator between the signal and background events, especially the QCD-multijet events. Depending on the matching of the lepton candidates to the generator particles, three different classes of lepton candidates are introduced from which the criteria to enrich real leptons are derived. For this analysis only electrons and muons are considered, while potential events with τ -leptons are assigned to the background events. It will be shown that the selection purity for electrons is 92.7% and the muon identification succeeds in 97.3% of all cases if either the electron or the muon is a decay product of a W-boson from a top-quark decay.

7.2.1 Lepton Classes

Before any criteria can be elaborated it must be explained which lepton candidate is assumed to originate from a real lepton and which is only faked. In addition, it must be separated among the real leptons, which is a decay product of a W-boson from a top-quark decay referred to as top electron or top muon in the following. For that reason two lists of generator particles are extracted from the Monte Carlo information per event, one containing all generator electrons, the other one all generator muons. If the distance in the $(\eta-\phi)$ -plane (ΔR) and the deviation of the reconstructed from the generated energy fulfill

the following relations the dedicated reconstructed electron or muon candidate is considered to be a real electron or muon candidate, respectively.

$$\begin{aligned} \Delta R &< 0.05 \quad (\text{distance criterion}) \\ \frac{|E_{reco} - E_{gen}|}{E_{gen}} &< 0.10 \quad (\text{energy criterion}) \end{aligned} \tag{7.1}$$

Driven by the fact, that the event preselection allows only leptons with $p_T > 20 \text{ GeV}/c$ generator leptons with $p_T > 15 \text{ GeV}/c$ within the preselection acceptance $|\eta| < 2.5$ are taken to make the matching procedure more efficient. In case of two generator leptons matching the same reconstructed candidate the one is favoured which has the smaller distance in the $(\eta-\phi)$ -plane.

So far this procedure separates real lepton candidates from fake ones. In order to split the real lepton candidates into top and non-top candidates the mother-daughter relations of the matched generator particle are exploited. If the mother of the latter is a W-boson which is the decay product of a top-quark the reconstructed candidate is a top candidate according to the introduced criteria, otherwise it is denoted as a non-top candidate.

In summary, the lepton identification distinguishes between three classes of candidates.

- Real Top Lepton Candidates (*top lepton candidates*, TL)
- Real Non-Top Lepton Candidates (*non-top lepton candidates*, NTL)
- Fake Lepton Candidates (*fake lepton candidates*, FL)

Before searching for identification criteria the list of all electron candidates is scanned for duplicates to obtain cleaner input data. These are electron candidates sharing one super-cluster with two tracks or one track pointing to two super-clusters. As in both cases two different candidates are reconstructed only this one is kept where the ratio of the energy of the super-cluster and the momentum of the track is closer to one. This removal option can already be switched on in the production of the TQAF-electrons.

7.2.2 Identification Criteria

The goal of a lepton identification is enrich the top lepton candidates, i.e. to distinguish these leptons which are direct decay products of a W-boson in the $t\bar{t}\gamma$ -decay chain from those leptons which are either faked by photons or jets, or might be delayed decay products of a b-meson in a jet, for instance. As the first step of that separation the kinematic distributions are checked. In Figure 7.1 the histograms of the transverse momentum of the different lepton candidate classes are compared. It is evident that for both lepton types the top components have in general a higher p_T than the non-top and fake ones. The entries in the lowest bin differ by a factor of two at least. Thus before more complex measures are considered the cut on the transverse momentum is increased from $20 \text{ GeV}/c$ to $30 \text{ GeV}/c$ for all lepton candidates.

Fig. 7.1

The two subsequent sections present details on the identification of electrons and muons. The criteria are optimized on the signal data-sets, i.e. the electron criteria are derived from the signal electron events, and the muon criteria from the signal muon events. The admixture of muons in electron events is about 5% while the number of electrons in signal muon events is about 30%. Even though the contribution of the other type is relatively large, especially for the signal muon events, their signatures are unambiguous so that these non-top and fake lepton candidates are reliably removed from the candidate lists as well even if they are not included into the studies of searching for identification criteria.

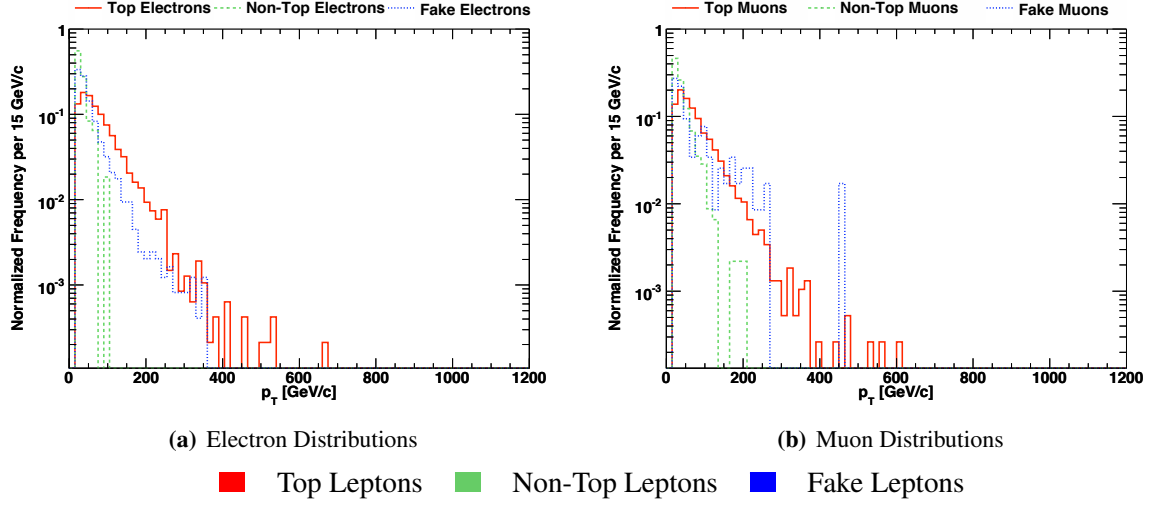


Figure 7.1: Distributions of the transverse momentum of the three lepton candidate classes before the lepton identification.

Criteria for Electron Identification

An electron candidate is considered to originate from a $t\bar{t}$ -decay if it is leaving an isolated signature in the silicon tracker and in the calorimeters. The following identification variables are examined in addition to the few loose identification criteria that have been applied in the electron reconstruction (see chapter 5.5).

Chp. 5.5

1. Tracker Isolation Variable

For the tracker isolation the scalar sum of the transverse momentum of all tracks is added which are closer than $\Delta R = 0.3$ compared to the reference track from the electron candidate. Then the p_T of the track of the electron candidate is subtracted. The resulting value describes the activity in the tracker in the vicinity of the electron candidate which should be zero for real and isolated electrons and should be larger than zero for fake and non-isolated ones.

2. Calorimeter Isolation Variable

The calorimeter isolation variables contains the sum of the energy of all electromagnetic and hadronic calorimeter towers that are positioned within a cone of $\Delta R = 0.3$ with respect to the track of the electron candidate. The energy of the super-cluster is detracted from this sum, so that this variable contains the remainder of the energy in that cone. As the energy of a calorimeter tower is not calibrated in contrast to the energy of the super-cluster this variable can be less than zero. This introduces another quality criteria. For the current analysis electron candidates are discarded if the ratio of this isolation variable to the total electron energy is less than -30% . This excludes candidates having a much larger energy correction than its total energy.

In general it is expected that the value for real and isolated electrons is close to zero, and well above zero for fake and non-isolated electrons.

3. Ratio of Hadronic over Electromagnetic Energy

The ratio of the hadronic over the electromagnetic energy (HoE) is zero for particles like the electron which lose its entire energy in the electromagnetic calorimeter. Thus, this variable allows to distinguish them from jets which ideally deposit their energy not only in the electromagnetic but also in the hadronic calorimeter giving $HoE > 0$.

The Figure 7.2 presents the histograms and graphs of these three variables¹. In each row the left his- Fig. 7.2

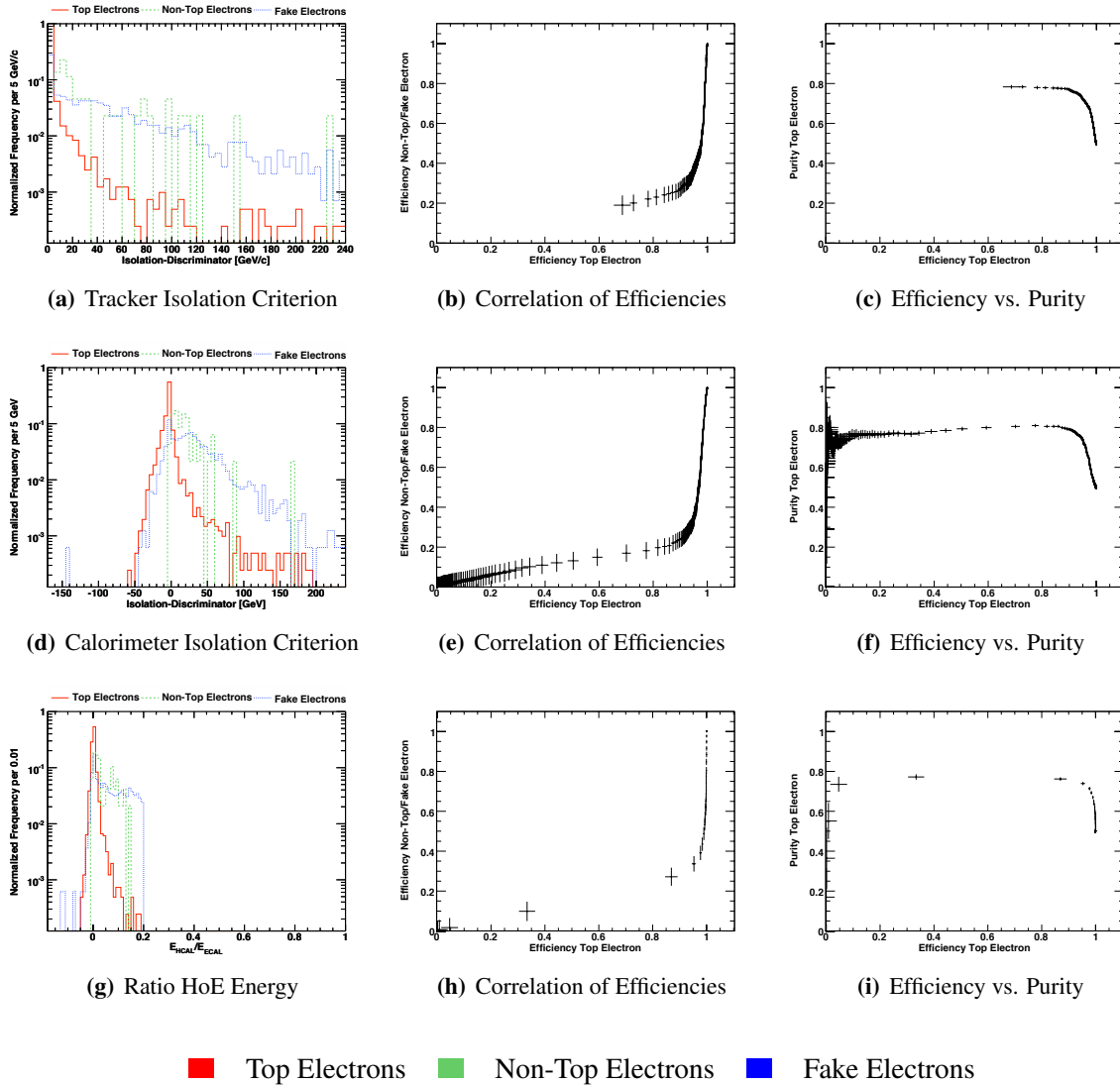


Figure 7.2: Distributions of the three electron identification variables for signal electron events. The left column reveals the normalized distributions for each of the three electron candidate classes. In the middle column the correlation of the top versus the combined non-top and fake selection efficiency of electron candidates is presented. The right column depicts the behaviour of the selection purity as a function of the selection efficiency of top electron candidates.

togram shows the probability distributions of the three types of electron candidates. In the middle the correlation between the efficiency to select a top electron candidate and the combined non-top and fake selection efficiency can be observed, if the cut on the dedicated variable is systematically varied. The right column concludes these investigations by showing how the selection purity of the top electron candidates varies with its selection efficiency.

Though the fake electron and partially the non-top electron candidates peak at the expected value for the top candidates, for each of the three variables large tails exist for the non-top and fake parts. They differ by several orders of magnitudes from the distributions of the top electron candidates which very steeply decrease in contrast. Thus, also proven by the initially nearly vertically falling efficiency correlations

¹The formulas used to calculate the errors on the selection efficiency and purity are attached in appendix C.

at high selection efficiencies for the top electron candidates, these three variables select very efficiently the top electrons candidates. This is supported by the dependence of the selection purity versus the selection efficiency of top electron candidates. It is strongly increasing when the selection efficiency is only minimally decreasing.

For each variable the cut value is chosen such that the product of selection efficiency and purity of top electron candidates is maximized.

$$\begin{aligned} \text{Tracker Isolation Variable} &< 6.0 \text{ GeV/c} \\ \text{Calorimeter Isolation Variable} &< 3.5 \text{ GeV} \\ \text{Hadronic over Electromagnetic Energy} &< 0.02 \end{aligned} \quad (7.2)$$

Chp. 7.2.3 The results of the electron identification are summarized in chapter 7.2.3. Having applied these cuts, the electron candidates lose their mere candidate status and become upgraded to actually identified reconstructed electrons according to the above discussed criteria. These cuts produce electrons that are not only reconstructed but also isolated particles in the detector.

Criteria for Muon Identification

The criteria for the muon identification are very similar to those for the electron identification because here an isolated signature in the silicon tracker and the calorimeters is required as well. But as high-energetic muons traverse both calorimeters to the muon system of the CMS detector they do not deposit a significant amount of energy on average in the calorimeters. Thus, the following isolation variables are taken into account for muon identification.

1. Tracker Isolation Variable

For the tracker isolation the scalar sum of the transverse momentum of all tracks in the silicon detector is added which are closer than $\Delta R = 0.3$ compared to the reference track of the muon candidate. Then the p_T of the track of the muon candidate is subtracted. The resulting value describes the activity in the tracker in the vicinity of the muon candidate which should be zero for real and isolated muons and should be larger than zero for fake and non-isolated ones.

2. Calorimeter Isolation Variable

The calorimeter isolation variables addresses the activity of the calorimeters in the vicinity of the muon candidate. Referred to the track of the muon candidate the sum of the energy of all calorimeter towers (electromagnetic and hadronic) that are positioned within a cone of $\Delta R = 0.3$ are summed up. The value for real and isolated muons is close to zero, and greater than zero for fake and non-isolated muons.

Fig. 7.3 The Figure 7.3 presents the histograms and graphs of the variables used for the muon identification.

The left column depict the probability distributions of the three types of muon candidates for both variables. In the middle the correlation between the efficiency to select a top muon candidate and the combined non-top and fake selection efficiency can be observed, if the cut on the dedicated variable is systematically varied. The right column concludes these investigations by showing how the selection purity varies with the selection efficiency for the top muon candidates.

Though the distribution of fake muon candidates peaks at the expected value for top muon candidates, there exist large tails which differ significantly from the distribution of the top muon candidates. The non-top muon candidates reveal the same larger tails and are systematically shifted towards higher isolation values which can be observed by the drop down of the curve with respect to the lowest bin. As for the

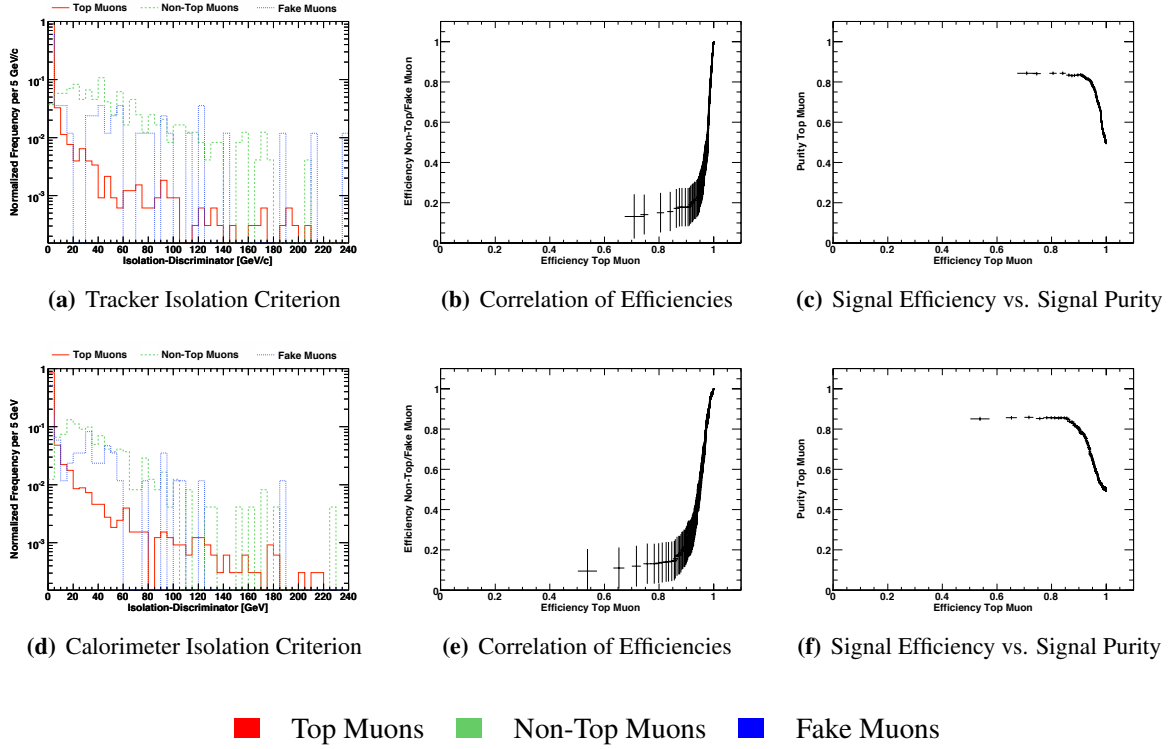


Figure 7.3: Distributions of muon identification variables for signal muon events. The left column reveals the normalized distributions for each of the three muon candidate classes. In the middle column the correlation of the top versus the combined non-top and fake selection efficiency of muon candidates is presented. The right column depicts the behaviour of the selection purity as a function of the selection efficiency of top muon candidates.

electron identification the selection efficiency correlation is falling steeply when the efficiency to select top muon candidates is still greater than 90%, though the error bars for the non-top/fake efficiency values become quite large because of the limited statistics of these candidates. At the same time the selection purity of top muon candidates is strongly increasing compared to the selection efficiency which is less sharply decreasing.

Even if the number of non-top and especially fake muon candidates is somewhat limited their signature compared to the top muon candidates is unique. Hence, an efficient muon selection based on the following cuts succeeds. As in the electron case the product of selection efficiency and purity of top muon candidates is maximized to obtain the cut values.

$$\begin{aligned}
 \text{Tracker Isolation Variable} &< 5.0 \text{ GeV}/c \\
 \text{Calorimeter Isolation Variable} &< 5.5 \text{ GeV}
 \end{aligned}
 \tag{7.3}$$

The results of the muon identification are summarized in chapter 7.2.3. Having applied these cuts, the muon candidates lose their mere candidate status and become upgraded to actually identified reconstructed muons according to the above discussed criteria. These cuts produce muons that are not only reconstructed but also isolated particles in the detector. Chp. 7.2.3

7.2.3 Results of the Lepton Identification

Tab. 7.1 The quantitative results of the lepton identification are indicated in Table 7.1 and Table 7.2. For both
 Tab. 7.2 lepton types these signature based methods gives top lepton selection purities well above 90% being better for muons than for electrons. The reason is that electrons can be mimicked by photons or jets leaving similar signatures in the detector, while muons are easier to recognize because of the special muon detector. This is emphasized by the rate of falsely identified, isolated muons in electron events which is negligible while the number of incorrectly observed electrons in muon events are mainly fake electrons which slightly degrades the selection purity of the top muon identification.

		Signal Electron Events (after cuts – before cuts)	Signal Muon Events (after cuts – before cuts)
Number of Events		3641 – 5798	2777 – 4224
Electrons	Top	3396 – 4095	not defined
	Non-Top	0 – 48	0 – 32
	Fake	268 – 1628	34 – 943
Muons	Top	not defined	2719 – 3278
	Non-Top	1 – 290	0 – 245
	Fake	0 – 36	42 – 85

Table 7.1: Number of leptons distinguished by the three classes passing the identification criteria separated for the electron and muon events (only signal data-sets).

	Signal Electron Events	Signal Muon Events
Event Selection Efficiency	62.8%	65.7%
Top Lepton Purity (only e or μ)	92.7%	98.5%
Top Lepton Purity (e and μ)	92.7%	97.3%
Average per Event	1.0069	1.0065

Table 7.2: Selection efficiency and purity for the lepton identification. The selection purity of the top lepton is calculated at first for top electrons in signal electron events and for top muons in signal muon events only. Then, a combined purity is determined if the number of identified muons in signal electron events and electrons in signal muon events is taken into account as well. The average per event gives the mean number of leptons (sum of electrons and muons) that are identified per event.

After applying the identification criteria on all electron and muon candidates in all data-sets the number of the leptons is counted as depicted in Figure 7.4. While there is almost only one lepton for signal events (see Table 7.2), it is obvious that the number of identified leptons is greater for a variety of background events, especially for the Z + Jets events where the Z-boson has decayed into a lepton pair, and the di-boson events including two heavy bosons where each of them potentially produces a lepton. Hence, an additional cut after the identification is implemented. Only events with exclusively one identified electron or muon are kept. This cut is less important for the muon events of the background data-sets because these have already been suppressed at the event preselection. The need for a reconstructed photon can easier be fulfilled in electron events because of the electron-photon ambiguity. At last it can be recognized that the lepton identification very efficiently reduces the γ + Jets and QCD-multijet events. Both data types do not contain a heavy boson from the hard interaction which can decay leptonically. Because of that the isolated and high-energetic lepton must be faked by any hadronic ingredient of these

events. The whole results of the lepton identification for the background data-sets in terms of event numbers are included in the tables attached in appendix A.

App. A

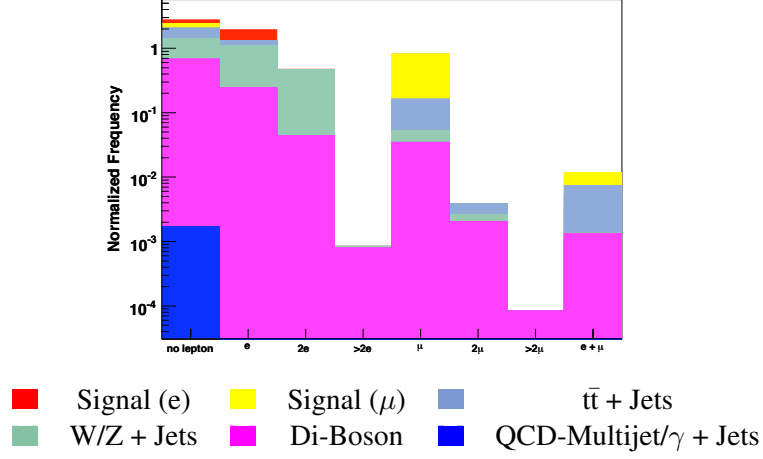


Figure 7.4: Number of leptons passing the identification criteria for the event samples. The histograms are stacked. The distribution for each data-set is normalized to unity and then scaled according to the relative fraction of the total cross-section of a data-set in case that it consists of more than one sub-data-set.

After the step of the lepton identification the analysis is split into an *electron event stream* and a *muon event stream* depending on the type of lepton that is found in an event. At this point the number of events with an identified electron is higher than that of muon events, especially for the background samples. This is still caused by demanding a high-energetic photon and the electron-photon ambiguity at the step of the event preselection. Hence, splitting the events does not smear the so far achieved better separation of signal and background events for the muon event stream.

7.3 The Photon Identification

The photon identification aims at selecting photons that are radiated off top-quarks out of the entire list of photon candidates. The separation criteria are separately investigated for the electron and the muon event streams because from the point of view of object reconstruction a photon is very similar to an electron. Apart from the different behaviour in the silicon tracker both particles ideally deposit their entire energy in the electromagnetic calorimeter. Thus, the presence of an electron might influence the identification of photons. It will be shown that not only the values of the discriminating variables but also the set of variables differ for the electron and the muon stream, but finally still leading to a comparable selection efficiency and purity. The photon identification provides furthermore a strong handle for the signal event selection. An isolated, high-energetic photon offers a very unique signature that is very well suited to reject all kinds of background events.

7.3.1 Photon Classes

Before starting the photon identification it must be defined what photon candidate is considered to be a candidate that is radiated off a top-quark (top photon candidate). Per event a list of generator photons is extracted from the Monte Carlo information and compared to the list of reconstructed photon candidates. The distance in the $(\eta-\phi)$ -plane ΔR and the deviation of the reconstructed from the generated energy are calculated. A reconstructed candidate is assumed to really originate from a generator photon when

these cut values are passed.

$$\begin{aligned} \Delta R &< 0.07 \quad (\text{distance criterion}) \\ \Delta E_{rel} = \frac{|E_{reco} - E_{gen}|}{E_{gen}} &< 0.15 \quad (\text{energy criterion}) \end{aligned} \quad (7.4)$$

At first, these two criteria separate real from fake photon candidates. If furthermore among the real photon candidates the matching generator photon is radiated off a top-quark this reconstructed object is promoted to the list of top photon candidates. For the matching only generator photons with a minimum energy of 15 GeV and $|\eta| < 2.5$ are considered. This saves computing time and only these can eventually match to any of the reconstructed photon candidates because the latter have already a minimum energy of 20 GeV within the same η -acceptance due to the event preselection. If a reconstructed candidate matches to more than one generator particle the one which has the lower ΔR -value is assumed to be the correct one.

After this step in the photon identification sequence, the entire list of candidates is split into three classes.

- Photon candidates, matching to a generator photon according to the cuts 7.4 which is radiated off a top-quark (*top photon candidates*, TP)
- Photon candidates, matching to a generator photon according to the cuts 7.4 which is not radiated off a top-quark (*non-top photon candidates*, NTP)
- Photon candidates, not matching to a generator photon according to the cuts 7.4 (*fake photon candidates*, FP)

Fig. 7.5

The matching cuts have been optimized such that the number of fake photon candidates which are more probably top photon candidates is kept low. In Figure 7.5 the accumulated distribution for the correlation of the matching criteria between the generator top photon and all reconstructed fake photon candidates per event are presented separated for the signal electron and signal muon events. By definition, there is no fake photon candidate closer than $\Delta R < 0.07$ and $\Delta E_{rel} < 0.15$ indicated by the white area in the lower left corner of the histograms. But it can be seen by the increased counting rates at the upper and the right edges of the white area that there are several photon candidates $\sim \mathcal{O}(10)$ which fulfill either only the distance or only the energy criterion. Therefore, they are considered to be fake photon candidates but are more probable top photon candidates as well. Tightening the cuts to $\Delta R < 0.05$ and the relative energy deviation down to 10% the number of these potentially falsely identified candidates would approximately be doubled.

Tab. 7.3

The challenge of a photon identification is now to reduce the number of non-top and fake photon candidates and to retain as much top photon candidates as possible. As it is indicated in Table 7.3 there are 4.9 (electron event stream) and 4.4 (muon event stream) photon candidates per event on average for the signal events. The mean number of photons for the background events is of a comparable size. When neglecting the physical background data-sets the average ratio of top photon candidates over the sum of non-top and fake-photon candidates is 6.8 (electron event stream) and 5.3 (muon event stream).

The chance to reduce the physical background events is estimated to be more realistic for the muon than for the electron stream. The number of top photon candidates is only approximately 20% lower, while the sum of non-top and fake photon candidates is smaller by a factor of approximately two to three for the most crucial background processes ($t\bar{t}$ + Jets and W + Jets) even before applying photon identification criteria.

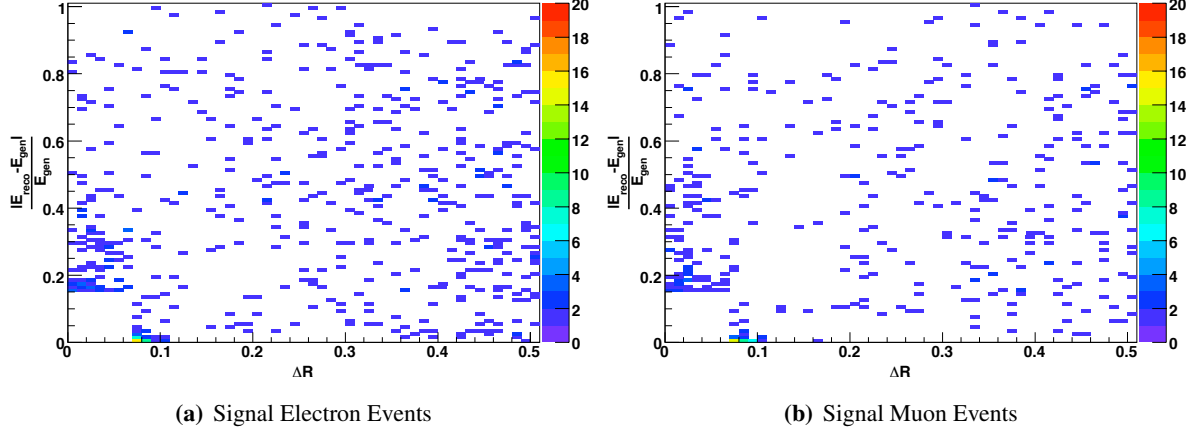


Figure 7.5: Correlation for the matching criteria between the generator top photon and each reconstructed fake photon candidate in an event. The higher counting rates at the upper and right edges of the white box in the lower left corner (real photon region) are more probable top photon candidates as well. They do not pass either the distance or the energy criterion and are thus sorted into the list of fake photon candidates.

	Electron Stream					Muon Stream				
	Events	TP	NTP	FP	$\emptyset$	Events	TP	NTP	FP	$\emptyset$
Signal Electron	398	251	56	1643	4.9	0	-	-	-	-
Signal Muon	2	2	0	5	3.5	298	205	45	1050	4.4
$t\bar{t}$ + Jets ($\cdot 10^3$)	294	1.133	43	1194	6.4	188	0.986	262	502	4.1
W + Jets ($\cdot 10^3$)	241	-	37	889	3.8	74	-	17	286	4.1
Z + Jets ($\cdot 10^3$)	99	-	13	358	3.7	10	-	3	36	3.9
WW	2046	-	279	6144	3.1	368	-	110	1091	3.3
WZ	1072	-	148	3277	3.2	100	-	17	234	2.5
ZZ	619	-	107	2239	3.8	30	-	13	74	2.9
$W\gamma$	42	-	47	105	3.6	4	-	4	8	3.0
γ + Jets	62	-	65	188	4.1	1	-	1	3	4.0
QCD-multijet ($\cdot 10^3$)	604	-	153	1827	3.3	125	-	3	518	4.2

Table 7.3: Total number and mean number per event of photon candidates before applying the photon identification, scaled to $\mathcal{L} = 10 \text{ fb}^{-1}$.

7.3.2 Generator Matching

Before starting the actual photon identification it is useful to know from which sources the photon candidates are originating. For that reason per event and separately for top, non-top and fake photon candidates each reconstructed photon candidate in the signal data-sets is compared to each stable generator particle² and depending on the matching result sorted into one of three sub-sets. The first one comprises these photon candidates matching to an arbitrary generator particle fulfilling both criteria of equation 7.4. The second and third ones contain these candidates when either the distance or the energy criterion matches only. The standalone energy criterion might be only a very weak hint on an agreement between a reconstructed candidate and a generator particle, because any generator particle which is pointing in a completely arbitrary direction can accidentally have nearly the same energy as the reconstructed photon candidate. Furthermore, it is explicitly not excluded that a photon candidate can match to more than one generator particle (double-counting). This provides information if maybe one of the top photon candidates would have matched to another particle than the generator photon that is radiated off the top-quark.

The results of these matching studies are depicted in the sub-figures (a) to (c) of Figure 7.6 – 7.8 for the signal electron events and in the sub-figures (a) to (c) of Figure 7.9 – 7.11 for the signal muon events. In addition, if it is found that the photon candidate matches to a generator photon it is checked which the mother particle of that generator photon has been. These charts can be observed in the second row of these figures (sub-figures (d) to (f)) again separated by the three sub-sets of matching conditions.

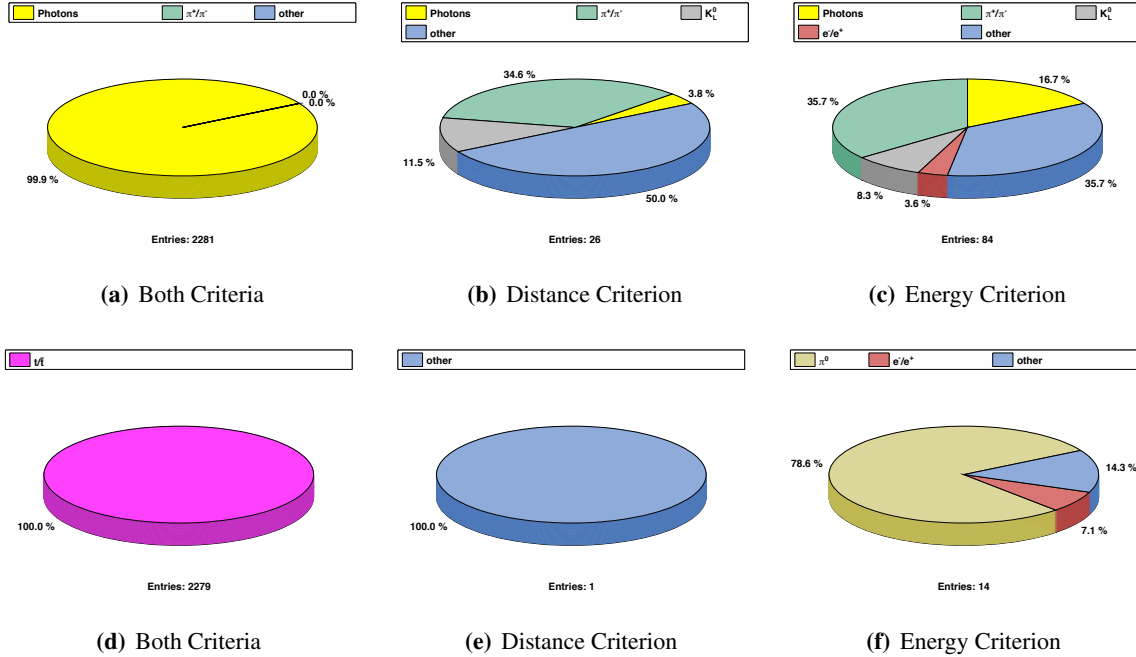


Figure 7.6: Signal Electron Events, Top Photon Candidates: Matching to generator particles (upper row) and the mother relation if the matched generator particle is a photon (lower row). The entries per chart are not scaled to a certain value of the integrated luminosity.

For both event types the fraction of top photon candidates fitting to another generator particle than the photon that is radiated off the top-quark is only about 5% when summing over all three matching sub-

²This does not mean that all of these particles do not decay. But their lifetime is so long that they are either absorbed by the detector or decay only having already left the detector. In other, more technical words these are the particle with status code 1 from the Monte Carlo generators.

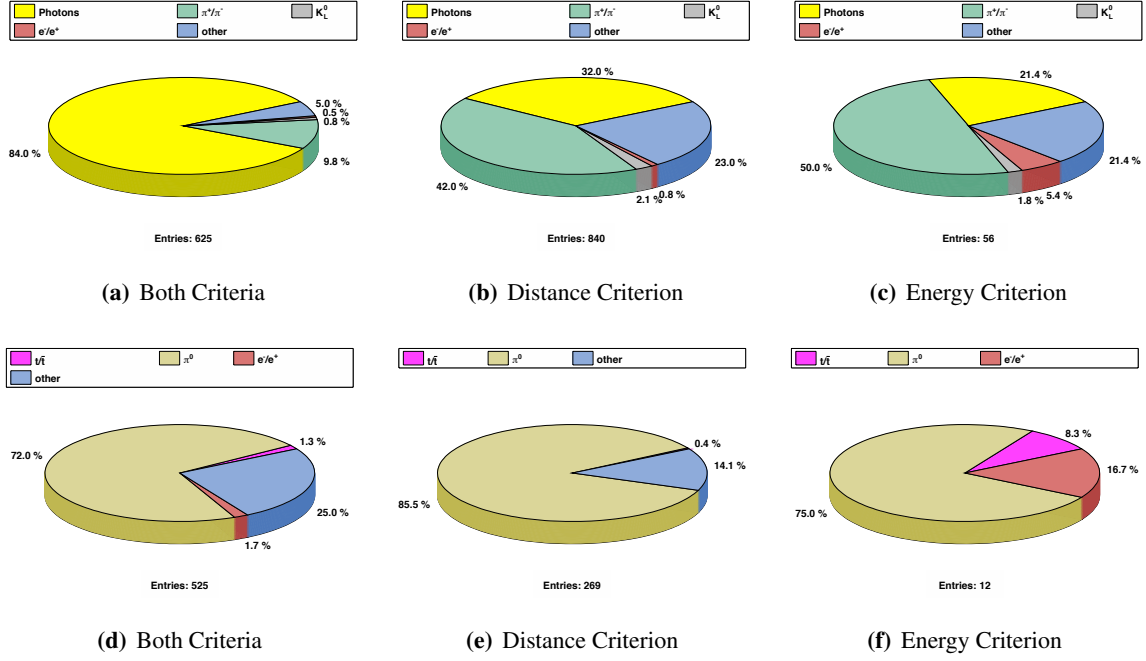


Figure 7.7: Signal Electron Events, Non-Top Photon Candidates: Matching to generator particles (upper row) and the mother relation if the matched generator particle is a photon (lower row). The entries per chart are not scaled to a certain value of the integrated luminosity.

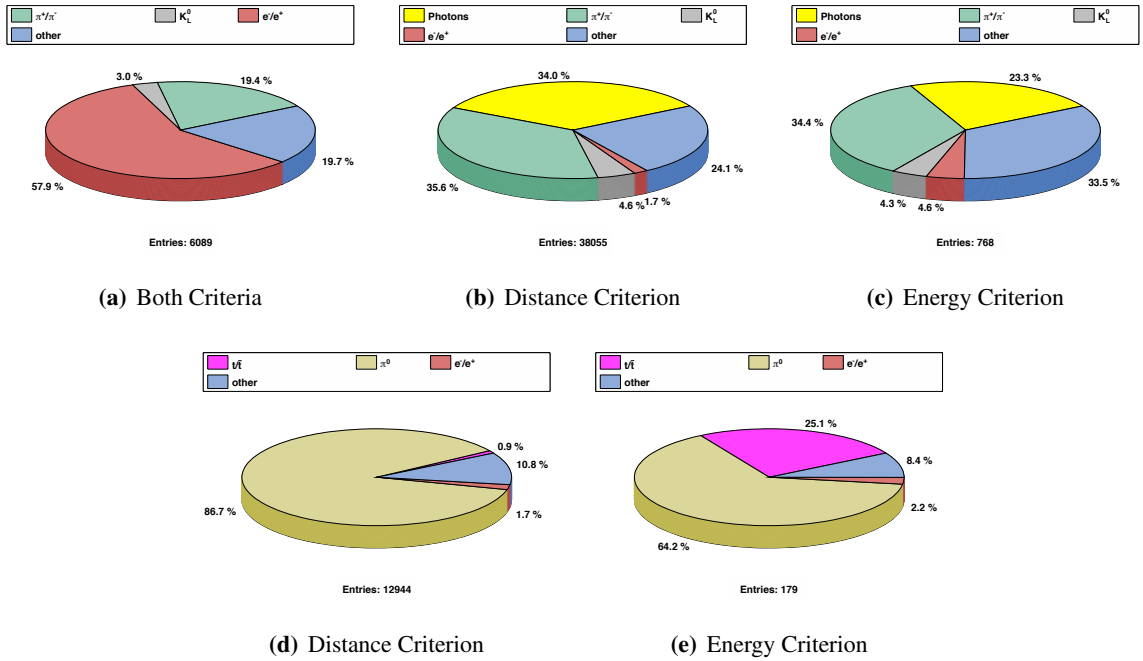


Figure 7.8: Signal Electron Events, Fake Photon Candidates: Matching to generator particles (upper row) and the mother relation if the matched generator particle is a photon (lower row). By definition there does not exist a chart for the mother relation fulfilling both criteria for the fake photon class. The entries per chart are not scaled to a certain value of the integrated luminosity.

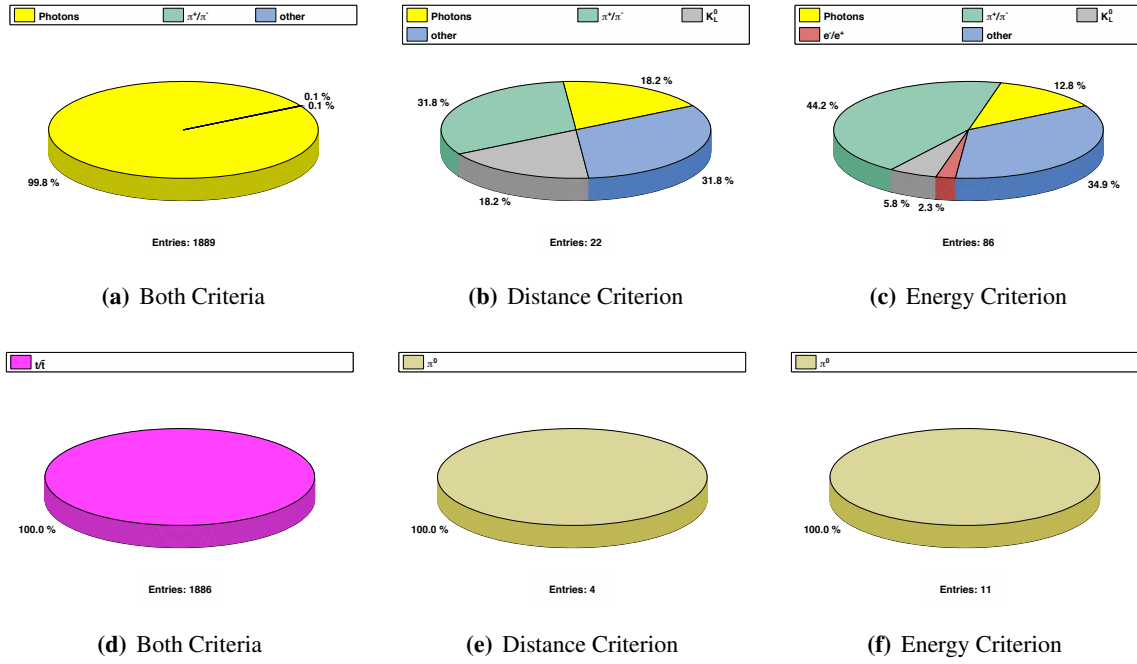


Figure 7.9: Signal Muon Events, Top Photon Candidates: Matching to generator particles (upper row) and the mother relation if the matched generator particle is a photon (lower row). The entries per chart are not scaled to a certain value of the integrated luminosity.

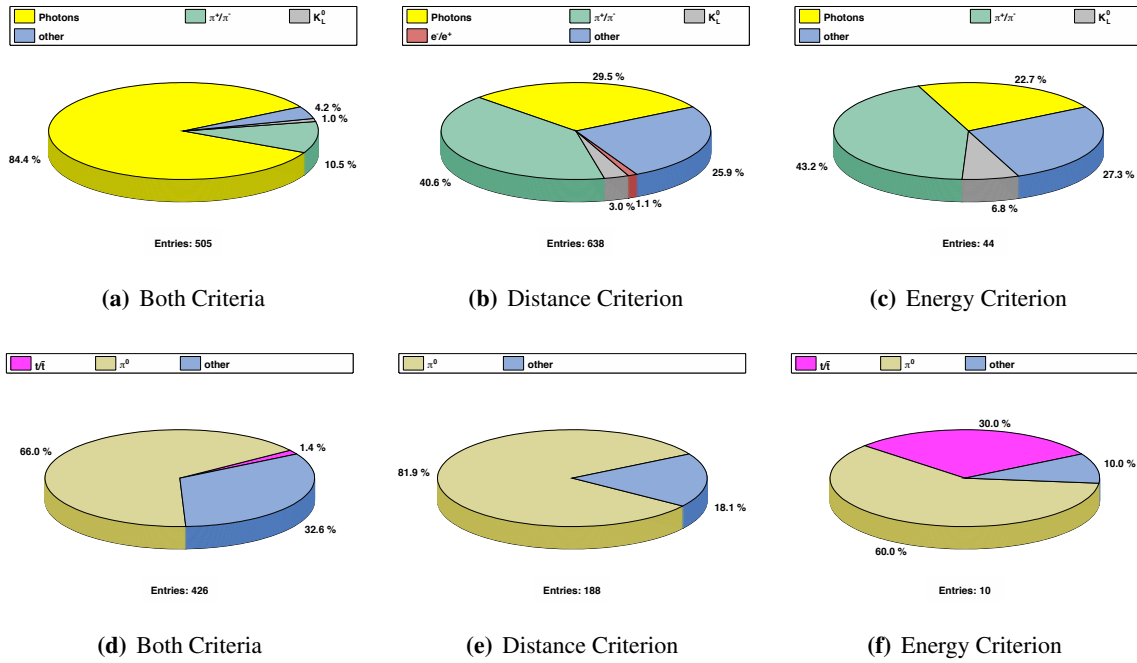


Figure 7.10: Signal Muon Events, Non-Top Photon Candidates: Matching to generator particles (upper row) and the mother relation if the matched generator particle is a photon (lower row). The entries per chart are not scaled to a certain value of the integrated luminosity.

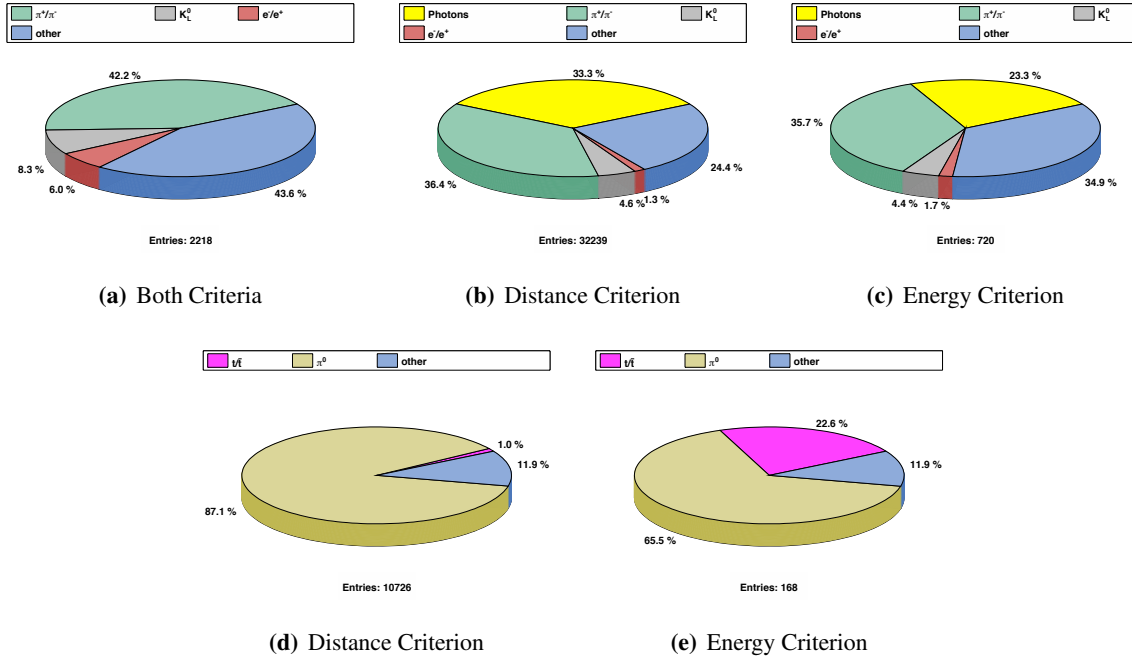


Figure 7.11: Signal Muon Events, Fake Photon Candidates: Matching to generator particles (upper row) and the mother relation if the matched generator particle is a photon (lower row). By definition there does not exist a chart for the mother relation fulfilling both criteria for the fake photon class. The entries per chart are not scaled to a certain value of the integrated luminosity.

sets. But for the majority of cases only the weak energy criterion is valid. Especially the number of photon candidates accepting both criteria for other particles than photons is only at the per-mille level. Hence, the top photon candidate class is very unambiguous with respect to the generator matching.

For the non-top photon candidates the situation is more diversified. The ambiguity for photon candidates that would perfectly match to another particle than a photon is approximately 16% for electron and muon events, while for matches due to only one of the criteria there is a fraction of about 40% when a photon candidate would fit to a $\pi^+\pi^-$ and less often to a K_L^0 . In about one third of all cases the candidate matches to a photon according to only one of the criteria. The numbers are slightly higher for electron compared to muon events.

When analyzing the relations for the fake photon candidates one clearly has to distinguish between electron and muon events. In the electron case respecting both criteria the majority of the photon candidates fits to generator electrons, and the second largest part consists again of a matching to $\pi^+\pi^-/K_L^0$ -mesons. If only one criterion is considered every third fake photon candidate matches to a generator photon, and about 40% are created by one of the light mesons. The agreement in muon events gives about the same numbers for the meson matching except from the case of both criteria. Generator electrons have produced only a minority of fake photons, but the $\pi^+\pi^-/K_L^0$ -part has increased to about 50%.

The studies off which particle the generator photons are radiated that have matched to any reconstructed candidate give the following results. At first, as a consistency check, it can be seen that the top photon candidates fulfilling both criteria are really photons radiated off top-quarks for both signal electron and muon events. Then, for the non-top and fake photon candidates in signal electron events only a small fraction can be identified as being radiated off an electron. This might be somewhat surprising because the photon radiation off the electron as the lightest charged particle should generally be favoured. But the electromagnetic reconstruction algorithms are built such that the electron and a potential radiated photon are merged together. Therefore, only in a very few cases the two objects are reconstructed separately

when they are separated so far that the merging fails. For the entire rest of the non-top and fake photon candidates a distinct tendency can be observed not only for the signal electron but also for the signal muon events. The majority of photons has the π^0 -meson as its mother particle. The physical explanation is that the π^0 -meson decays after $8.4 \cdot 10^{-17}$ s to 98.8% into a photon pair, so that its decay products are detected.

Fig. 7.5

In the end, an aspect must be emphasized, which has already been found in the discussions on Figure 7.5. The non-top photon candidates and even more pronounced the fake photon candidates contain a fraction of photon candidates which are radiated off top quarks. They could not be identified as such because they lie outside the classification interval (equation 7.4).

The above described procedure has been applied to the $t\bar{t}$ + Jets background events as well which signature is similar to the signal events. The results have been confirmed.

For all kinds of processes the challenge is to find measures and cut values not only to distinguish real from fake photon candidates but to separate top photon candidates from non-top ones, too. In summary, the task of the photon identification is to reduce

1. non-top and fake photon candidates (electron and muon streams) that are decay products of π^0 -mesons or charged, hadronic jets, because the π^+/π^- -mesons of the generators most probably end in jets in the detector, and
2. fake photon candidates (electron stream) that are electrons.

Thus, the identification criteria must base on information of the silicon tracker and both calorimeters as well. Different criteria will be studied in the next part and it will be shown that the matching of fake photon candidates to electrons leads to different measures for the electron and muon event streams.

7.3.3 The Photon Object

In order to distinguish between top and non-top and fake photon candidates, respectively, numerous measures derived from the silicon tracker and the calorimeters are systematically examined to find the most appropriate ones which differ most and are least correlated. To facilitate this scan of variables a dedicated C^{++} -class for photon objects has been implemented (see Figure 7.12). Adapting the same

Fig. 7.12

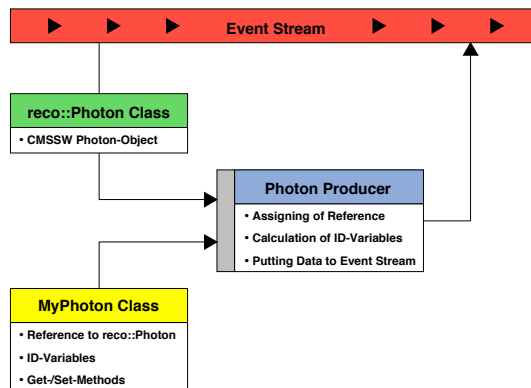


Figure 7.12: Scheme how the improved photon class (MyPhoton) and the photon producer interferes with CMSSW photon object (reco::Photon) and the CMSSW event data stream.

Chp. 7.1

idea as for the reconstructed objects of the TQAF (chapter 7.1) this private photon class contains a reference to the standard photon object as it is provided by the CMS reconstruction software. By that all

information about the kinematic variables, the super-cluster, the cluster shape etc. are preserved. But as it does not contain any isolation information and is missing a number of combined variables, these are added to the improved photon object.

The new object is created and pushed into the event stream, i.e. the data stream that is finally stored to an output file, by means of two additional classes. The first one (MyPhoton) administers the core object of that class, containing the reference, the new variables and the methods to set and access these values. The second class, the photon producer, is build around that object class in the sense that it is producing these objects. The routines to calculate the variables are exported to that class. Furthermore as it will be seen in the following, there is more than one version calculated per variable by varying input parameters which is the size of a ΔR -cone around a photon, for instance. This facilitates the systematic analysis to find the best cut value for a given measure.

The measures that have been investigated to identify top photon candidates can be ordered into three categories. If not explained differently, in any of the defined variables a ΔR -distance is always referred to the middle line of the photon candidate given by its momentum vector.

1. **Tracker Isolation Variables** - As photons are neutral particles they are not expected to retain a track in the silicon inner tracker, while electrons and jets do. Several isolation variables are calculated to exploit these correlation.

- p_T -Sum Criteria

The scalar p_T -sum of all tracks that lie in a given cone of size $C \in \{0.2, 0.3, 0.4, 0.5\}$ in the $(\eta-\phi)$ -plane. This measure should ideally zero for photon candidates and well above zero for electrons and jets.

$$\sum_{\Delta R < C} p_T \quad (7.5)$$

- ΔR -Track Criteria

The distance to the N^{th} nearest track, $N \in \{1, 2, 3, 4\}$, with a minimum transversal momentum of $p_T^0 \in \{1.0 \text{ GeV}/c, 1.5 \text{ GeV}/c\}$. As photons are not expected to have a track match, the next track should be not too close to an isolated photon candidate.

$$\Delta R^{\text{track } N} \text{ with } p_T^{\text{track}} > p_T^0 \quad (7.6)$$

- N-Tracks Criteria

The number of tracks in a cone of size $C \in \{0.2, 0.3, 0.4, 0.5\}$ in the $(\eta-\phi)$ -plane while the track must have a minimum transversal momentum of $p_T^0 \in \{1.0 \text{ GeV}/c, 1.5 \text{ GeV}/c\}$. This measure is very similar to the last one while it is counting the tracks in the vicinity of the photon candidate which is optimally zero.

$$N_{\text{tracks}} \text{ for } \Delta R < C \text{ and } p_T^{\text{track}} > p_T^0 \quad (7.7)$$

2. **Electromagnetic Calorimeter Variables** - These variables use the different shower shapes created in the electromagnetic calorimeter. As a single photon deposits about 94% and 97% of its energy in a cluster of 3×3 and 5×5 cells of the electromagnetic calorimeter six different ratios of these energy clusters are considered.

- Shower Shape Criteria

$$\frac{E_{22}}{E_{33}}, \frac{E_{22}}{E_{44}}, \frac{E_{22}}{E_{55}}, \frac{E_{33}}{E_{44}}, \frac{E_{33}}{E_{55}}, \text{ and } \frac{E_{44}}{E_{55}} \quad (7.8)$$

In the optimal case the shower of a photon candidate should be very narrow so that most of the energy is already deposited in the smaller cluster, and the outer cells only contribute to a small extent, such that the ratio is close to one. This can separate photons from jets, as the latter are spread more widely in the calorimeter, so that the outer cells contain a significant amount of energy pulling the ratio away from one. Furthermore these ratios can eliminate these real photon candidates which are decay products of π^0 -mesons if the two do not hit the same cell. In the best case both of them arrive at different but close cells in the electromagnetic calorimeter so that the main part of their energy is released in the smaller cluster. But the cluster of the other is adding a large amount of energy to the bigger cluster of the first one, so that the ratio is again smaller than one.

3. **Combined Calorimeter Variables** - An important tool to exclude jets from the photon candidate list is to calculate the ratio of energy that is deposited in the hadronic calorimeter behind the electromagnetic super-cluster.

- Hadronic over Electromagnetic Energy (HoE) Criteria

$$\frac{E_{\text{HCAL}}^{\Delta R < C}}{E_{\text{Super-Cluster}}} \quad (7.9)$$

The energy entries of all hadronic calorimeter towers within a cone of size $C \in \{0.2, 0.3, 0.4, 0.5\}$ are summed up and compared to the electromagnetic energy. A photon should be stopped inside the electromagnetic calorimeter so that a ratio should ideally be zero. On the other hand most of the jets reach the hadronic calorimeter and lose the rest of their energy only then. Even if not the biggest part of the energy is detected by the hadronic calorimeter towers the ratio becomes greater than zero.

In order to extract the most appropriate identification criteria the distributions of the top, non-top and fake photon candidates are compared. Having added the histograms for non-top and fake photon candidates, each of the remaining two histograms is normalized to one. Then the cut values are systematically varied and the selection efficiency of the top, of the combined non-top and fake photon candidates as well as the selection purity of the top photon candidates are calculated. The results of these examinations separated for electron and muon events are presented in the next section.

7.3.4 Identification Criteria

In order to decide if a criterion is useful for the identification of top photon candidates three arguments are considered. At first the selection efficiency of top photon candidates must fall less steeply compared to the non-top/fake selection efficiency if the cut values become tighter. Secondly, the selection purity of the top photon candidates should be as high as possible, and thirdly, the correlation of a criterion to all others should ideally be zero. This avoids that measures contribute to the identification which have almost no influence on the reduction of non-top and fake photon candidates because other measures have already removed the candidate.

Therefore, at first the graphs of the selection efficiency of top candidates against the non-top/fake efficiency are considered for all isolation criteria. Per group this criterion is further investigated which reveals the highest top efficiency and at the same time the lowest non-top/fake efficiency to achieve the best rejection for non-top and fake photon candidates. Then, the correlations between the chosen criteria are considered so that finally a set of four identification criteria is remaining for each lepton event stream. The best cut values are determined by maximizing the product of selection efficiency and selection purity for the top photon candidates. Again the importance to split the events into an electron and a muon

stream will be proven, because not only the cut values but also only three out of the chosen four criteria coincide.

Electron Events

According to the above mentioned rules Figure 7.13 compares the selection efficiency of the top photon candidates versus the combined non-top and fake selection efficiency for each group of the identification criteria. Except for the p_T -sum criterion for each group the criterion is selected which gives the steepest curve. For the p_T -sum the cone size has been chosen to be $\Delta R < 0.3$ instead of $\Delta R < 0.2$. This gives better results for the identification as the two curves intersect at a selection efficiency of top photon candidates of about 80%. From there towards lower values of the efficiency this criterion has a more powerful separation. Fig. 7.13

In order to check that all groups of isolation variables contribute to the photon identification the correlation matrix for the chosen measure of each group is depicted in Figure 7.14. At first, it is obvious that the correlations between the N-tracks criterion and the other criteria is the greatest for all three types of photon candidate classes. Thus, it is not integrated into the set of cuts that are used for the photon identification. For the rest of the measures some correlations to the HoE-criterion especially for non-top photon candidates are observed. But as the number of fake photon candidates must essentially be reduced this criterion is kept because for this type of photon candidates the correlations are less pronounced. Fig. 7.14

Having identified the most appropriate criteria Figure 7.15 compares in detail the results for the top, non-top and fake photon candidates from which adequate cut values are determined. In each row the left histogram presents the distributions for the three candidate classes normalized to unity. The second graph reviews the correlation of both selection efficiencies, and the right graph shows the selection purity versus the selection efficiency of the top photon candidates. The two latter graphs are derived by systematically varying the respective cut value of the dedicated measure. It is evident that there is an overlap of the selection variables between the top and the non-top/fake candidates but the shapes for any criterion differ obviously such that a proper separation is possible. That is supported by the correlation of the top versus the non-top/fake selection efficiency which fall steeply, and the selection purity of the top candidates as a function of their selection efficiency which is strongly increasing. The performance of the shower shape criterion (E_{22}/E_{55}) is slightly worse compared to the other criteria. Fig. 7.15

The best cut for each measure is chosen by maximizing the product of selection efficiency and selection purity of the top photon candidates resulting in the following set for the electron event stream.

p_T -Sum Criterion	$\sum_{\Delta R < 0.3} p_T$	$<$	$7.0 \text{ GeV}/c$	(7.10)
ΔR -Track Criterion	$\Delta R^{\text{track } 1}, p_T^{\text{track}} > 1.5 \text{ GeV}/c$	$>$	0.12	
Shower Shape Criterion	$\frac{E_{22}}{E_{55}}$	$>$	0.86	
HoE Criterion	$E_{\text{HCAL}}^{\Delta R < 0.2} / E_{\text{Super-Cluster}}$	$<$	0.10	

Muon Events

As for the electron stream the events of the muon stream have been compared contemplating the dependencies of the selection efficiencies of the top and the combined non-top and fake photon candidates, when the cut values are varied. The results of these systematic studies are presented in Figure 7.16. Again the best measure of each group is determined by the steepest curve shape. At this point a difference between the electron and the muon stream occurs. As an electron requires a track match and fake photon candidates are mimicked by electrons the ΔR to the first closest track mainly selects the matched track in case of the electron event stream. For the muon event stream this electron-photon fake is less Fig. 7.16

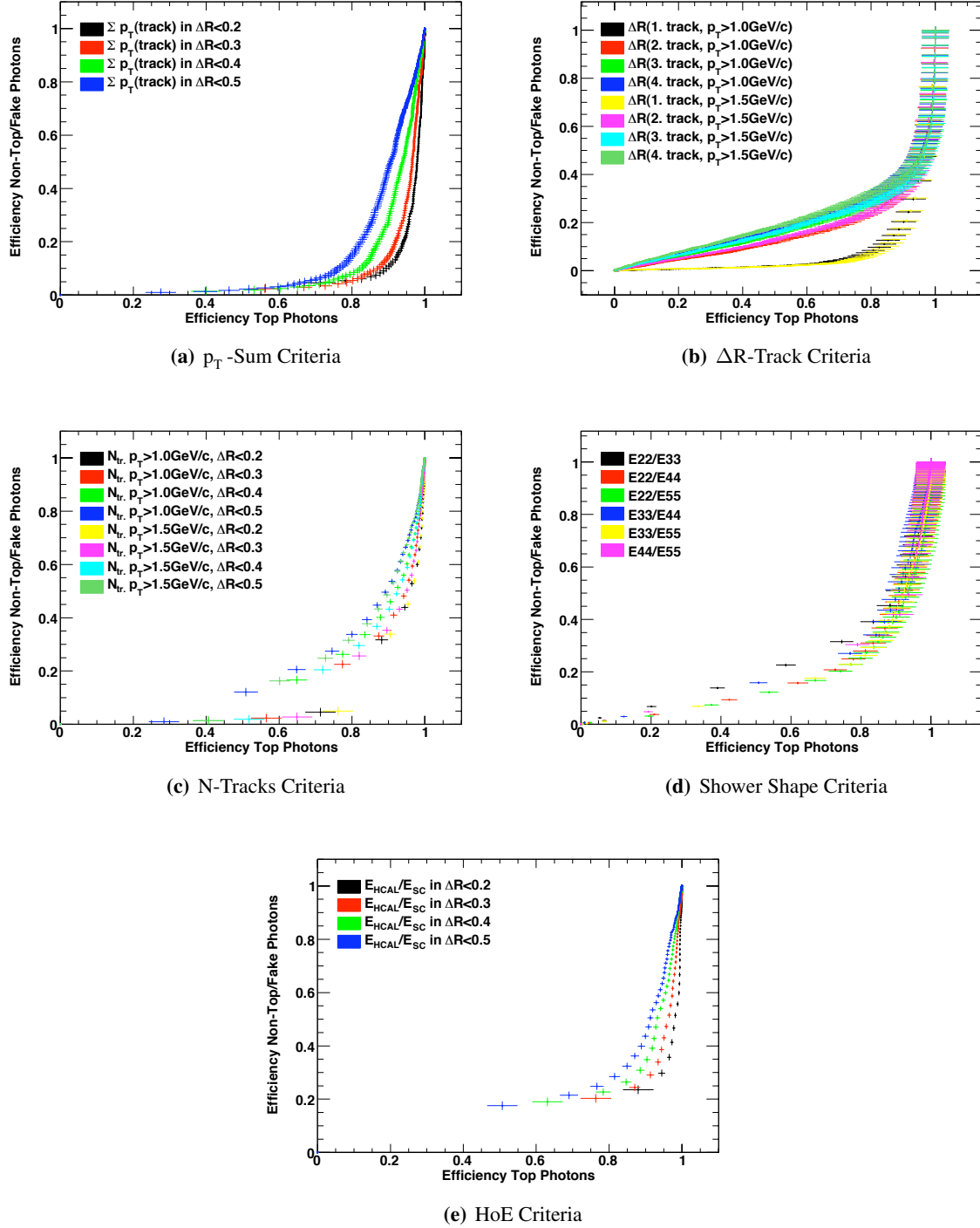


Figure 7.13: Signal Electron Events: Correlation of the selection efficiency of top photon candidates versus the selection efficiency of non-top/fake photon candidates for the various types of identification criteria.

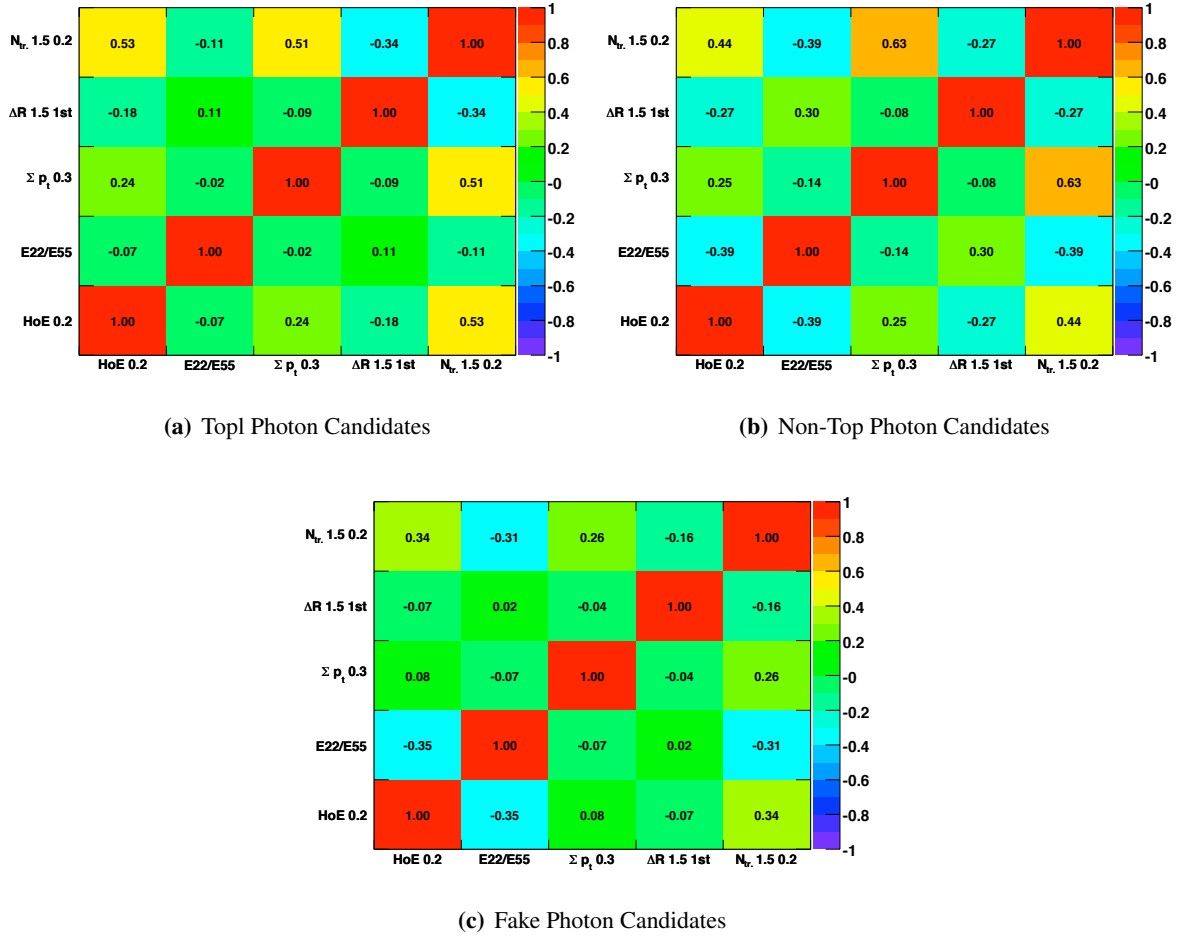


Figure 7.14: Signal Electron Events: Correlation matrices showing the correlation coefficients of the chosen identification criteria. The N-tracks criterion is strongly correlated to the other criteria for each type of photon candidate class and for that reason it is rejected.

pronounced so that this criterion respects any arbitrary track that is adjacent to the photon candidate. Thus, the distance in the $(\eta-\phi)$ -plane to the second closest track has been chosen for the muon event stream because the selection efficiency of this criterion is comparable to the other criteria of that group but is less correlated to the other variables. The correlation matrices for the chosen variable of each group are depicted in Figure 7.17. The advantage that the variables are not correlated for the fake photon candidates again dominates over the correlations for the top and non-top photon candidates which are especially observed for the HoE criterion. The large correlation between the N-tracks criteria and the other ones is respected by discarding the first.

For the selected variables the detailed distributions for top, non-top and fake photon candidates are depicted in Figure 7.18. Again, the left column presents the distributions of the three classes normalized to unity. The middle column gives information about the combined non-top and fake selection efficiency versus the top one, and in the right column the selection purity versus the selection efficiency of top photon candidates is presented. The two latter graphs are determined by systematically changing the respective cut value. Following the same arguments as for the electron event stream, selection variables are chosen for the muon event stream that differ evidently in their shape for top and non-top/fake photon candidates. In addition, the selection efficiency correlations are steeply changing and the selection purity of the top photon candidate is already relatively high when the selection efficiency of top candidates has

Fig. 7.17

Fig. 7.18

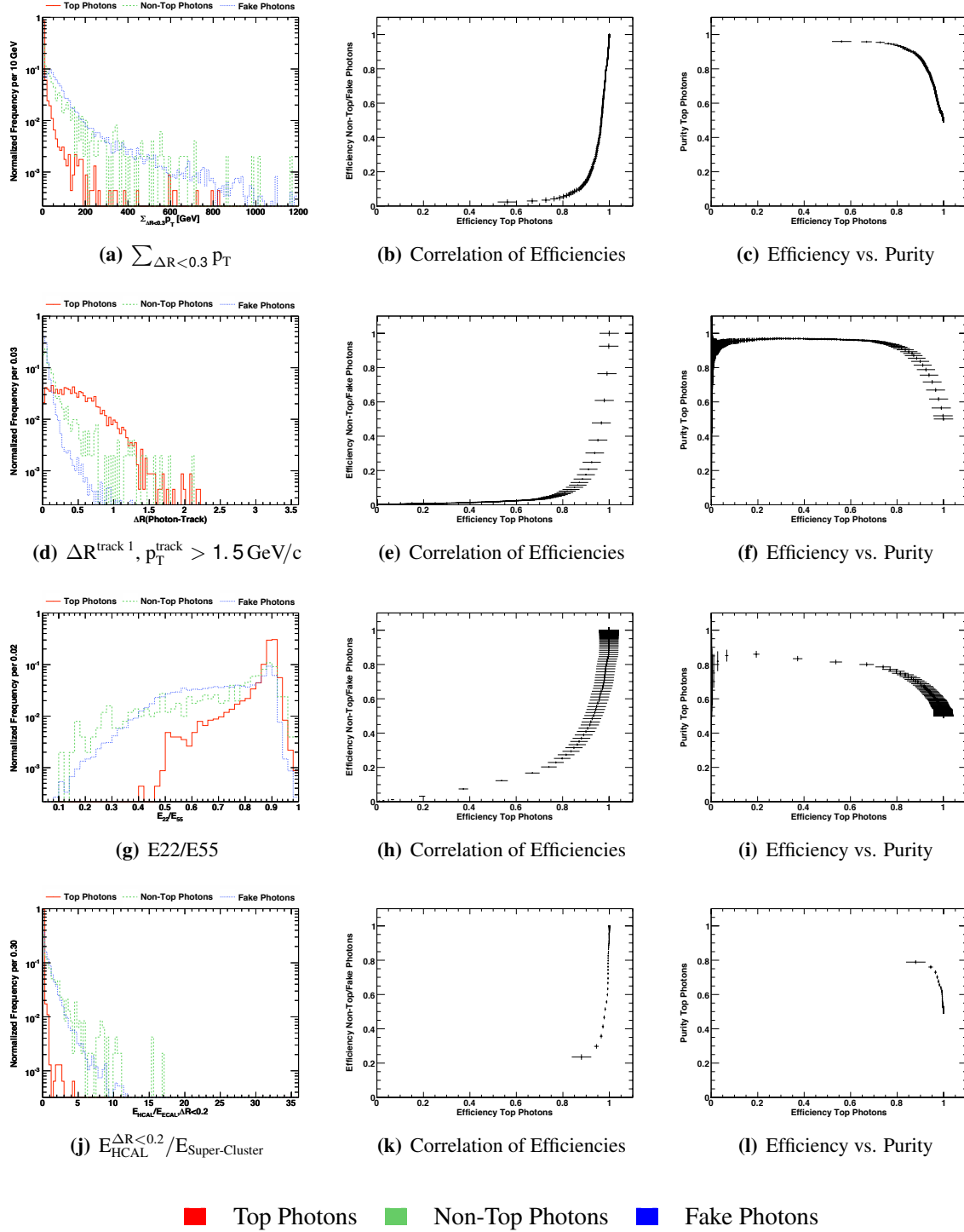


Figure 7.15: Distributions of photon identification variables for electron events. The left column reveals the normalized distributions for each of the three photon classes. In the middle column the correlation of the efficiencies of the top and the combined non-top/fake photon candidates is presented. The right column depicts the behaviour of the selection purity as a function of the selection efficiency of top photon candidates.

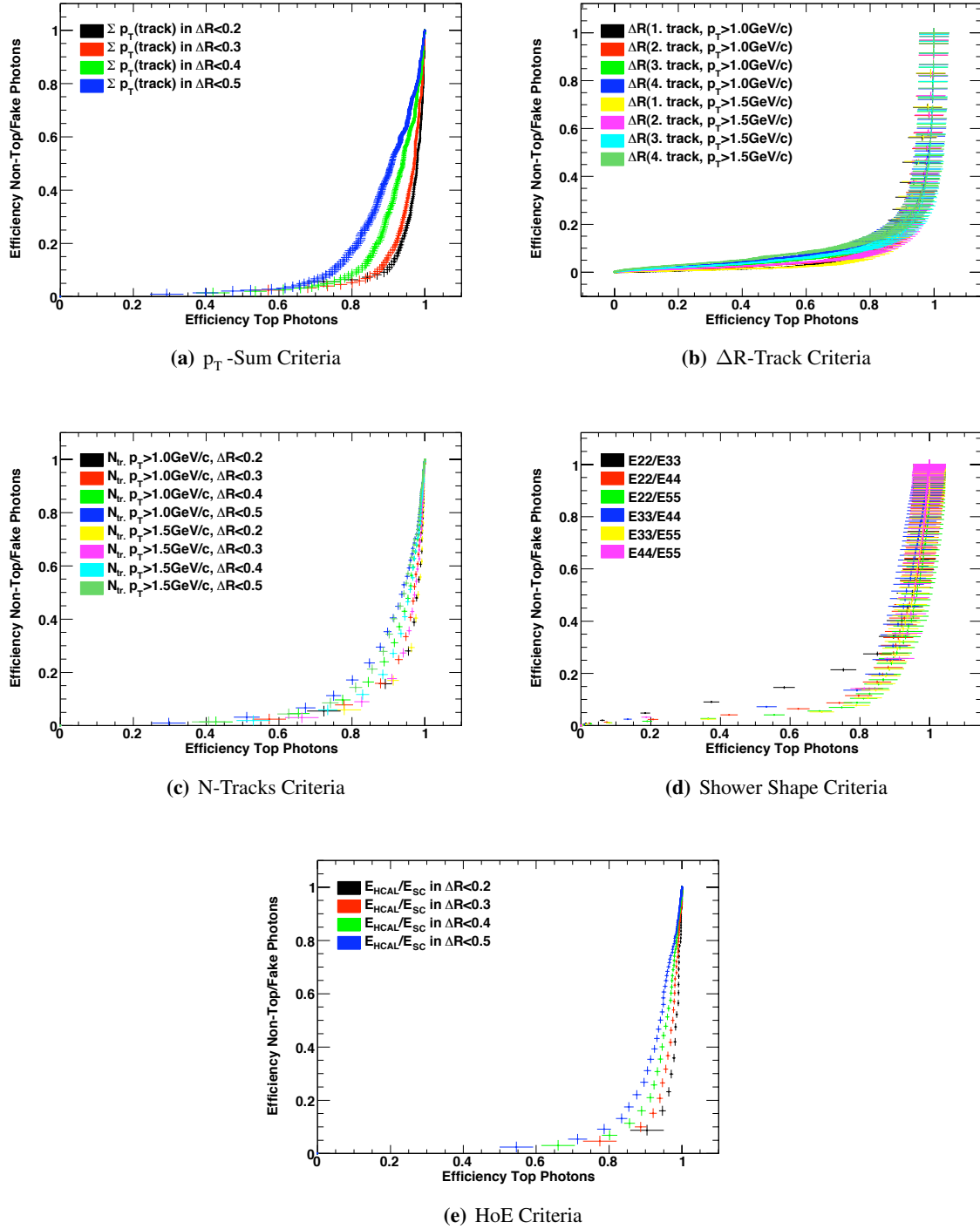


Figure 7.16: Signal Muon Events: Correlation of top photon efficiencies versus non-signal/fake photon efficiencies for the various types of identification criteria.

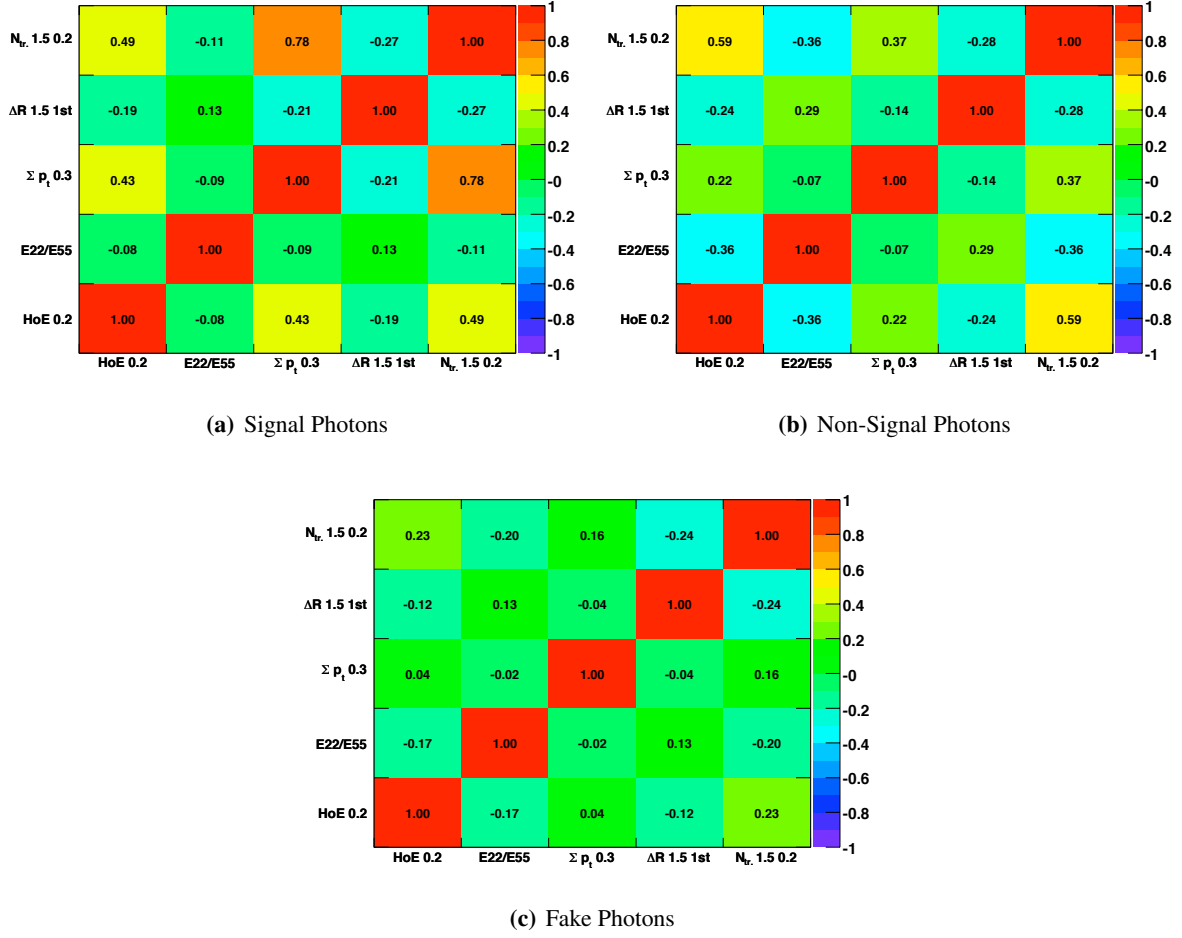


Figure 7.17: Signal Muon Events: Correlation matrices showing the correlation coefficients of the chosen identification criteria. Due to its correlation to the others the N-tracks criterion is removed from the list of possible identification criteria.

not yet fallen down too much. Again the performance of the HoE criterion is somewhat worse compared to the others.

Maximizing the product of the selection efficiency and selection purity of the top photon candidates gives the following cuts of the photon identification in the muon event stream.

$$\begin{array}{llll}
 p_T\text{-Sum Criterion} & \sum_{\Delta R < 0.3} p_T & < & 5.0 \text{ GeV}/c \\
 \Delta R\text{-Track Criterion} & \Delta R^{\text{track } 2}, p_T^{\text{track}} > 1.5 \text{ GeV}/c & > & 0.22 \\
 \text{Shower Shape Criterion} & \frac{E_{22}}{E_{55}} & > & 0.87 \\
 \text{HoE Criterion} & E_{\text{HCAL}}^{\Delta R < 0.2} / E_{\text{Super-Cluster}} & < & 0.08
 \end{array} \tag{7.11}$$

7.3.5 Results of the Photon Identification

Having applied all mentioned cuts, the photon candidates lose their mere candidate status and become upgraded to actually identified reconstructed photons according to the above discussed criteria. These cuts produce photons that are not only reconstructed but also identified as isolated particles in the de-

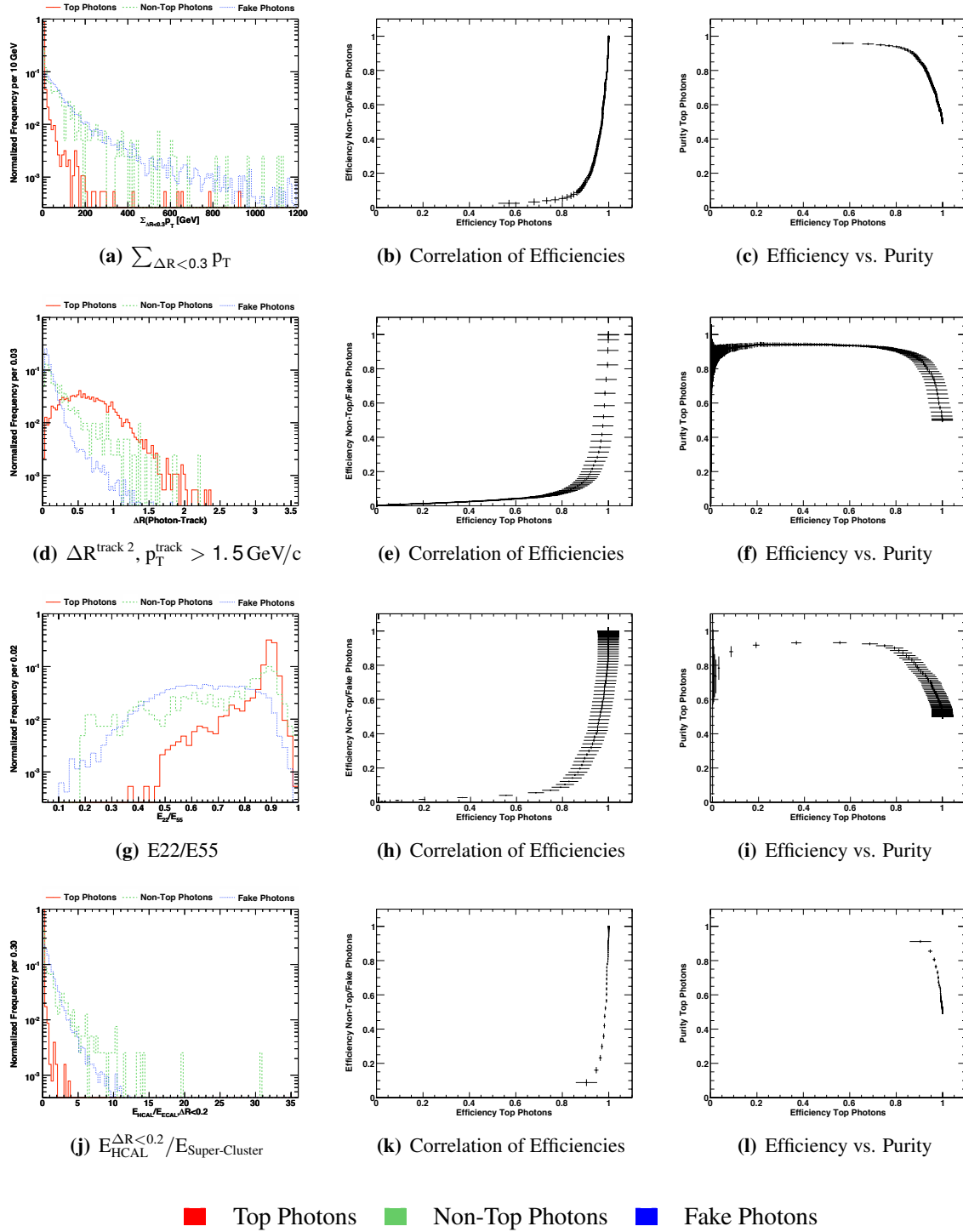


Figure 7.18: Distributions of photon identification variables for signal muon events. The left column reveals the normalized distributions for each of the three photon classes. In the middle column the correlation of the efficiencies of the top and the combined non-top/fake photon candidates is presented. The right column depicts the behaviour of the selection purity as a function of the selection efficiency of top photon candidates.

Tab. 7.4 tector. The numerical results of the photon identification are summarized in Table 7.4 and Table 7.5, while appendix A contains detailed event numbers when splitting the background data-sets to its single constituents.

App. A

	Signal Electron Events	Signal Muon Events
Event Efficiency	42.1%	43.3%
Top Photon Purity (only e or μ)	88.1%	92.3%
Top Photon Purity (e and μ)	88.1%	92.1%
Average per Event	1.041	1.020

Table 7.4: Selection efficiency and purity of the photon identification for signal electron and muon events. At first, the selection purity of the top photon is calculated for top photons in the electron event stream of signal electron and the muon event stream of signal muon events only. Then, the combined purity is calculated including the opposite stream as well. The average per event gives the mean number of photons that are identified per event.

	Electron Event Stream				Muon Event Stream			
	Events	TP	NTP	FP	Events	TP	NTP	FP
Signal Electron	167	154	6	15	0	0	0	0
Signal Muon	0	0	0	0	130	122	2	8
$t\bar{t}$ + Jets ($\cdot 10^3$)	14	0.70	4.46	10.51	6.73	0.71	3.13	3.02
W + Jets ($\cdot 10^3$)	13	-	5.82	7.36	2.40	-	1.74	1.04
Z + Jets ($\cdot 10^3$)	12	-	3.09	8.64	0.54	-	0.46	0.08
WW	91	-	43	48	14	-	12	5
WZ	77	-	24	53	5	-	0	5
ZZ	63	-	23	51	5	-	5	5
$W\gamma$	23	-	22	1	2	-	2	0
γ + Jets	2	-	2	1	0	-	0	0
QCD-multijet ($\cdot 10^3$)	28	-	15.00	12.65	0	-	0	0

Table 7.5: Number of photons after the photon identification scaled to $\mathcal{L} = 10 \text{ fb}^{-1}$. A convenient statistical interpretation in case that zero events are left for any data-set is given in chapter 9.

At first, it can be derived from the results of the signal events that the selection purity is higher and the mean number of photons left per event is lower for the muon event stream than for the electron stream. There is still an ambiguity between electrons and fake photons which is discussed and resolved in chapter 7.4.1.

Chp. 7.4.1

When the selection results of the background events are considered it must again be emphasized that the muon stream gives better results than the electron stream. The number of events is still significantly higher especially for the non- $t\bar{t}$ data-sets because of the again larger number of fake photons. By means of the photon identification the events of the muon stream are so far suppressed by approximately one order of magnitude compared to the electron stream. This aspect become even more evident when examining the selection results of the $t\bar{t}$ + Jets events. For the muon stream it has become the largest contribution to the background while for the electron stream the W + Jets, Z + Jets and QCD-multijet sample contribute still to the same order of magnitude. Hence, the photon identification is another important handle to reduce the background data. This is pointed out as well by another remarkable aspect. Even before any assumptions about a global event reconstruction are made the contribution of

the QCD-multijet and γ + Jets-background starts to get eliminated, especially for the muon event stream.

The global success of the photon identification is underlined by Figure 7.19 and Figure 7.20. The remaining fake photons in the signal electron and signal muon samples are investigated to which generator particles the fake photons match according to the criteria of equation 7.4. The same procedure as in chapter 7.3.2 has been applied. As it can be seen there is still an electron contamination for the signal electron events 7.19 (a), but regarding the more adequate distance criterion the majority ($> 85\%$) of fake photons match to a photon, which is either a decay product of a π^0 -decay or actually a generator top photon. For the muon sample a very similar situation can be observed expect for the electron contribution which is negligible.

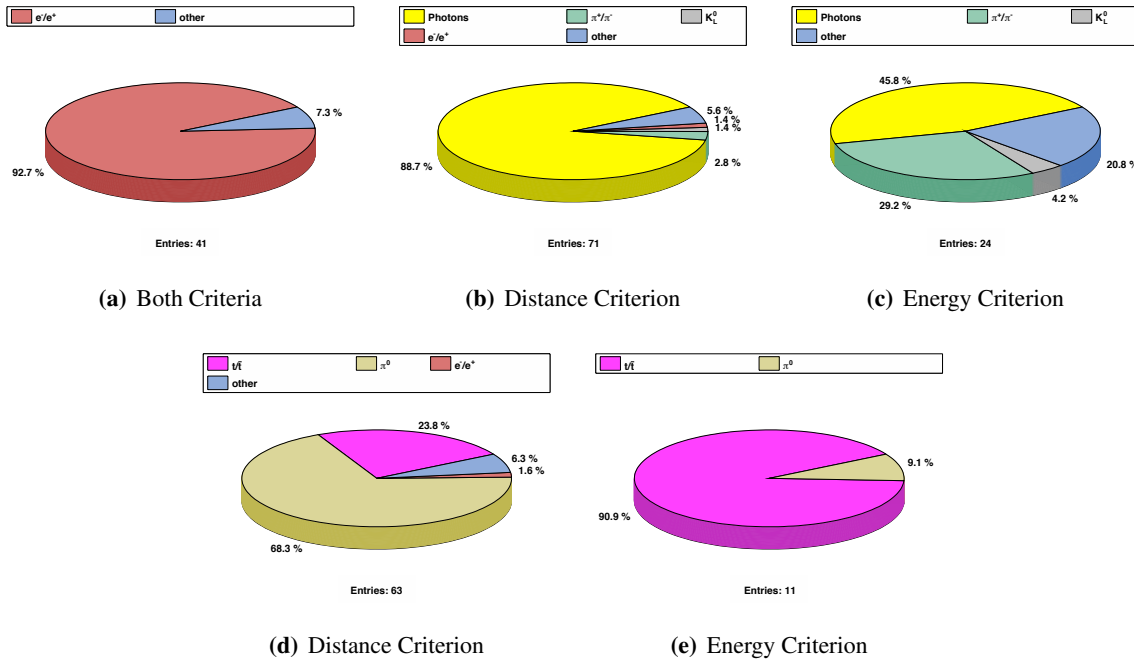


Figure 7.19: Signal Electron Events, Fake Photons after Photon Identification. The entries per chart are not scaled to the luminosity of $\mathcal{L} = 10 \text{ fb}^{-1}$.

At last, the distance between the photons in the $(\eta-\phi)$ -plane has been considered, assuming that two photons from a π^0 -decay are close together but sufficiently separated that they are resolved as two objects with a small distance ΔR . This handle to reduce the π^0 -background has proven to be implicitly included in the identification variables by the shower-shape ratio $E22/E55$. If there is a second photon close to another one then the energy content in the 5×5 -square of electromagnetic calorimeter cells does not fall steeply enough compared to the 2×2 -square. Thus, both photon candidates are already rejected by this criterion.

In summary, the photon identification is very efficient because real photons are indeed identified even if they are not exclusively radiated off a top-quark. Both event streams suffer from the irreducible admixture of photons from π^0 -sources. The results are so far better for the muon stream than for the electron stream because there is still a significant overlap between the reconstructed electron and photon objects.

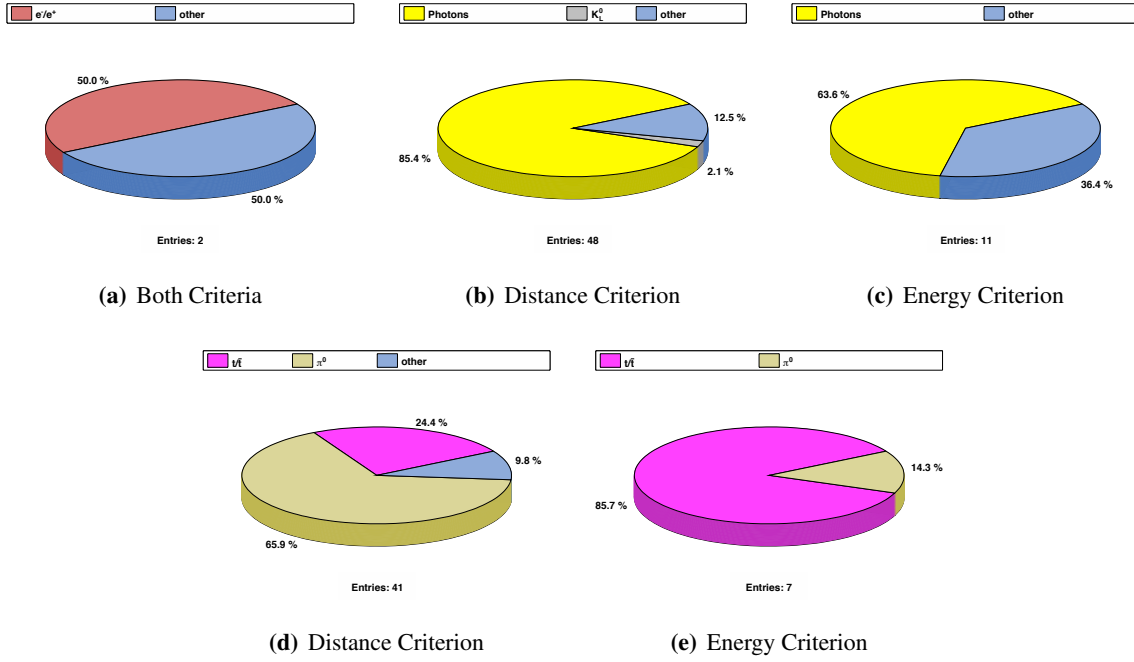


Figure 7.20: Signal Muon Events, Fake Photons after Photon Identification. The event numbers are not scaled to the luminosity of $\mathcal{L} = 10 \text{ fb}^{-1}$.

7.4 The Particle Correlations

Chp. 5.5

As described in chapter 5.5 every reconstructed electromagnetic super-cluster becomes a photon candidate. If these are further upgraded to electron candidates or included in the calorimeter towers of jet candidates they are not removed from their candidate list. Thus, having identified electrons and photons the correlations between these three types of objects are investigated to resolve these ambiguities.

7.4.1 Correlations between Photons and Electrons

Fig. 7.21

In order to examine the correlations between photons and electrons their distance in the $(\eta-\phi)$ -plane and their energy ratio are plotted in Figure 7.21, separated for the three classes of photons. It can be observed

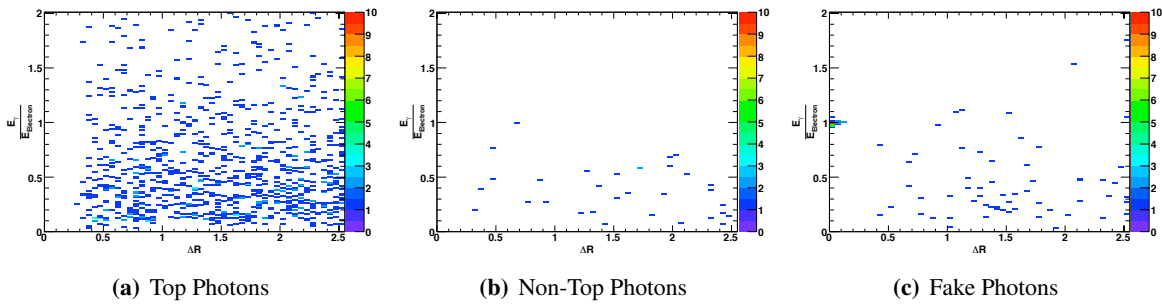


Figure 7.21: Correlation between the ΔR -distance in the $(\eta-\phi)$ -plane and the ratio of the electron to the photon energy for signal electron events. While top (a) and non-top (b) photons are only randomly correlated to electrons, there is an agreement between fake photons and electrons (c).

that there is an increased counting rate close to the point (0,1) for fake photons, while for top and non-top photons the histograms are only randomly populated. That means that there exist fake photons pointing in the same direction as the electrons having very similar energy values. Thus, these fake photons can be identified as electrons and therefore an additional two-dimensional cut is introduced. It is demanded that a photon must lie outside the following area in the $(\Delta R - E)$ -plane when being compared to an electron

$$\begin{aligned} \Delta R &< 0.15 \\ \frac{E_\gamma}{E_{\text{Electron}}} &\in \{0.9, 1.1\}. \end{aligned} \quad (7.12)$$

In order to answer the question why these correlations still exist after the electron and photon identification, both explicitly respecting a tracker isolation criterion, the matching of the electron candidates to a track in the silicon tracker can give a hint. There is no tight cut on the ratio of the super-cluster energy and the momentum of the track (E/p) imposed on the electrons candidates. Though there is an excess at unity, that is proven by the distribution in Figure 7.22 revealing the E/p -ratio of the electrons, which match to a photon according to the above introduced cuts, there exist several electrons which only poorly match to a track. Fig. 7.22

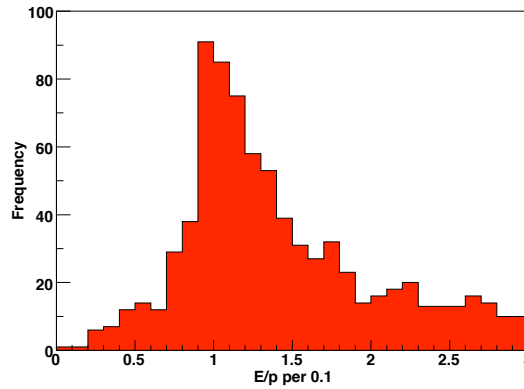


Figure 7.22: Ratio of the electron super-cluster energy to the momentum of the matched track. Only electrons are integrated into that distribution that lie inside the area defined by equation 7.12. In order to increase the statistics not only the signal electron but also the $t\bar{t}$ + Jets-events are included in the histogram.

The correlations between photons and muons have also been investigated, but no agreement has been observed.

7.4.2 Correlations between Photons and Track Seeds

As it has been explained in chapter 5.5 a photon candidate must have a close track match to be upgraded Chp. 5.5 to an electron candidate. If the matching criteria are not fulfilled there is still the possibility that the super-cluster and by that the photon candidate is pointing towards a track seed. As this can be the origin of a track which is not assigned to the super-cluster the tracker isolation criteria might fail in dropping these photons. The last aspect of the photon identification is therefore to delete any still remaining photon out of the collection which is pointing to a track seed. By that about 36% of the fake photons are removed, while the list of top and non-top photons is reduced by only approximately 1% for the signal electron events.

For signal muon events the impact of this correlation is less pronounced. About 10% of the fake photons are dropped and again only approximately 1% of the top and non-top photons are deleted.

7.4.3 Correlations between Photons and Jets

The correlation between photons and jets are not meant to get a more precise identification of photons but to remove jets which are photons more probably. In order to disentangle this correlation the dependence of the distance in the $(\eta-\phi)$ -plane between a photon and a jet versus the electromagnetic energy fraction of the total jet energy is studied, as it is shown in Figure 7.23(a) for the signal electron events and Figure 7.24(a) for the signal muon events. The increase in the density of points in the upper left corner consists to about 75% for signal electron events and almost completely for signal muon events of correlations between top photons and jets. Thus, this correlation is not capable of further reducing the number of non-top or fake photons without losing too many top photons, but clearly indicates that there are jets being more probable photons.

This is emphasized when comparing the direction of the reconstructed photons to the generator quarks of the $t\bar{t}$ -decay are shown being the origin of the high-energetic jets in the signal data-sets. In Figure 7.23(b) and Figure 7.24(b) the distance in the $(\eta-\phi)$ -plane between the reconstructed photons and the generator quarks. The low excess in the first bin is somewhat more pronounced for signal electrons events but not significant enough to explain the entire correlation. Therefore, the assumption to remove these jets because they are actually photon is considered to be correct.

Finally, a jet is dropped from the collection if it fulfills the two following criteria with respect to any reconstructed photon.

$$\begin{aligned} \Delta R &< 0.10 \\ \frac{E_{EM}}{E_{Jet}} &> 0.85 \end{aligned} \quad (7.13)$$

The importance to resolve this correlation becomes obvious when taking into account the index of the dropped jet in a p_T -ordered list (see Figure 7.23(c) and Figure 7.24(c)). Even the highest-energetic jets are affected. This is relevant because in the next step of the analysis the leading four or five jets are kinematically combined to a W-boson and a top-quark. As the physical model expects these jets to be derived from high-energetic hadronic objects, the correlation between photons and jets modifies and improves the input data for this stage.

7.4.4 Correlations between Electrons and Jets

The correlation between electrons and jet have already been implemented in the Top Quark Analysis Framework. If the distance between an electron candidate and a jet is less than $\Delta R = 0.3$ then the jet has been discarded if the the electron is respecting a preliminary isolation against high-energetic tracks (sum over track p_T must not exceed 3 GeV/c in that isolation cone).

7.4.5 Results of the Particle Correlation Investigations

The entire cleaning procedure can reduce the numbers of photons and jets per event such that either no more photon is present in an event or the number of jets is less than four. Hence, these events are deleted. Table 7.6 and Table 7.7, respectively, give new results on the event and photon numbers replacing the results of chapter 7.3.5, having now finally finished the photon identification.

The mean photon number for both stream has been lowered and the selection purity of top photons has increased (compare Table 7.4). Although the results for the electron stream are still slightly worse than for the muon stream both values are converging. That indicates that for both streams the same types of photons remain, either top photons or they stem from the irreducible background of π^0 -meson decays.

This is furthermore confirmed when contemplating the background events. First and foremost due to the removal of fake photons, the events of the electron stream are reduced by at least a factor of two for the

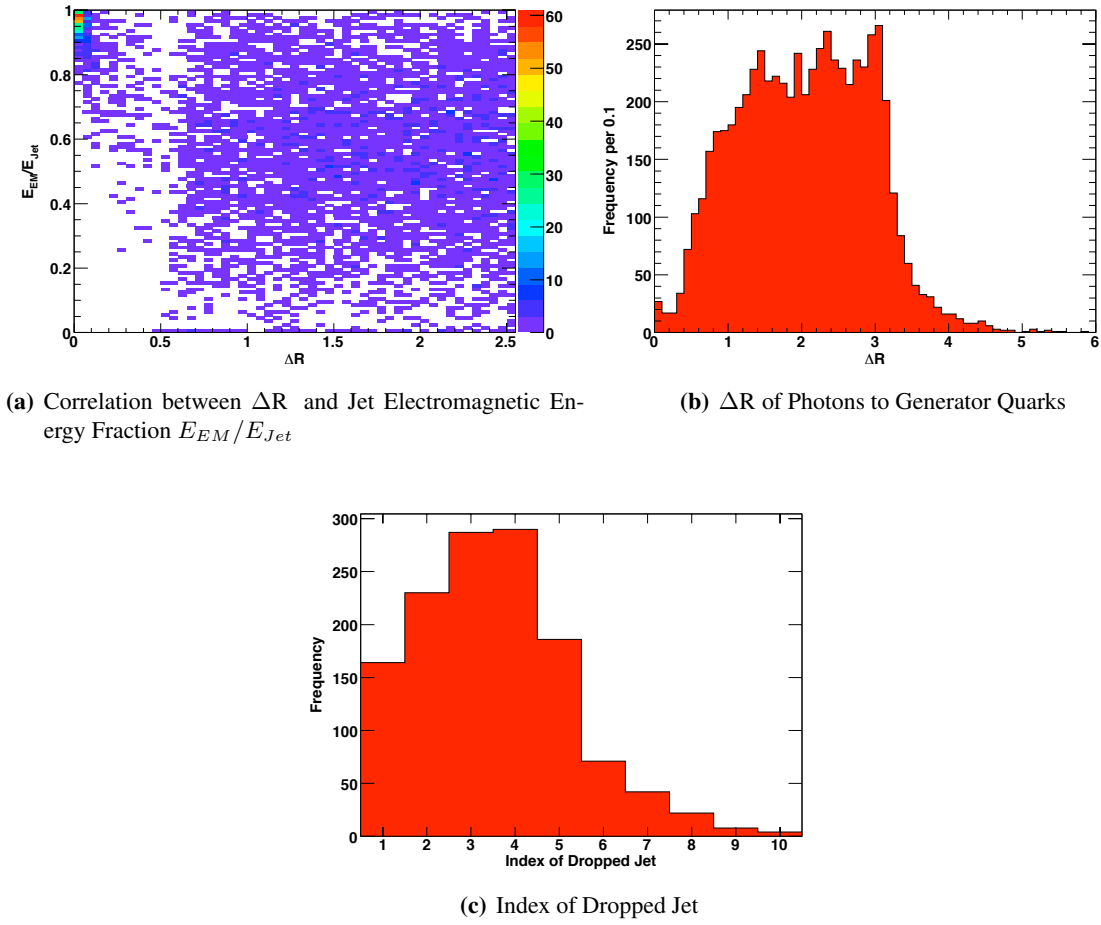


Figure 7.23: Signal Electron Events: Results of the investigation of correlations between photons and jets. In (a) the correlation between the distance in the $(\eta-\phi)$ -plane ΔR of the photon to the jet versus the electromagnetic energy fraction of the jet is shown. The figure (b) indicates that there is a real correlation because the photons do not match to the generator quarks which have formed the jets. In (c) the indices of the jets in a p_T -ordered list is shown which are removed due to the correlation.

	Signal Electron Events	Signal Muon Events
Event Efficiency	35.3%	39.6%
Signal Purity (only e or μ)	90.9%	92.8%
Signal Purity (e and μ)	90.9%	92.7%
Average per Event	1.023	1.016

Table 7.6: Selection efficiency and purity of the photon identification after resolving particle correlations for signal electron and muon events. At first, the selection purity of the top photon is calculated for top photons in the electron stream of signal electron events and the muon stream of signal muon events only. Then the combined purity is calculated including the opposite stream as well. The average per event gives the mean number of photons that are identified per event.

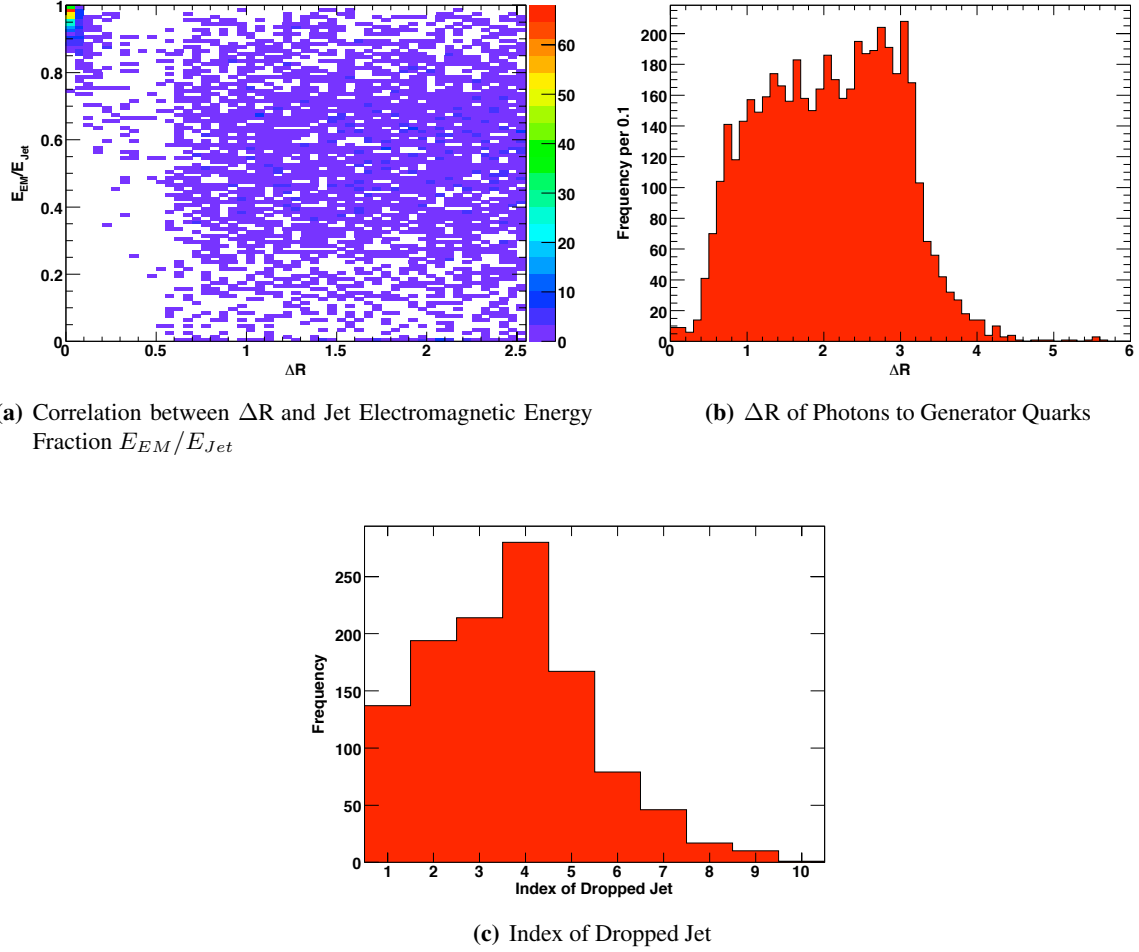


Figure 7.24: Signal Muon Events: Results of the investigation of correlations between photons and jets. In (a) the correlation between the distance in the $(\eta-\phi)$ -plane of the photon to the jet versus the electromagnetic energy fraction of the jet is shown. The figure (b) indicates that there is a real correlation because the photons do not match to the generator quarks which have formed the jets. In (c) the indices of the jets in a p_T -ordered list is shown which are removed due to the correlation.

$t\bar{t}$ + Jets and W + Jets samples up to several orders of magnitude for the QCD-multijet data-sets. The modifications of the muon stream data are less pronounced, so that the same converging tendency can be observed for the background data as for the signal samples.

In summary, at this stage of the event selection the di-boson, γ + Jets, and QCD-multijet samples are reduced to the same order of magnitude or even below as the signal samples even before the $t\bar{t}$ -signature of the events is exploited. In contrast, the contribution from the $t\bar{t}$ + Jets and the W/Z + Jets data-sets are still enhanced by a factor of more than 10 for the electron stream compared to the signal electron events. For the muon stream the situation can be evaluated somewhat more optimistic because the W/Z + Jets contribution is significantly lower than for the electron stream. But the remaining number of $t\bar{t}$ + Jets events is comparable for both event streams again indicating what the most important and dominating background will be.

	Electron Stream				Muon Stream			
	Events	TP	NTP	FP	Events	TP	NTP	FP
Signal Electron	140	131	5	8	0	0	0	0
Signal Muon	0	0	0	0	118	111	2	6
$t\bar{t} + \text{Jets}(\cdot 10^3)$	7.66	0.60	3.45	3.73	5.67	0.68	2.85	2.25
W + Jets ($\cdot 10^3$)	5.46	-	3.37	2.14	2.07	-	1.50	0.68
Z + Jets ($\cdot 10^3$)	1.75	-	0.96	0.82	0.42	-	0.37	0.07
WW	29	-	12	17	14	-	12	2
WZ	36	-	18	18	5	-	0	5
ZZ	9	-	9	0	2	-	2	0
$W\gamma$	8	-	7	1	2	-	2	0
$\gamma + \text{Jets}$	1	-	1	1	0	-	0	0
QCD-multijet	138	-	0	138	0	-	0	0

Table 7.7: Number of photons after photon identification and implementing the particle correlations scaled to $\mathcal{L} = 10 \text{ fb}^{-1}$. A convenient statistical interpretation in case that zero events are left for any data-set is given in chapter 9.

Chapter 8

The Event Reconstruction

THE present chapter is devoted to the entire reconstruction of the events according to the semileptonic $t\bar{t}$ -decay signature. At first, properties of the participating jets are investigated to exploit the b-tag information and to check if tighter kinematic cuts are necessary. Then the subject of the missing transverse energy is addressed, before the still separated objects are re-connected according to the decay model of Figure 5.2(c) to create a semileptonic $t\bar{t}$ event solution. The intermediate steps of the $t\bar{t}$ -decay are reconstructed by searching for these two jets which have most likely been the decay products of the hadronically decaying W-boson, and by identifying a third jet so that a top-quark is reconstructed. In the end, additional cuts are presented to judge the quality of the event reconstruction. It will be shown that all background events not including $t\bar{t}$ -pairs will disappear after these steps. Fig. 5.2(c)

8.1 The Identification of b-Jets

The decay of a top-quark in the Standard Model requires the existence of a b-quark that is detected as a b-jet in the detector. The reconstruction and identification of a b-jet in the detector is dominated by the life-time of the associated b-hadron which only decays at an average distance of approximately 0.5 mm from the interaction point. If the tracks of charged particles in b-jets are traced back to their origin, they should cross at this displaced secondary vertex, while jets from light quarks or gluons should point to the primary vertex. The limits on this spatial difference are set by the efficiency of track reconstruction, by the track resolution and by the efficiency to determine the primary vertex as reference point.

For the present analysis the *track counting algorithm* [68] has been used to identify b-jets. This algorithm Ref. [68] evaluates the impact parameters for each track associated to a jet and satisfying some quality criteria (e.g. number of hits, quality of the track fit, etc.). The discriminating variable of the algorithm is the significance of the impact parameter which is stored in a descending list. It has been shown in [68] that Ref. [68] the discriminator value of either the second or the third track in that list provides the best separation between b-jet and the other jet types. The first one give results with a high selection efficiency, while the second one optimizes the purity of the b-jet identification. The entries for the b-tag discriminator can be positive or negative depending on the relation of the reconstructed jet axis and the impact parameter vector. If both are pointing in the same hemisphere defined by the jet axis then the impact parameter is positive, otherwise negative. That implies that negative discriminator values belong to tracks not fitting well to the jet, either due to bad track reconstructions or due to badly identified jet directions or primary vertices.

In Figure 8.1 the distributions of the b-tag discriminator variable (significance of second track) are shown Fig. 8.1 for the highest and second highest b-tag. The existence of b-jets is kinematically suppressed if they are created in a radiation process due to the relatively large mass of the b-quark ($m_b \approx 4.25 \text{ GeV}/c^2$).

The b-tag values especially for the data-sets including vector bosons (W/Z + Jets, di-boson) have their maximum at lower b-tag values than the top-quark events (signal and $t\bar{t}$ + Jets). There the b-quark, and in the following the b-jet, is a daughter particle of the top-quark and occurs naturally from its decay.

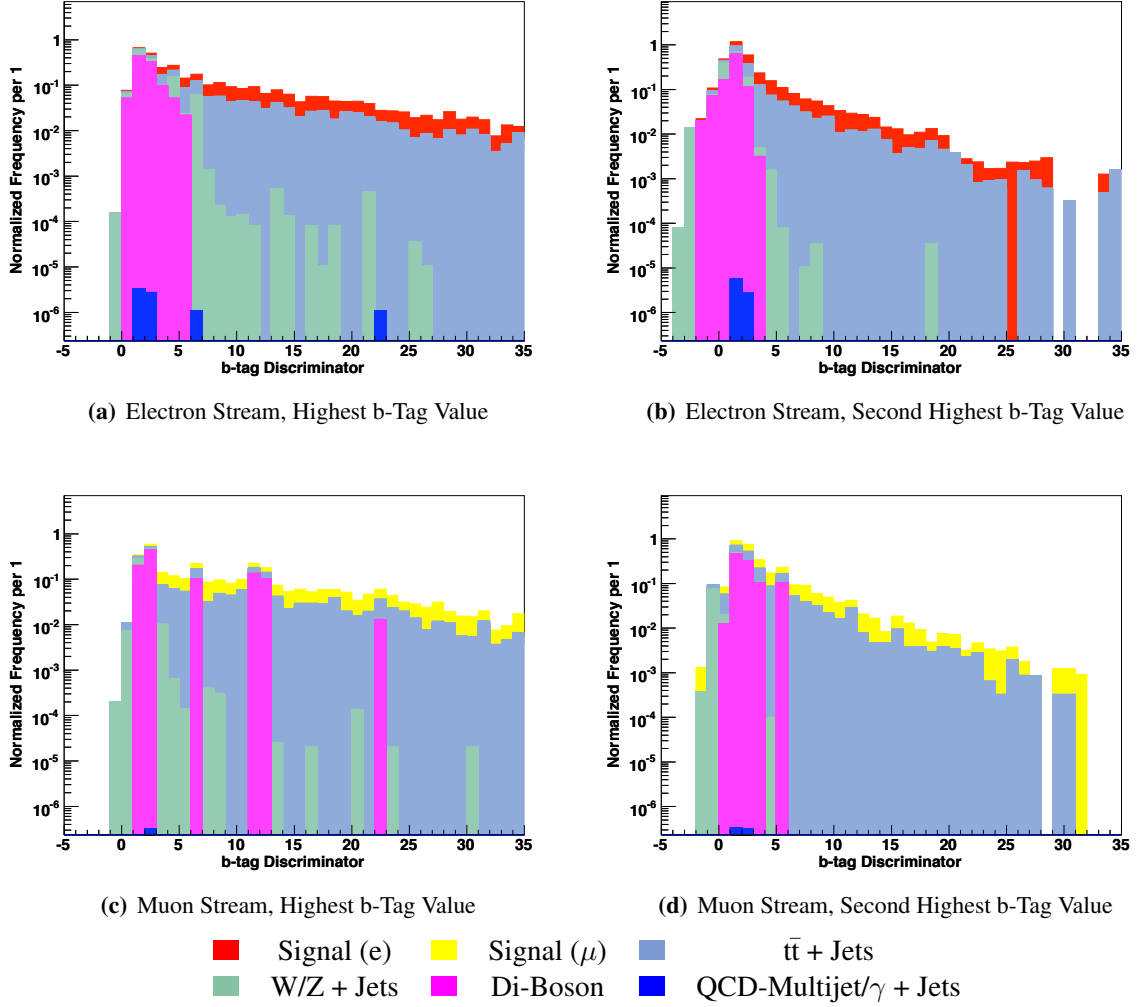


Figure 8.1: Distributions of the highest and second highest b-tag discriminator (significance of second track) values for the different data samples reveal that the non- $t\bar{t}$ data-sets have systematically lower entries than the $t\bar{t}$ -samples where a b-jet is not created by radiation but by the top-quark decay. The histograms are stacked. The distribution for each data-set is normalized to unity and then scaled according to the relative fraction of the total cross-section of a data-set in case that it consists of more than one sub-data-set.

The distributions for the highest and the second highest b-tag discriminator values reveal a characteristic behaviour. While the highest b-tag discriminator value tends towards higher values, the histogram of the second highest value shows a sharp limit between the $t\bar{t}$ - and non- $t\bar{t}$ -samples. Thus, a cut on the second highest b-tag values has been introduced.

$$B_{2\text{nd}} > 5.0 \quad (8.1)$$

If the second jet is required to pass this cut the jet with the highest b-tag is implicitly tagged as a b-jet with a discriminator value of that value or greater. After applying this cut signal and $t\bar{t}$ + Jets-events remain almost exclusively. Only single events of the W/Z + Jets data-sets have passed this b-tag cut in the electron event stream.

While the distinction between the $t\bar{t}$ -events and non- $t\bar{t}$ samples is efficiently working already for the discriminator using the significance of the second track, no additional separation between the signal and the $t\bar{t}$ + Jets samples is possible when using the discriminator basing on the significance of the third track. For both samples and analysis streams the efficiency is about 30% for the first case, and approximately 10% for the latter. Therefore, the second track discriminator is selected to retain more signal events.

8.2 The Jet p_T -Distributions

After requiring jets to be b-tagged almost all non- $t\bar{t}$ events are rejected and in the following the focus is exclusively concentrated on a probable distinction between the remaining signal and the $t\bar{t}$ + Jets events. As the latter might include a variety of jets that originate from radiation processes it is possible that there is a difference in the transverse momentum distributions. Though this investigation is a method in the realm of top-physics [69] for the separation of $t\bar{t}$ -events from non- $t\bar{t}$ -events, it can be tested if it is nevertheless applicable for the distinction between semileptonic $t\bar{t}\gamma$ -events and mainly dileptonic $t\bar{t}$ + Jets-events. Ref. [69]

In Figure 8.2 and Figure 8.3 the p_T -distributions of the leading four jets are compared. If a jet is created by gluon radiation rather than by the hard interaction or a hadronic decay then its momentum is expected to be lower on average. Therefore, as most background processes involve only one high-momentum reaction or decay, which is also concurring with the need for one high-energetic lepton, the jets are assumed to be created by these radiation processes. For the $t\bar{t}$ + Jets-events there is only a difference to the signal events expected if the dileptonic decay channel still dominates the $t\bar{t}$ + Jets-sample. This is the only channel which gives two jets from top-quark decays and can keep the number of high-energetic jets below four if not more than one additional jet is involved in the hard interaction. Fig. 8.2
Fig. 8.3

This has not been observed and a statistical analysis of the number of dileptonic events in that sample gives an admixture of 21% (18%) for the electron (muon) stream in total. This is not negligible but the fraction with zero or one additional high-energetic jets is only 3% for both cases, so that an evident difference in the p_T -spectra is not expected.

In summary, no further kinematic cut on the transverse momentum of the jet objects is applied.

8.3 The Missing Transverse Energy

The missing transverse energy (MET) is a measure that accounts for the kinematic balance of the event in the plane transverse to the beam axis. As the p_T of the colliding partons is to a first approximation nearly zero, each object which is detected with a transverse momentum must be kinematically compensated by at least one other object. Thus, the MET is calculated as the missing counter-part for an entire event to balance the total sum of all detected transversal contributions.

The physical origin can be found in three sources. At first, neutrinos or other new rarely interacting particles can carry away energy and momentum. In the semi-leptonic $t\bar{t}$ -decay channel a neutrino is created via a W-boson decay. This is the best case because MET is a real physical observable then. But detector effects can contribute to that measure as well by imperfections like dead readout channels or by an imperfectly hermetical coverage in the forward directions. Objects can either pass through or leave the detector without a signal being registered. Finally, the resolution of the energy measurement in the calorimeters determines how precisely the MET can actually be measured. In case of the CMS detector this is the most important contribution so that any distribution of the MET in the energetic range of the neutrino originating from a top-quark decay is smeared by that effect. Hence, it is not expected to observe any difference in the MET spectra.

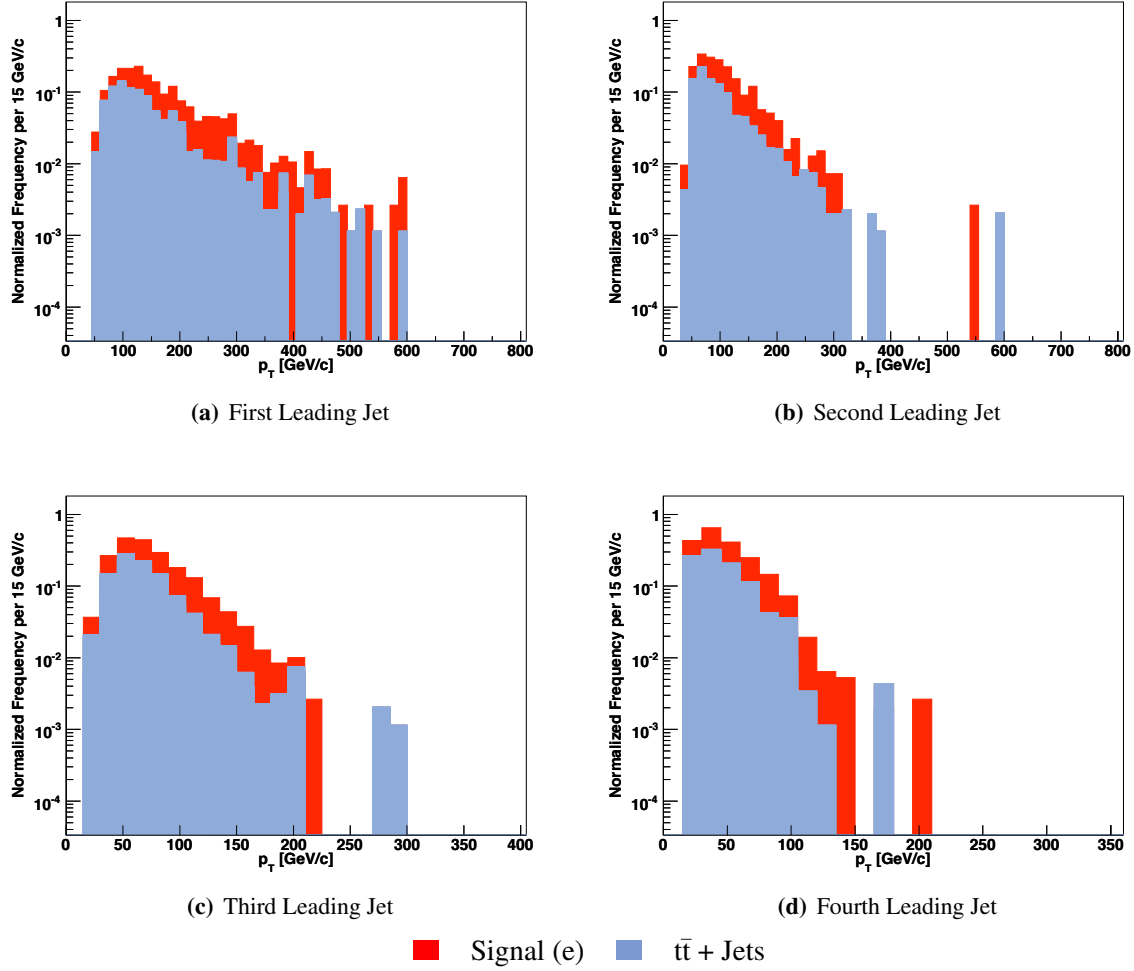


Figure 8.2: Distributions of the p_T -spectra of the four leading jets for the electron stream. As the admixture of the di-leptonic $t\bar{t}$ + N Jets ($N = 0,1$) events is small no significant difference in the shape is visible. The histograms are stacked. The distribution for each data-set is normalized to unity and then scaled according to the relative fraction of the total cross-section of a data-set in case that it consists of more than one sub-data-set.

Fig. 8.4

In Figure 8.4 the distribution of the missing transverse energy is depicted. The shape is the same for the signal and the $t\bar{t}$ + Jets samples. Even if the detector resolution effect are neglected the expectations are met because the remaining fraction of the all-hadronic $t\bar{t}$ -decay channel at this stage of the analysis is approximately 0.2%. This is theoretically the only source that does not create a neutrino when decaying and therefore can lead to any difference between the signal and $t\bar{t}$ + Jets events.

In summary, no cut on the missing transverse energy is implemented.

8.4 The Kinematic Event Solutions

For the kinematic event solutions it is exploited that some of the jets are kinematically correlated. The hadronically decaying top-quark has produced three jets, out of which two are daughter objects of a W-boson. When combining the correct ones these kinematic information can be regained and potentially used to suppress events with an admixture of jets which originate from radiation effects and are therefore

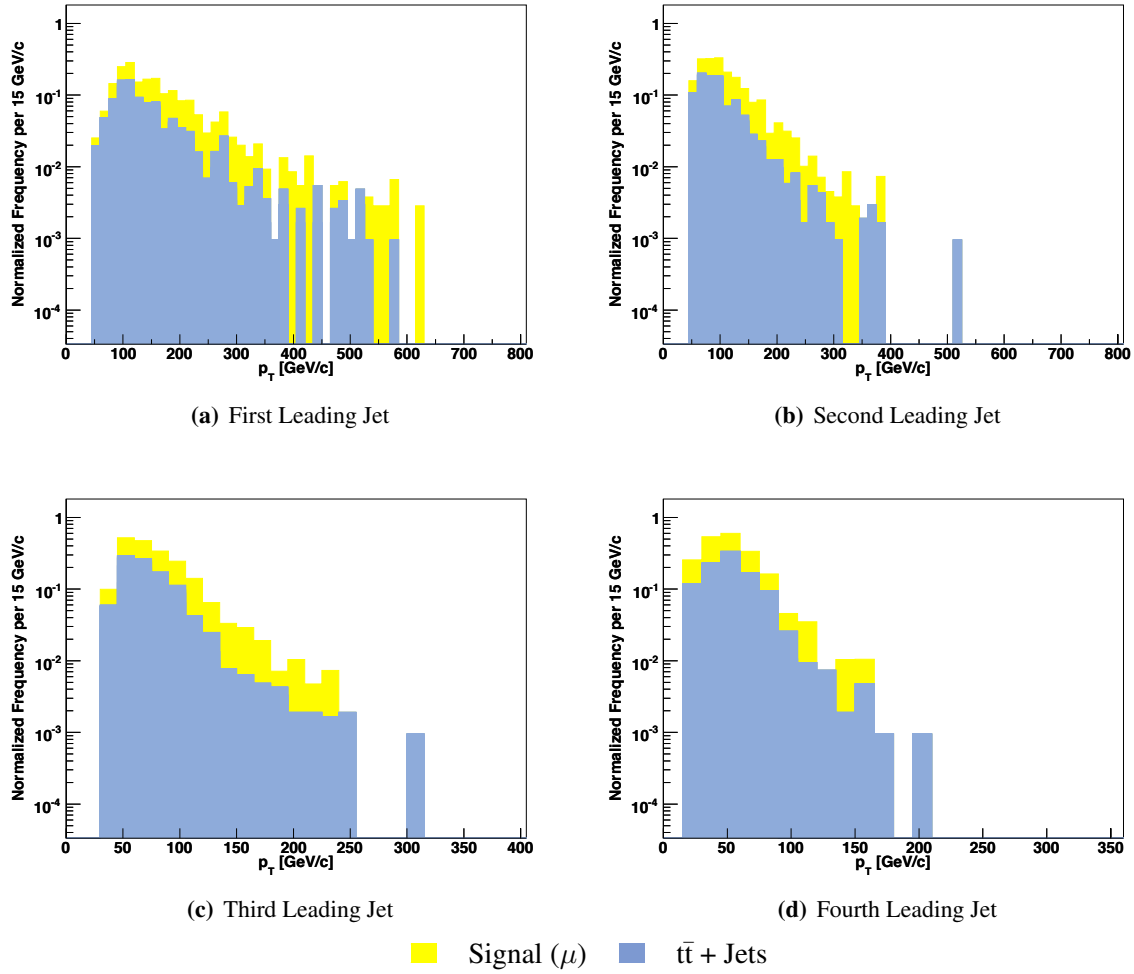


Figure 8.3: Distributions of the p_T -spectra of the four leading jets for the muon stream. As the admixture of the dileptonic $t\bar{t}$ + N Jets ($N = 0,1$) events is small no significant difference in the shape is visible. The histograms are stacked. The distribution for each data-set is normalized to unity and then scaled according to the relative fraction of the total cross-section of a data-set in case that it consists of more than one sub-data-set.

uncorrelated. This reconstruction process is done when generating the event kinematic solutions.

The main routine to build different event solutions is a kinematic fitting algorithm [70]. This package ^{Ref. [70]} is a common tool incorporated in the CMS software and it is interfaced by the second layer of the Top Quark Analysis Framework (see chapter 7.1). An adjustable number of jets, starting with this one with ^{Chp. 7.1} the highest p_T , is taken and the algorithm combines them systematically, respecting the semileptonic top-quark-pair decay hypothesis. During that fit routine the kinematic parameters of the jets are varied within the ranges of the resolutions provided to the fitting procedure as input data. The algorithm can be configured furthermore by setting kinematic constraints. In the present analysis two of them are used.

- W-boson Mass Constraint

The invariant mass of two jets must be as close as possible to $M_W = 80.35 \text{ GeV}/c^2$.

- Top-Quark Mass Constraint

The invariant mass of the latter two combined with a third jet must be as close as possible to $m_t = 175.0 \text{ GeV}/c^2$.

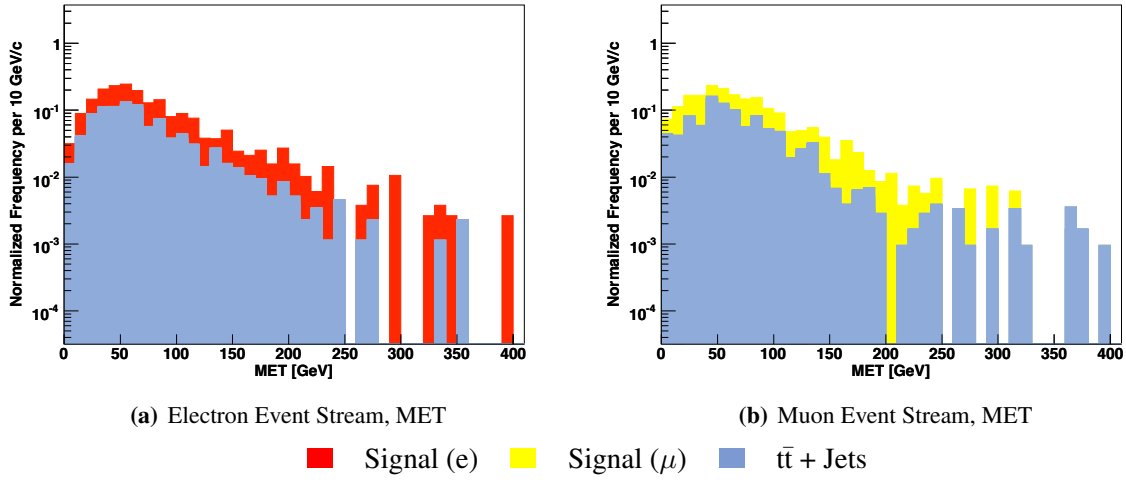


Figure 8.4: MET histograms for the electron (a) and the muon stream (b). The number of all-hadronic $t\bar{t}$ + Jets events is negligible. Hence, no cut on the MET is necessary. The histograms are stacked. The distribution for each data-set is normalized to unity and then scaled according to the relative fraction of the total cross-section of a data-set in case that it consists of more than one sub-data-set.

After the kinematic fit a reconstructed jet is identified for each hadronic object of the semileptonic $t\bar{t}$ -decay and the energy of the jet is corrected. As there are different correction functions for the various types of jet flavours each jet is corrected with respect to its assumed identity in the $t\bar{t}$ -decay chain.

Ref. [71]

This procedure is followed for each combinatorial variation of the jets so that a measure must be evaluated to determine the best solution. A jet likelihood ratio [71] is exploited to get the most probable of all calculated solutions. In the end, the chosen solution gives an event-specific response which jet has most probably been originating from which parton in the semileptonic $t\bar{t}$ -decay.

The main parameter for the fit algorithm is the number of input jets. It must be discussed what the optimal value is to build the kinematic event solutions. In general, the probability to get the correct solution is higher the more jets are taken into account. As the number of combinations per event is determined by the faculty of the number of accepted jets the computing time per event increases extraordinarily such that the number of jets must be limited to a reasonable value. Physical arguments can be taken into consideration to determine a convenient number. As a parton from the $t\bar{t}$ -decay has on average a higher energy than a parton from a radiation process the four correct jets are found more probably at the beginning of an energy- or momentum-ordered list.

Fig. 8.5

In order to test the hypothesis that already the four leading jets contain the hadronic $t\bar{t}$ -decay information it is compared to the combinations including five leading jets. It can be recognized from Figure 8.5 that the peaks of the reconstructed masses of the W-boson and the top-quark fit better to the known masses of the particles and are sharper for the solutions with five jets than with four jets. That indicates that in some cases the leading four jets do not contain the entire kinematic information. But nevertheless, stable results on the reconstructed masses of the W-boson and the top-quark are obtained.

Fig. 8.5

Before deciding which number is actually chosen a criterion to assess the quality of the event solution is introduced. Therefore, not only the ability to reliably reconstruct the masses but also the selection efficiencies of signal and $t\bar{t}$ + Jets events is compared before the number of input jets is fixed.

It is expected that even if selecting the most probable four jet solution the mass distributions show tails mainly towards high reconstructed masses (see Figure 8.5). They arise when the most probable solution does not represent the true combinations either due to a wrong jet assignment or when the leading four jets do not contain the entire information on the hadronic part of the $t\bar{t}$ -decay. Therefore, two additional

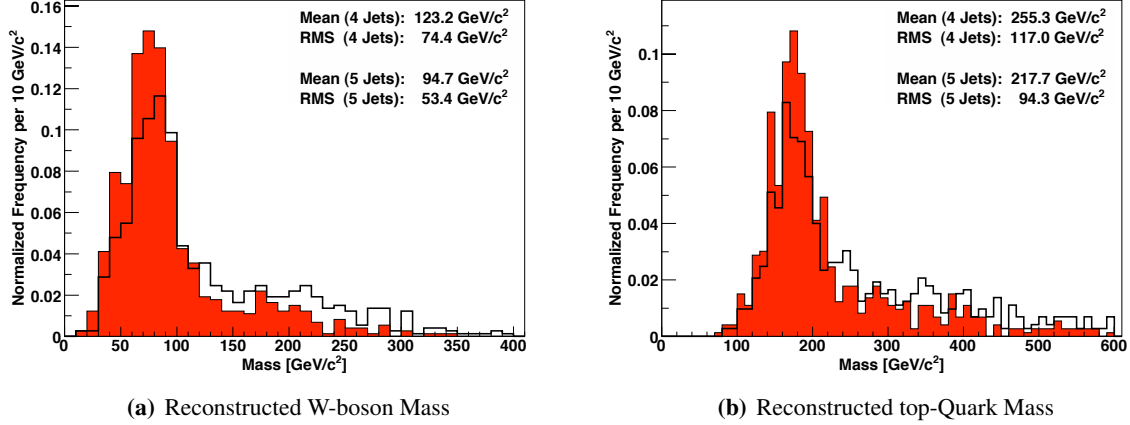


Figure 8.5: Normalized distributions of the reconstructed masses of the W-boson (a) and the top-quark (b) for the hadronic decay side (electron and muon streams merged). The filled histogram indicates the distributions for five jet solutions, and the open histogram gives the result for four jet solutions. The mean and the RMS are not extracted from a fit to the mass peaks but calculated for the entire distribution to show that the tail towards higher mass values is receding.

cuts on the mass windows of the reconstructed W-boson and the top-quark are introduced.

$$\begin{aligned} 50 \text{ GeV}/c^2 &< M_W^{reco} < 110 \text{ GeV}/c^2 \\ 150 \text{ GeV}/c^2 &< m_t^{reco} < 200 \text{ GeV}/c^2 \end{aligned} \quad (8.2)$$

The importance of the quality criteria can be understood when comparing the selection efficiencies for signal and $t\bar{t}$ + Jets events in Table 8.1. In general, it can be observed that the more additional Tab. 8.1

	Electron Event Stream		Muon Event Stream	
	4 Jet Solutions	5 Jet Solutions	4 Jet Solutions	5 Jet Solutions
Signal Events (e)	28%	36%	-	-
Signal Events (μ)	-	-	31%	39%
$t\bar{t}$ + 0 Jets	65%	63%	60%	68%
$t\bar{t}$ + 1 Jets	32%	39%	30%	36%
$t\bar{t}$ + 2 Jets	23%	34%	17%	30%
$t\bar{t}$ + 3 Jets	14%	37%	21%	28%
$t\bar{t}$ + $\geq$ 4 Jets	8%	18%	10%	20%

Table 8.1: Selection efficiencies of the event solution for the signal and $t\bar{t}$ + Jets events. The event solutions take either the four or five leading jets as input.

radiative hard jets occur in an event, the lower the selection efficiency becomes after applying these cuts, because the probability that one of the additional jets enters the leading jets rises if the total number of reconstructed jets is larger. This is independent of the number of input jets for the kinematic solutions, and leads to reasonable suppressions of at least the high multiplicity jet bins ($t\bar{t}$ + $\geq$ 2 Jets) compared to the signal events. Thus, the five jet solutions are still favoured. But in contrast, the selection efficiency increases when five instead of four jet solutions are chosen, because the probability to pick the correct combination of jets is larger for the first case. The increase of the signal efficiency ($\approx$ 25 – 30%) is deteriorated by the larger increase especially for the high multiplicity jet bins ($t\bar{t}$ + $\geq$ 2 Jets, $\approx$ 100%). Therefore, the four jet solutions are taken for the further analysis.

Fig. 8.5

This does not have to be the final solution because tighter quality cuts can be discussed for the five jet solutions as the mass peaks of the W-boson and the top-quark are reconstructed more sharply. This can potentially compensate the increase of the selection efficiencies due to the higher number of input jets. From Figure 8.5 it becomes evident that the mass distribution of W-boson mass is more altered than the top-mass histogram when changing between four and five jet solutions. The RMS of the W-boson mass distribution decreases by more than 30%, while the RMS of the top-quark mass distributions decreases by only about 20%. Thus, for the five jet solutions the upper and lower cut are separately modified by $10 \text{ GeV}/c^2$ towards the mass of the W-boson while the interval for the top-quark mass is kept constant. The effect of the increase of the signal efficiency is again observed, but also the even larger increase for the high multiplicity jet bins of the $t\bar{t}$ + Jets events when the five jet solutions are compared to the four jet solutions. The numerical results are similar to the above quoted. Following the same arguments as in the last paragraph this choice of the event reconstruction is abandoned as well.

In summary, it is concluded to select the four jet solutions for the kinematic event reconstruction respecting the above cited quality cuts.

8.5 The Correlations between Photons and Reconstructed Top-Quarks

In this section it is studied if there are any directional correlations between the photons and the reconstructed top- and antitop-quark by calculating the ΔR -distance from their reconstructed momentum directions. The results for top photons and fake photons are illustrated in Figure 8.6. As there is a causal

Fig. 8.6

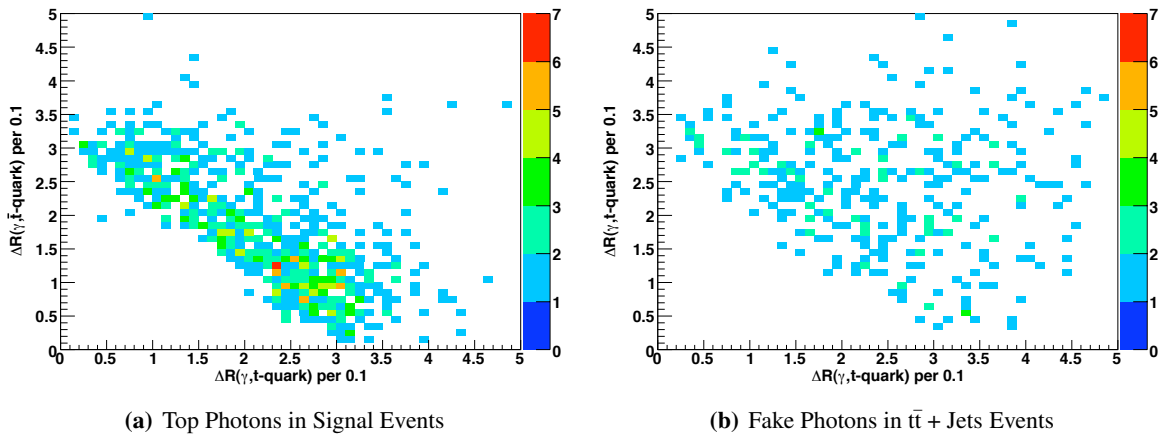


Figure 8.6: (a) Correlation between the reconstructed top photons and the reconstructed directions of the top- and antitop-quarks for the signal electron and muon events. (b) The same distributions is depicted for fake photons from the $t\bar{t}$ + Jets samples.

connection between the top photons and the reconstructed $t\bar{t}$ -system a ΔR -correlation for these photons can be observed by the increased counting rates along a virtual straight line in the histograms of the top photons. If the photon is radiated off one of the top-quarks, the momentum conservation holds for this subsystem. Thus the existing structure can be explained by the constraints imposed by the three-body system. If the system is boosted to the Lorentz-invariant frame where the $t\bar{t}$ -system has no longitudinal momentum component, i.e. $p_z^{t\bar{t}} = 0$, the limit separating the lower left corner from the distributions can be understood. It arises when the photon is lying in the same plane as both top-quarks. This gives the minimal value for the sum of both values of ΔR , such that the lower left corner is not populated. All other radiation processes with a longitudinal contribution lead to larger values of ΔR and extend the distribution to the upper right corner. The recoil of the photon radiation is accepted by the top-quark

which is constrained by the photon energy. Therefore, the density structure of the distributions for the top photons is created when the photon energy spectrum is folded into the distribution.

On the other hand the fake photons are not correlated to the top-quarks and the ΔR -distance relations are uniformly distributed. The fake photons can accidentally lie in the same plane as the $t\bar{t}$ -system such that again the minimal sum of both values of the correlation exclude the lower left corner of the histogram. But the energy spectrum is not folded into the distribution because the three particles are not linked by momentum conservation. Additionally, the admixture of fake photons being rather top photons (see Figure 7.19 and Figure 7.20) is too low in order to generate a notable correlation.

Fig. 7.19

This correlation gives a hint how an improved separation between top and fake photons can be achieved. But due to the low statistics no additional cut on this measure has been implemented. In principle the top photon distribution should extend to the upper right corner if the recoil and therefore the ΔR -values to the top-quarks become larger. But the number of very high-energetic photons is limited. Hence, before cutting on that measure the number of signal events should be increased to get a more complete picture of that distribution. At last, it must be investigated if a simple cut-based separation is actually advisable. When cutting out the upper right corner the high-energetic photons are removed. The implications on the photon spectra must be carefully studied.

Fig. 7.20

8.6 The Results of the Identification of $t\bar{t}\gamma$ -Events

The building of the event kinematic solutions and the application of quality criteria to the results represent the last step of the analysis chain to identify $t\bar{t}\gamma$ -signal events and remove background events. The

Table 8.2 contains the final numbers on the events that remain after the entire signal event identification. The results on the detailed sub-samples can again be seen in appendix A.

App. A

	Electron Stream				Muon Stream			
	Events	SP	NSP	FP	Events	SP	NSP	FP
Signal Electron	11.5	10.4	0.4	1.0	0	0	0	0
Signal Muon	0	0	0	0	11.8	11.0	0.0	0.9
$t\bar{t}$ + Jets	515.7	48.9	236.8	241.7	358.8	54.2	182.9	121.8

Table 8.2: Number of events at the end of the event selection. The numbers are scaled to $\mathcal{L} = 10 \text{ fb}^{-1}$.

The mean number of identified photons per event is about 1.03 (electron event stream) and 1.01 (muon event stream) for the signal events. Equal numbers, 1.02 and 1.00, respectively, hold approximately for the $t\bar{t}$ + Jets sample. For all non- $t\bar{t}$ samples the remaining event numbers is zero, and hence the figures are omitted from the table.

In summary, the key results of the analysis can be pointed out.

- The separation of $t\bar{t}$ events from the relevant physical background succeeds very efficiently. Subject to the discussion in chapter 9 the non- $t\bar{t}$ events could be significantly reduced, while at the same time $t\bar{t}$ -events remain. This signature based method obtains comparable results to other semileptonic studies [69]. The quality of the chosen strategy can be underlined when considering that the inclusive $t\bar{t}$ + Jets sample has been reduced to a nearly exclusive data-set which contains 85.5% events of the semileptonic (e- or μ -decay channel) type while the rest is dominantly of the dileptonic or semileptonic (τ -channel) type. Chp. 9
Ref. [69]
- As far as the photon identification is concerned the majority of the reconstructed and identified photons are real photons. They are either radiated off a charged particle which is preferential the

top-quark in case of the signal data-sets or they are decay products of π^0 -mesons. That latter is the main source of fake photons which has been proven for the $t\bar{t}$ + Jets events as well. These events have been generated without the explicit requirement for a hard photon so that they provide an independent cross-check for the reliability of the results on the photon identification. In addition, it is worthwhile to repeat that the fake photons contain a small admixture of photons radiated off top quarks. All these statements hold for both types of event streams.

- The separation of the events into an electron and muon stream gives the most important difference concerning the numerical results of that analysis. For the muon stream the results of that signature based analysis are slightly more promising than for the electron stream because of the possible physical ambiguity of the photon and the electron for the latter events. Neglecting a possible influence of the non- $t\bar{t}$ samples, the signal-to-background ratio is 1/45 for the electron stream and 1/30 for the muon stream.

Though the event rates are still dominated by the photons from π^0 -meson decays in $t\bar{t}$ + Jets events a first investigation is done in chapter 10 to create and compare different photon spectra. But before, the results on the non- $t\bar{t}$ data-sets are statistically reviewed to estimate their contribution to the photon spectra. The preliminary result that zero non- $t\bar{t}$ events remain after the event selection might be assigned to the fact that the number of generated events is far below the equivalent of $\mathcal{L} = 10 \text{ fb}^{-1}$ (integrated luminosity) and thus the phase space of these events is only roughly covered.

Chapter 9

The Statistical Results of the $t\bar{t}\gamma$ -Analysis

As it has been alluded in the end of chapter 8.6 the results for the background samples giving zero events after the event selection are still debatable. It is not possible to generate an appropriate number of events¹ due to the limited available computing power so that the entire phase space is well modeled. Therefore, in this chapter the contribution of the non- $t\bar{t}$ events are estimated based on the available data to get a sensible statistical treatment of these data-sets. Later, when real experimental data are available the contribution of these high-statistics background samples can be derived from these data using a similar method. Chp. 8.6

9.1 The Method of Decomposing the Total Selection Efficiency

The *Method of Decomposing the Total Selection Efficiency* is based on the fact that the total selection efficiency of a data-set is a composition of the selection efficiencies of the single filter steps included in the whole analysis chain. Ideally, the single filters are independent from each other. Then, the total selection efficiency factorizes into the product of the single selection efficiencies. This implies that the total selection efficiency of a data-set does not depend on the order in which the filters are applied. This is an important feature for the following examinations, and therefore the present analysis is divided into four independent filter steps.

1. Event Preselection
2. Lepton Identification
3. Photon Identification, Resolution of Particle Correlations, and Kinematic Event Solution (called *three-filter-stage* in the following)
4. Object Selection (b-tagging etc.).

The third item comprises three single filter stages. But as the output of the kinematic event solution strongly depends on the cleaning procedure of the particle correlations and the latter is again dependent on the results of the photon identification (see chapters 7.3 and 7.4), they should not be considered as single filter steps. Hence, the total selection efficiency for the present analysis is calculated as Chp. 7.3

Chp. 7.4

$$\varepsilon_{total} = \underbrace{\varepsilon_{Preselection}}_{\varepsilon_1} \cdot \underbrace{\varepsilon_{LeptonID}}_{\varepsilon_2} \cdot \underbrace{\varepsilon_{Three-Filter-Stage}}_{\varepsilon_3} \cdot \underbrace{\varepsilon_{ObjectSelection}}_{\varepsilon_4} \quad (9.1)$$

¹ A convenient sample of QCD-multijet events would contain $10^5 - 10^6$ times more events than the about five million events that have been analyzed for the present analysis.

It is now an inherent property of the analysis chain that the number of events is continuously reduced by the subsequent filter steps. This can end in a scenario that at a certain point in the analysis chain the number of events is reduced so far that only a few event remain as input for the next stage. Then, it can happen that this step cuts out all remaining events. Therefore, the total number of events and the total selection efficiency becomes zero. This has happened in the present analysis for the non- $t\bar{t}$ background samples.

In order to estimate the number of background events that would have passed the present analysis, the independence of the order of the filter stages is exploited. This ensures that a certain filter stage which has previously output zero events now gets a sufficiently large number of input events if it is run at a different position in the analysis chain. Hence, a single selection efficiency of this stage can be calculated differing from zero. Then, this result is inserted into equation 9.1 so that finally total selection efficiencies for all non- $t\bar{t}$ data-sets can be estimated.

In detail, the following actions are taken into account to apply this method.

1. The selection efficiencies for the event preselection ε_1 and lepton identification ε_2 can be adapted without modifications.
2. The selection efficiency of the three-filter-stage ε_3 is derived without the inclusion of the filter on the object selection in the analysis chain.
3. The selection efficiency of the object selection stage ε_4 is derived when this filter is applied immediately after the lepton identification.

Tab. 9.1 In Table 9.1 the final results on the selection efficiencies and the number of remaining events after the $t\bar{t}\gamma$ -analysis are summarized.

	Event Selection Efficiency					Events
	ε_1	ε_2	ε_3	ε_4	ε_{total}	
Signal (e)	$5.85 \cdot 10^{-1}$	$6.24 \cdot 10^{-1}$	$9.77 \cdot 10^{-2}$	$2.97 \cdot 10^{-1}$	$1.06 \cdot 10^{-2}$	11.5
Signal (μ)	$4.21 \cdot 10^{-1}$	$6.53 \cdot 10^{-1}$	$1.22 \cdot 10^{-1}$	$3.24 \cdot 10^{-1}$	$1.09 \cdot 10^{-2}$	11.8
$t\bar{t}$ + Jets (e)	$2.54 \cdot 10^{-1}$	$2.06 \cdot 10^{-1}$	$6.45 \cdot 10^{-3}$	$2.72 \cdot 10^{-1}$	$9.18 \cdot 10^{-5}$	515.7
$t\bar{t}$ + Jets (μ)	$2.54 \cdot 10^{-1}$	$1.31 \cdot 10^{-1}$	$6.04 \cdot 10^{-3}$	$3.17 \cdot 10^{-1}$	$6.37 \cdot 10^{-5}$	358.8
W + Jets (e)	$1.31 \cdot 10^{-3}$	$4.43 \cdot 10^{-1}$	$5.50 \cdot 10^{-3}$	$9.92 \cdot 10^{-3}$	$3.16 \cdot 10^{-8}$	13.1
W + Jets (μ)	$1.31 \cdot 10^{-3}$	$1.36 \cdot 10^{-1}$	$3.71 \cdot 10^{-3}$	$2.04 \cdot 10^{-2}$	$1.35 \cdot 10^{-8}$	5.6
Z + Jets	$6.67 \cdot 10^{-3}$	$4.74 \cdot 10^{-1}$	$3.13 \cdot 10^{-3}$	$6.34 \cdot 10^{-2}$	$6.27 \cdot 10^{-8}$	2.1
WW	$7.29 \cdot 10^{-3}$	$2.90 \cdot 10^{-1}$	$3.96 \cdot 10^{-3}$	$2.27 \cdot 10^{-2}$	$1.90 \cdot 10^{-7}$	0.2
WZ	$9.14 \cdot 10^{-3}$	$2.57 \cdot 10^{-1}$	$6.55 \cdot 10^{-3}$	$2.18 \cdot 10^{-2}$	$3.36 \cdot 10^{-7}$	0.2
ZZ	$1.53 \cdot 10^{-2}$	$2.63 \cdot 10^{-1}$	$1.07 \cdot 10^{-2}$	$3.57 \cdot 10^{-2}$	$1.54 \cdot 10^{-6}$	0.2
$W\gamma$	$3.34 \cdot 10^{-1}$	$2.74 \cdot 10^{-1}$	$6.77 \cdot 10^{-3}$	$2.42 \cdot 10^{-2}$	$1.50 \cdot 10^{-5}$	0.7
γ + Jets	$2.19 \cdot 10^{-6}$	$2.73 \cdot 10^{-2}$	$2.91 \cdot 10^{-3}$	$1.77 \cdot 10^{-2}$	$3.07 \cdot 10^{-12}$	< 0.1
QCD-Multijet	$4.45 \cdot 10^{-7}$	$2.85 \cdot 10^{-3}$	$1.84 \cdot 10^{-5}$	$5.09 \cdot 10^{-4}$	$2.37 \cdot 10^{-15}$	1.4

e and μ combined

Table 9.1: Single and total event selection efficiencies of the data-samples and number of events in $\mathcal{L} = 10 \text{ fb}^{-1}$.

At first, it can be seen for the signal data-sets that the present analysis does essentially not distinguish between the electron and muon event streams. Even if the single selection efficiencies for the different

stages differ, the total selection efficiency is the almost the same having applied all cuts and constraints. This statement holds almost perfectly for the signal data-sets, but it is valid up to a factor of less than two for the $t\bar{t}$ + Jets being very similar to the signal events.

This observation gives rise to another approximation for the non- $t\bar{t}$ data-sets. The above described decomposition does not succeed without splitting the three-filter-stage when the electron and muon stream are considered separately because the number of events in the muon stream is in general too low. Therefore, only the sum of background events for both streams is estimated and then divided into a contribution to the electron and the muon stream by assigning the half of the sum to each stream. This is acceptable for all non- $t\bar{t}$ but the W + Jets data-sets because even if the background events are not equally distributed to the two event streams the miss-assignment would be only of the order of less than one event.

The situation in case of W + Jets events is somewhat more complex because this is the non- $t\bar{t}$ data-set which is greater by approximately one order of magnitude compared to the sum of all other background events and therefore contributes to the highest extent. But here, the number of events in both event streams is large enough. The remaining events can be estimated separately for the electron and muon stream according to the above described procedure.

Then, it must be noted that for the QCD-multijet events the three-filter-stage must be split because the number of available events is too low to obtain a non-vanishing value for the total efficiency, even if the electron and muon event streams are merged. As explained above, the three different stages are not independent, but it can be understood that the following numerical result does not underestimate the contribution of the QCD-multijet events. In this particular case, the ambiguities between photons and jets are not resolved. As the particle correlation filter normally deletes high-energetic jet candidates which are more probably photons, the number of jets can become less than four and thus the event is dropped (see chapter 7.4.3). If this filter is not included, more events remain because the number of jets is preferentially larger. Therefore, the selection efficiency for the kinematic event reconstruction step increases as well, and the selection efficiency of the three-filter-stage rises. Hence, the background estimation of the QCD-multijet data-set represents an upper limit.

Chp. 7.4.3

In Figure 9.1 the number of remaining event for an integrated luminosity $\mathcal{L} = 10 \text{ fb}^{-1}$ (last column of Table 9.1) is transformed into observable cross-sections normalized to the signal cross-section of 218 fb^{-1} . This figure must be compared to Figure 5.6 showing the same distribution before the event selection.

Fig. 9.1

Tab. 9.1

Fig. 5.6

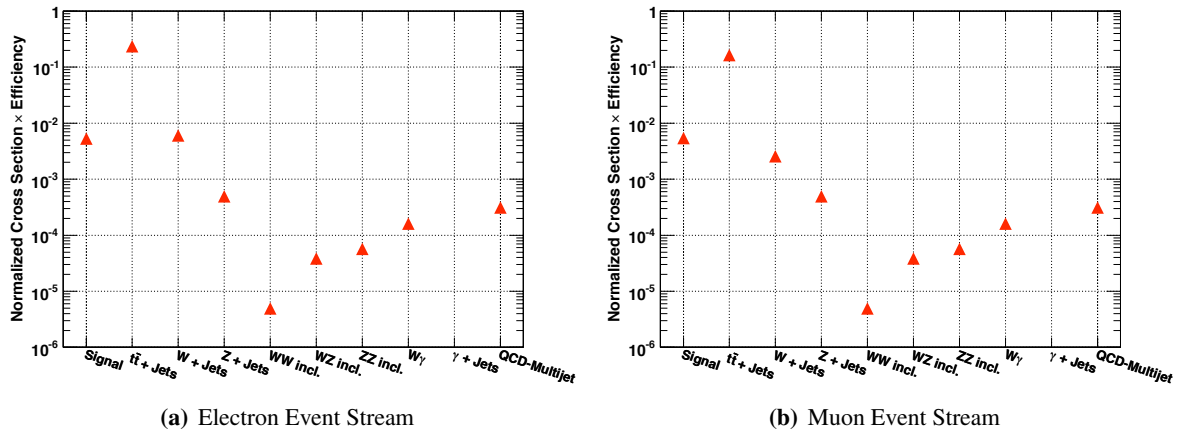


Figure 9.1: Comparison of the cross-sections normalized to the signal cross-section of 218 fb^{-1} and scaled to the selection efficiency for the electron event stream (a) and the muon event stream (b).

It is illustrated that after the event selection the efficiency to select signal events is about 1% (sum of electron and muon event streams). As far as the non- $t\bar{t}$ is concerned these contributions are suppressed by at least one order of magnitude for both event streams, except for the $W + \text{Jets}$ events. For this data-set the number of remaining events is efficiently reduced in general such that the number becomes almost equal to the number of signal events. But when splitting them into the electron and muon event streams the number is slightly higher by about 15% for the first, while these background events are lower by approximately a factor of two for the last.

In contrast, the relative excess of $t\bar{t} + \text{Jets}$ events could be reduced only from approximately 2500 to about 40 such that the final data are dominated by this sample. The results are again somewhat more promising for the muon than for the electron stream. But it is evident that the number of $t\bar{t} + \text{Jets}$ events is that large for an integrated luminosity $\mathcal{L} = 10 \text{ fb}^{-1}$ that a mere statistical fluctuation on that data-set can suffice to mimic a signal on top of the $t\bar{t} + \text{Jets}$ background like it is described by the signal events. Hence, the present $t\bar{t}\gamma$ -analysis does not seem to be sensitive to a contribution of events where a photon is radiated off a top-quark.

This can again be confirmed when the new signal-to-background ratios (S/B) are calculated, now including the estimated number of background events. The results are indicated in Table 9.2. The ratio are

Tab. 9.2

	Electron Event Stream	Muon Event Stream
S/B without non- $t\bar{t}$ Events	1/45	1/30
S/B with non- $t\bar{t}$ Events	1/46	1/31
$t\bar{t} + \text{Jets}$ Events/non- $t\bar{t}$ Events	33/1	45/1

Table 9.2: Signal-to-background ratio for the $t\bar{t}\gamma$ -analysis without and with the contribution of the non- $t\bar{t}$ events. Furthermore the ratio of $t\bar{t} + \text{Jets}$ and non- $t\bar{t}$ events is calculated.

only marginally reduced further by the non- $t\bar{t}$ data-sets. The muon stream gives still a higher ratio than the electron stream.

But another ratio is of a particular interest for the following studies of the photon spectra that can be derived from the selected events. More clearly for the muon than for the electron stream, the S/B ratio between the $t\bar{t} + \text{Jets}$ and the non- $t\bar{t}$ events is much larger than one. Hence, the $t\bar{t}\gamma$ -event selection has identified a $t\bar{t}$ -data-set which is dominating over its background. This gives rise to distinguish two scenarios in the $t\bar{t}\gamma$ -analysis.

- *Signal A (SigA)*

The first scenario is a rather unphysical one. It neglects the contribution of the $t\bar{t} + \text{Jets}$ data in order to study the non- $t\bar{t}$ background with respect to the signal sample. In a real experiment this scenario is not sensitive to any modest modification of the coupling between the photon and top-quark because the $t\bar{t} + \text{Jets}$ events must definitively be taken into account. It can become a relevant scenario if the fake photons from π^0 -meson decays are efficiently identified compared to the top photons so that the $t\bar{t} + \text{Jets}$ background can be further reduced.

- *Signal B (SigB)*

The $t\bar{t} + \text{Jets}$ events contain a 10%- and 15%-fraction (electron stream and muon stream, respectively) of information about photons that are radiated off top-quarks (see Table 8.2). Hence, the $t\bar{t} + \text{Jets}$ events can replace the signal events as the source of events containing top photons. This scenario does not neglect any data-set and can eventually be used to assess the sensitivity of this analysis with respect to modifications of the top photon component. As the non- $t\bar{t}$ background is suppressed with respect to that data-set an impact of the top photon component on the photon spectra can become visible.

Tab. 8.2

At last, it is worthwhile to emphasize that the analysis is strongly relying on the dedicated signal data-sets. All identification criteria and cut values especially for the photon identification can only be derived from a data-set which is enriched with high-energetic photons that are radiated off top-quarks. If taking the $t\bar{t} + \text{Jets}$ events as signal from the very beginning no adequate results can be achieved because the absolute number of top photons is too low.

9.2 The Application of the Estimated Limits to the Photon Spectra

Before the remaining events are integrated into photon spectra it must be explained how the estimated event number of the non- $t\bar{t}$ data-sets are transformed into a spectrum when actually no events are left after the event selection. In the present case the photon spectra extracted right after the lepton identification are taken and uniformly scaled by the event selection efficiencies of the following analysis steps such that the limits on the efficiencies of Table 9.1 are achieved in total. This ensures that the correct number of events per data-set contributes but still giving sufficient events to arrive at steady shapes of the single spectra. By that procedure each event is included and only scaled with the probability to be selected. This procedure must be applied for the QCD-multijet, $\gamma + \text{Jets}$, and di-boson samples because only here sufficient events are left. Tab. 9.1

This is not a perfect solution because the spectra contain photon candidates rather being jets leading to high-energetic tails of the distributions. These do not exist anymore after the photon identification. For the $W + \text{Jets}$ and $Z + \text{Jets}$ data-sets it can be afforded to scale the events after the entire photon identification has taken place. This still gives relatively steady shapes of the photon distributions and more precise results for the photon spectra.

But not only the event selection efficiency but also the photon identification rate must be respected. At the stage of scaling the background events, the mean number of photons per event can be larger than 1.03 (electron event stream) and 1.02 (muon event stream) which are the numbers that have been reached after the final steps of the photon identification (see Table 8.2). Therefore the numbers of photons per event are scaled as well such that 1.03 and 1.02 photons per event are included in the spectra on average. Tab. 8.2

Chapter 10

The Photon Energy Spectra

THIS chapter presents the photon spectra which are generated from the events that have passed the event selection. According to the two cases that have been introduced in chapter 9 the Standard Model spectra neglecting either the $t\bar{t}$ + Jets events (Signal A) or the dedicated signal sample (Signal B) are shown. The second subject of this chapter is devoted to an investigation how the coupling of the photon to top-quarks modifies the energy spectra if it differs from the Standard Model prediction. For that reason, a Monte Carlo generator has been adopted which allows for the modification of the coupling between the photon and the top-quark [72]. The separation of the Standard Model top-quark with an electromagnetic charge $q_{top} = 2/3$ from a scenario where the top-quark has $q_{top} = -4/3$ is investigated. Chp. 9
Ref. [72]

10.1 The Standard Model Photon Energy Spectra

In Figure 10.1 the Standard Model photon energy spectra are shown in the upper row for the Signal A and in the lower row for the Signal B case. In accordance with Figure 9.1 and Table 9.2 it can be observed that the $t\bar{t}$ -contribution strongly influences the spectra. In case of the Signal A the $t\bar{t}\gamma$ -fraction is at the same order as the non- $t\bar{t}$ background at $\mathcal{L} = 10 \text{ fb}^{-1}$, at typical value for the integrated luminosity obtained in one year of running in the low luminosity phase of the LHC ($\mathcal{L} = \mathcal{O}(10^{33} \text{ cm}^{-2} \text{ s}^{-1})$). The Signal B spectra are scaled to $\mathcal{L} = 1 \text{ fb}^{-1}$. It is evident that a proper separation between the $t\bar{t}$ -contribution and the non- $t\bar{t}$ samples seems to be feasible already with this lower integrated luminosity. For both cases the muon events stream provides better results with respect to the separation from background events than the electron event stream. Fig. 10.1
Fig. 9.1
Tab. 9.2

Then, the high-energetic tails of the di-boson and QCD-multijet/ γ + Jets-events especially for the electron stream must be discussed. Even if the absolute number of events is not very large it must be repeated that these tails stem from the uniform scaling procedure of these events (see chapter 9.2). The contents of these bins is more related to jets faking photons rather than consisting of photon candidates that would pass the photon identification. This part could not be cut out because in that case the number of remaining events would be too low for these data-sets in order to achieve a relatively steady shape of the spectra. This explanation is proven as well by the samples which have undergone the photon identification ($t\bar{t}$ -samples, W + Jets and Z + Jets samples) and do not show these pronounced tails. Chp. 9.2

In general, neither Signal A nor Signal B can distinguish further between $t\bar{t}$ -events and non- $t\bar{t}$ events, because the shapes of the photon energy distributions for the various background samples resemble these of the $t\bar{t}$ -events. The majority of the photons of all events tend to be in the low-energetic part of the histograms. Throughout the entire rest of this chapter it will be studied what can be deduced from the spectra about the electromagnetic charge of the top-quark due to the admixture of photons that are radiated off top-quarks in the $t\bar{t}$ -contents of the photon spectra.

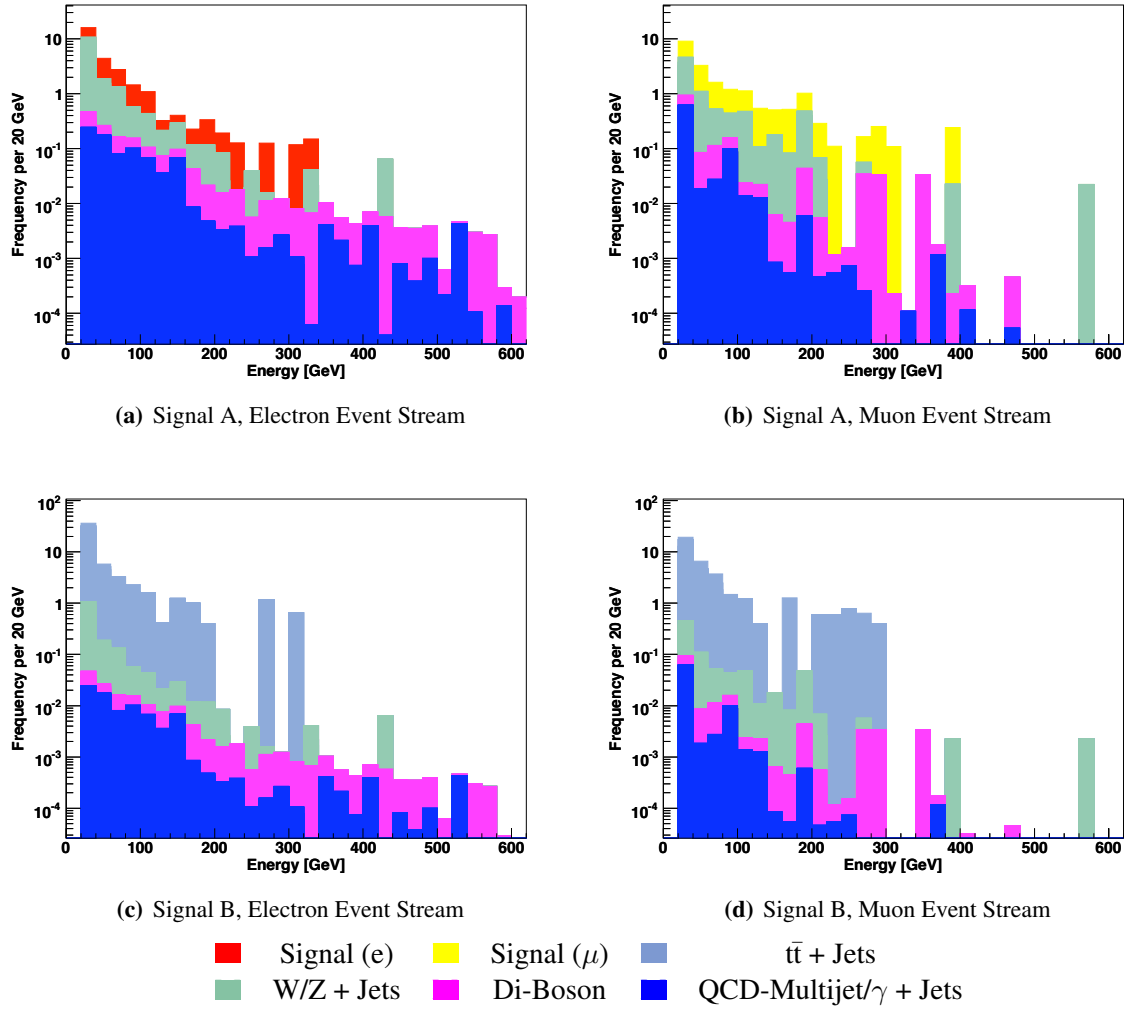


Figure 10.1: Standard Model photon energy spectra separated for the electron event stream (left column) and muon event stream (right column). The upper row histograms (a) and (b) are scaled to an integrated luminosity $\mathcal{L} = 10 \text{ fb}^{-1}$ and reveal the spectra neglecting the $t\bar{t}$ + Jets contributions (Signal A). The lower row histograms (c) and (d) are scaled to an integrated luminosity $\mathcal{L} = 1 \text{ fb}^{-1}$ and consider the $t\bar{t}$ + Jets events as signal sample (Signal B).

10.2 The Non Standard Model Photon Energy Spectra

Having elaborated the Standard Model spectra it is now studied how the photon energy spectra would be altered if the coupling between the photon and the top-quark differs from the Standard Model prediction. The future aim of such a study could be to provide a number of spectra that could be compared to proton-proton collision data in order to measure the coupling constants when comparing the latter to the template spectra histograms.

Fig. 10.2

Mathematically the coupling of the photon to the top-quark (Figure 10.2) is hidden in the transition matrix element for $t(p) \rightarrow t(q) + \gamma(k)$. While the modification does not affect the particle currents the vertex factor is changed by a Lorentz-invariant parametrization

$$\Gamma_\mu(p, q, k) = -ie\{\gamma_\mu [F_{1V}(k^2) + \gamma_5 F_{1A}(k^2)] + \frac{\sigma_{\mu\nu}}{2m_t} (p + q)^\nu [iF_{2V}(k^2) + \gamma_5 F_{2A}(k^2)]\}, \quad (10.1)$$

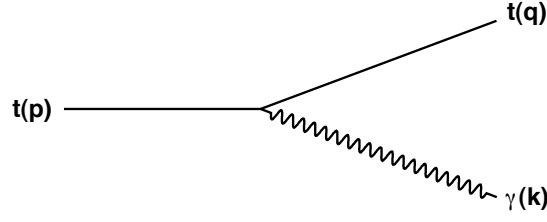


Figure 10.2: Kinematic structure of the $t\bar{t}\gamma$ -vertex.

where e is the positive electromagnetic charge, m_t the mass of the top-quark and p, k , and q the Lorentz-vectors of the particles at the vertex. The functions $F_{1,V}$ and $F_{1,A}$ are the vector and axial vector form factors, and $F_{2,V}$ and $F_{2,A}$ represent the magnetic and electric dipole form factors g_t and d_t , respectively.

At Born-level the Standard Model couplings are predicted as follows.

$$\begin{aligned}
 F_{1,V}(0) &= -\frac{2}{3} = Q_t \\
 F_{1,A}(0) &= 0 \\
 F_{2,V}(0) &= 0 = Q_t \frac{g_t - 2}{2} \big|_{g_t=2} \\
 F_{2,A}(0) &= 0 = \frac{2m_t}{e} d_t \big|_{d_t=0}
 \end{aligned} \tag{10.2}$$

As these form factors are included in the transition matrix element possible modifications immediately affect the value of the matrix element, and thus the phase space distributions of these photons that couple to top-quarks. Therefore, it is possible to investigate alternative sets of coupling parameters to compare them to the Standard Model configuration.

For the current investigation the form factor $F_{1,V}(0)$ that is related to the quantum number of the electromagnetic charge is changed such that the top-quark gets a charge quantum number $q_{top} = -4/3$ (compare the discussion in chapter 5.1). The next section describes how this modifications is introduced into the photon energy spectra that are so far derived for a Standard Model top-quark with $q_{top} = 2/3$. Chp. 5.1

10.2.1 Signal Rescaling by the Baur-Generator

The vertex structure of equation 10.1 is implemented in a matrix element generator programmed by U. Baur and collaborators [72]. It allows for the calculation of a $t\bar{t}\gamma$ -final state for proton-proton- collisions at $\sqrt{s} = 14$ TeV. The output of this process generation is a text file which mainly contains the kinematic relations of the particles and the value of the squared matrix element $|M|^2$ as shown in the following for an arbitrary event. Ref. [72]

	E	p_x	p_y	p_z
<i>Parton 1</i>	1833.4	0.0	0.0	1833.4
<i>Parton 2</i>	5627.9	0.0	0.0	-5627.9
<i>top - quark 1</i>	5579.9	221.7	-94.7	-5572.3
<i>top - quark 2</i>	419.8	-187.2	102.2	316.4
<i>Photon</i>	1461.6	-25.5	-7.5	1461.4

$$|M|^2 \quad 1.078 \cdot 10^{-7} \tag{10.3}$$

This text file is read and the information about the photon as well as the event weight given by $|M|^2$ are stored in a ROOT-file in order to be processed further. The strategy to introduce the information about

Fig. 10.3(a)

the varied coupling parameters is now to run this generator at first with the Standard Model configuration and then probing an alternative scenario. In Figure 10.3(a) the normalized photon energy spectra for the two values of the electromagnetic charge of the top-quark $q_{top} = 2/3$ and $q_{top} = -4/3$ are depicted. Each entry is weighted by the value of the matrix element so that the difference between the Standard Model and the alternative model becomes obvious.

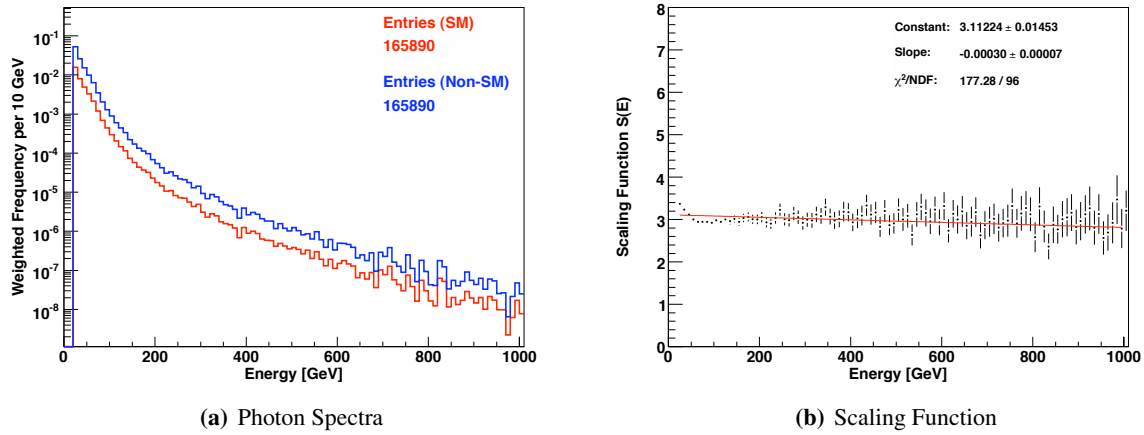


Figure 10.3: (a) Photon spectrum for a top-quark with $q_{top} = 2/3$ (lower curve) and with $q_{top} = -4/3$ (upper curve). Each entry is weighted by the value of the matrix element $|M|^2$ and scaled by the total number of events (non-stacked histograms). (b) The ratio of the two curves gives the scaling function for top photons which is fitted by a linear function.

Fig. 10.3(b)

Fig. 10.1

The ratio of these two spectra is determined and a scaling function $S(E)$, depending on the photon energy E , is calculated (see Figure 10.3(b)). Finally, a linear fit to this function is used to scale the fraction of top photons in the reconstructed energy spectra (Figure 10.1). This provides the new energy spectra for the alternative scenario with $q_{top} = -4/3$.

In detail, the scaling function is deduced according to the following ratio.

$$S(E) = \frac{N_{SM}}{N} \frac{\frac{dN(E)}{dE}}{\frac{dN_{SM}(E)}{dE}} \quad (10.4)$$

It describes how the energy spectrum $\frac{dN(E)}{dE}$ of the alternative scenario deviates from the Standard Model reference spectrum $\frac{dN_{SM}(E)}{dE}$. Each differential spectrum contains the matrix element-weighted distributions of the photon energy. The first quotient in the equation is meant to make the histograms independent from the number of events. The scaling function itself is determined by dividing the entries of the n^{th} bin (B^n) of the alternative spectrum by the n^{th} bin (B_{SM}^n) of the Standard Model spectrum resulting in a discrete scaling function. The statistical error on that ratio per bin is deduced by error propagation

$$\Delta S^n(E) = S^n(E) \cdot \sqrt{\left(\frac{\Delta B^n}{B^n}\right)^2 + \left(\frac{\Delta B_{SM}^n}{B_{SM}^n}\right)^2 + \left(\frac{\Delta N}{N}\right)^2 + \left(\frac{\Delta N_{SM}}{N_{SM}}\right)^2}. \quad (10.5)$$

Ref. [73]

According to [73] the error on the matrix element due to higher order corrections, i.e. the error on the weight per photon energy value, can be several 10%. The implications of this error on the scaling function must be included in an investigation of the systematic error.

The result of the linear fit to the ratio of the discrete scaling values read as follows.

$$S(E) = 3.11224 - 0.01453 \cdot E \text{ [GeV]} \approx 3.1 \quad (10.6)$$

This scaling function is nearly energy-independent which is a physical property of the charge coupling parameter that has been changed. The first expectation that the function would be equal to four is not met because the effects of initial state radiation and interference effects are included in the results as well. Hence, the scaling function is not only the squared ratio of the two charge values. In principle, the shape of the spectrum can be altered as well when varying a different coupling parameter. In that case a convenient fitting function must be chosen in order to correctly transform the top photon fraction of the spectra to the new coupling scenario.

10.3 The Limits on the Electromagnetic Charge of the Top-Quark

The Figure 10.4 indicates the spectra for the Signal A and Signal B cases separated by the electron and the muon event streams. The histograms are not stacked so that one can see how the spectra for $q_{top} = -4/3$ rise above the distributions for $q_{top} = 2/3$. For Signal A and Signal B the non- $t\bar{t}$ background is constant for each lepton stream and has been subtracted from the spectra such that only the difference between the alternative scenario and the Standard Model one is visible in the histograms. Fig. 10.4

For the Signal A the photon rate is evidently increased throughout the entire spectrum because it contains about 90% top photons. Even if this results is very interesting because the two scenarios differ significantly from each other it must be repeated that the Signal A does not contain the $t\bar{t}$ + Jets events which must be included and would smear the entire effect.

For the Signal B the modifications affect only approximately 10% and 15% (electron and muon event stream, respectively) and could be explained principally by statistical fluctuations as well. But it must be noted that in that case all fluctuations must lead to higher counting rates. But as these signal scenario is closer to reality a detailed statistical analysis will show that the distinction between a Standard Model top-quark and a top-quark with an electric charge of $-4/3$ is possible if only the integrated luminosity becomes large enough. For that reason limits on the top photon contribution to the photon spectra are calculated at a confidence level (CL) of 95% using the Feldman and Cousins method [74]. Ref. [74]

The signal and the background events shall be described by Poisson distributions. Thus, the probability to observe n events given that s signal events and b background events are expected reads as follows.

$$P(n|s, b) = \frac{(s + b)^n}{n!} e^{-(s+b)}. \quad (10.7)$$

In order to get the total number of observed events the signal and background events are summed up. The background events contain the non- $t\bar{t}$ events according to Table 9.1, whereas the signal events must be split into two parts according to the ratio of top and non-top/fake photons (see Table 8.2). The first part comprise the fraction of $t\bar{t}$ -events containing a top photon, and the second one contributes to the spectrum with a non-signal or fake photon which is therefore added to the background events. The distinction is necessary because only the signal fraction is affected by the scaling by means of equation 10.6 when the charge hypothesis is changed, while the non-signal and fake parts remain constant. Tab. 9.1
Tab. 8.2

Following the idea of the Feldman and Cousins method the confidence belts are constructed assuming that only background events are expected, i.e. only non- $t\bar{t}$ -events and $t\bar{t}$ -events with non-signal or fake photons. In Figure 10.5 for $q_{top} = 2/3$ and Figure 10.6 for $q_{top} = -4/3$ the confidence belts are depicted. Each figure contains the confidence belt for Signal A and Signal B, further distinguished into the electron and the muon event streams. The total number of observed events that is predicted by the present analysis is drawn as a dashed vertical line, and the intersections of this line with the confidence belt provide the limits on the number of signal events at 95% CL. Fig. 10.5
Fig. 10.6

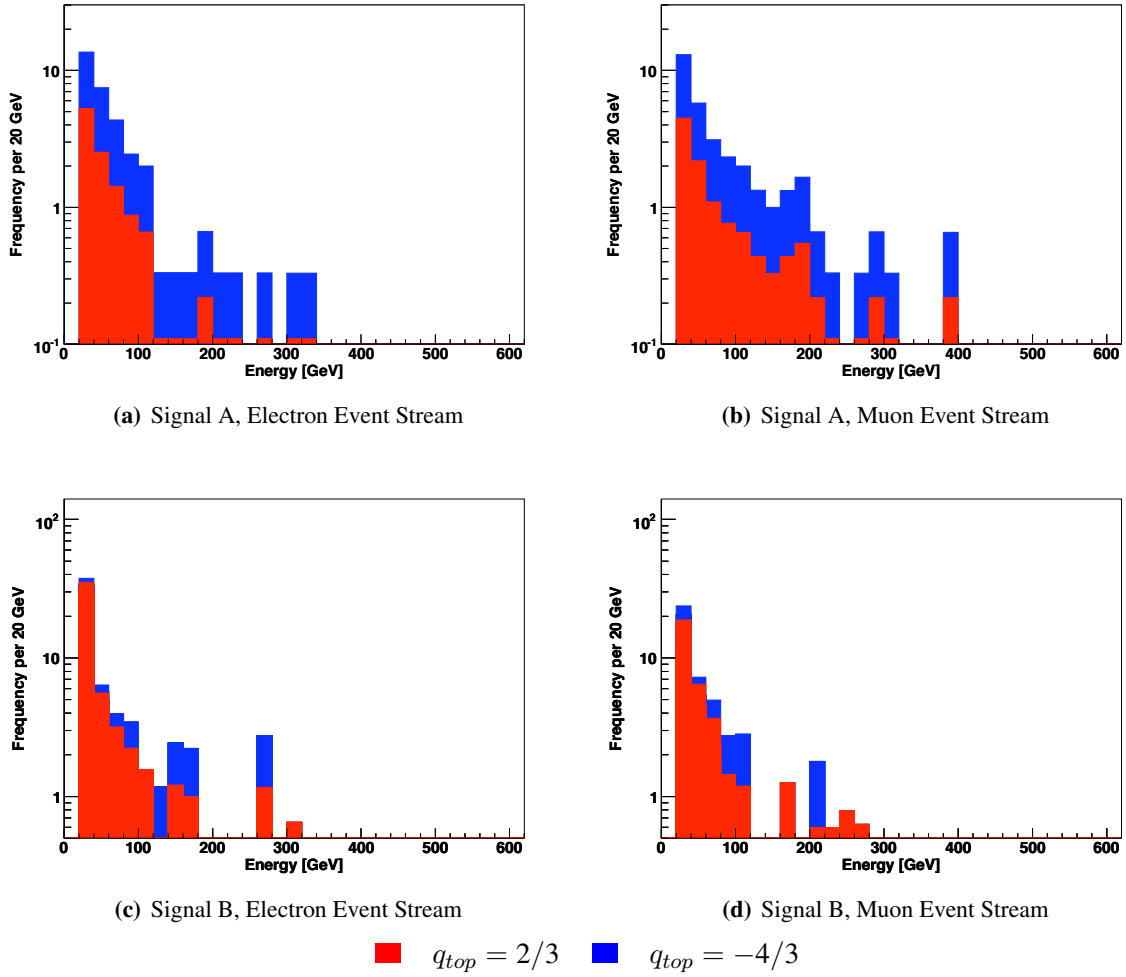


Figure 10.4: Comparison between the photon spectra for $q_{top} = 2/3$ and $q_{top} = -4/3$ (non-stacked histograms). The non- $\bar{t}t$ background has been subtracted because it is the same for both charge scenarios in Signal A ($\mathcal{L} = 10 \text{ fb}^{-1}$) and Signal B ($\mathcal{L} = 1 \text{ fb}^{-1}$), respectively.

In addition, the expectation values of the upper and lower limits are determined. For that reason a Poisson distribution is folded into the confidence belt construction which accepts the number of observed events (all signal and background events) as input for the mean value. These values provide information about the mean limits if an equal analysis is performed on more than one data-set, and can serve as a numerical result to compare the sensitivity of a different analysis on the same data-set.

Tab. 10.1

The numerical results on these statistical evaluations are summarized in Table 10.1. They will be discussed later when the number of events is converted into limits on the electromagnetic charge of the top-quark.

For that reason the dependence of the scaling function on the charge parameter must be determined. Two additional spectra have been created, assuming that the top-quark has either $q_{top} = -1/3$ or $q_{top} = -1$. Though there is no popular physical model for these scenarios, the scaling functions for the top photons are again approximately energy-independent. In summary, the scaling parameters reveal a quadratic dependence on q_{top} as it is shown in Figure 10.7.

Fig. 10.7

The data points are fitted to a quadratic function $S(q)$. It is used to convert the limits on the signal event

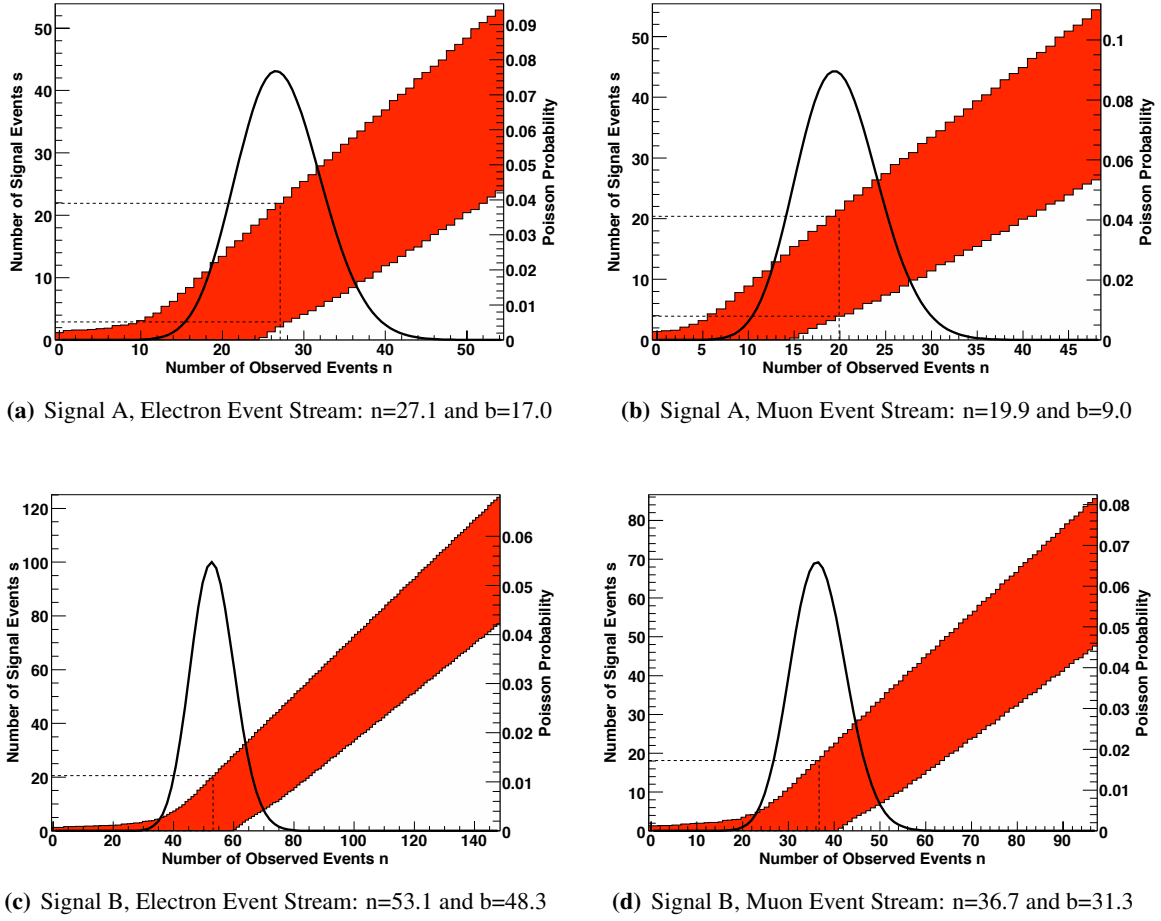


Figure 10.5: Confidence belts (95% CL) for the number of signal events for $q_{top} = 2/3$. The background events are the composition of non- $t\bar{t}$ events and $t\bar{t}$ -events with non-signal and fake photons. The dashed line indicates the number of expected events of the present analysis for the Signal A and B scenarios (upper and lower row, respectively). The intersection of this line with the belt give the interval for the number of signal events.

		Signal A				Signal B			
q_{top}		Lower Limit		Upper Limit		Lower Limit		Upper Limit	
+2/3	e-Stream	2.9	< 3.2 >	21.9	< 22.5 >	0.0	< 0.9 >	20.6	< 22.0 >
	μ -Stream	3.9	< 4.6 >	20.4	< 22.5 >	0.0	< 1.1 >	18.1	< 20.1 >
-4/3	e-Stream	19.4	< 19.8 >	46.4	< 47.8 >	3.2	< 4.8 >	32.1	< 33.2 >
	μ -Stream	22.4	< 23.1 >	47.4	< 49.5 >	5.9	< 7.1 >	32.1	< 33.2 >

Table 10.1: Limits on the signal event number for Signal A and Signal B. The numbers in brackets represent the expectation values on the upper and lower limits when folding a Poissonian probability function with the curves on the upper and lower limits.

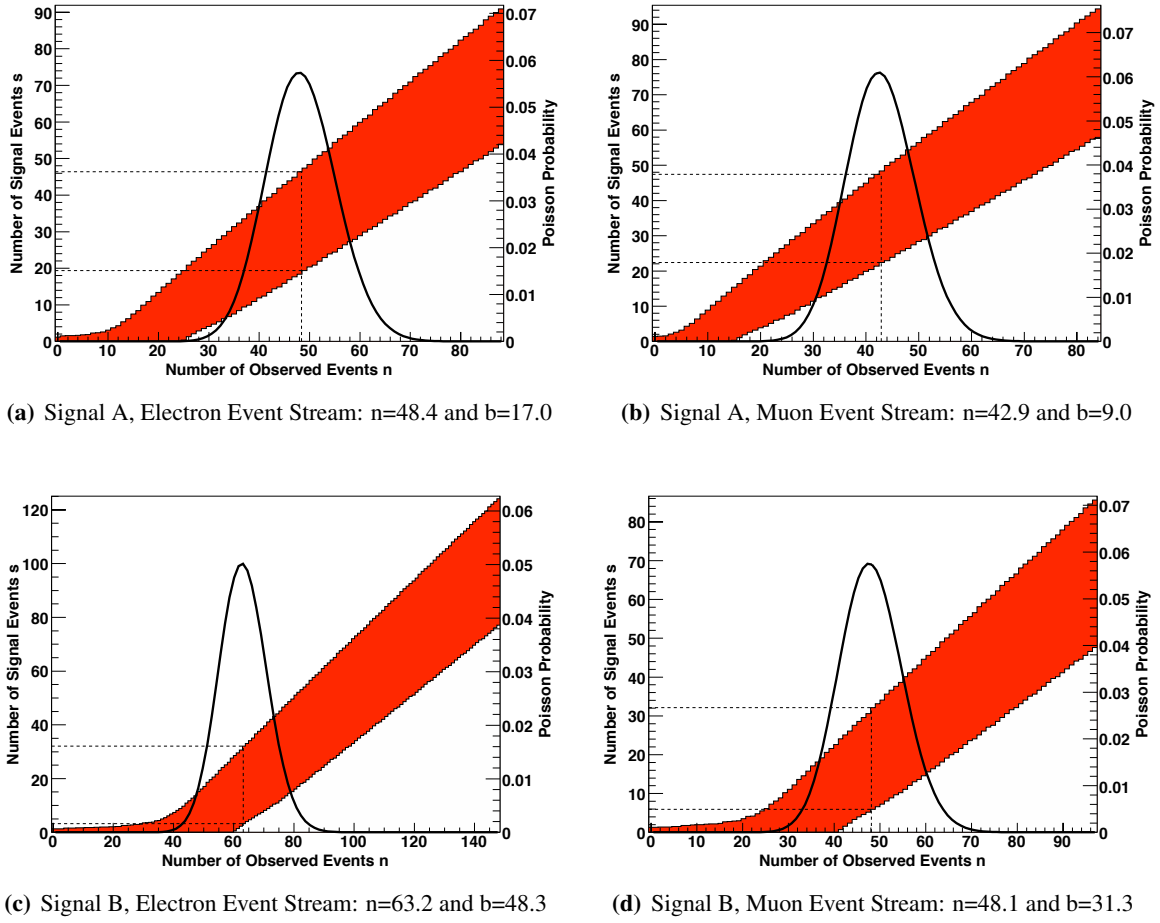


Figure 10.6: Confidence belts (95% CL) for the number of signal events for $q_{top} = -4/3$. The background events are the composition of non- $t\bar{t}$ events and $t\bar{t}$ -events with non-signal and fake photons. The dashed line indicates the number of expected events of the present analysis for the Signal A and B scenarios (upper and lower row, respectively). The intersection of this line with the belt give the interval for the number of signal events.

numbers to limits on the electromagnetic charge of the top-quark.

$$\begin{aligned}
 N(q) &= N_0 \cdot S(q) \\
 &= N_0 \cdot (aq^2 + bq + c) \\
 \Leftrightarrow q &= -\frac{b}{2a} + \sqrt{\frac{b^2}{4a^2} + \frac{1}{a} \left(\frac{N(q)}{N_0} - c \right)}
 \end{aligned} \tag{10.8}$$

The parameters a, b and c are taken from the quadratic fitting function and the number of signal events $N(q)$ at a certain value for the charge q are compared to the expected number of signal events N_0 for the Standard Model spectrum as predicted by the present analysis. The negative solution is omitted because this analysis is only sensitive to the absolute value of the electromagnetic charge of the top-quark.

Using this equation the limits on the electric charge of the top-quark can be determined. The result is graphically shown in Figure 10.8 for Signal A at $\mathcal{L} = 10 \text{ fb}^{-1}$ and for Signal B at $\mathcal{L} = 1 \text{ fb}^{-1}$ separated for the electron and muon stream at 95% CL. The lower bar for each scenario represents the limits for a top-quark with $|q_{top}| = 2/3$, and the upper bar is valid for $|q_{top}| = 4/3$.

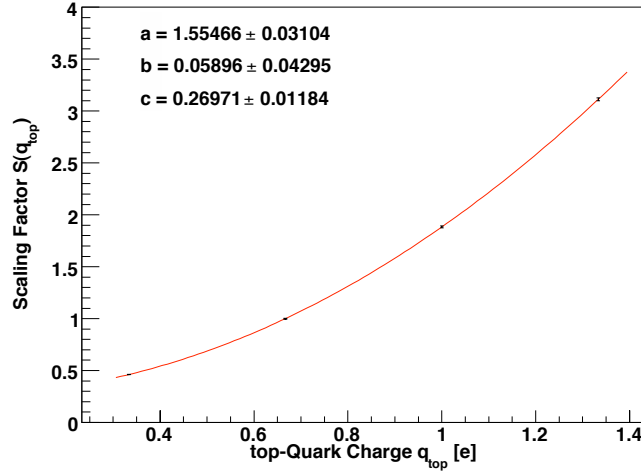


Figure 10.7: Functional dependence of the scaling factor versus the electromagnetic charge quantum number. The data points at $1/3$ and 1 do not correspond to a physical model but show that the scaling factor depends quadratically on q_{top} .

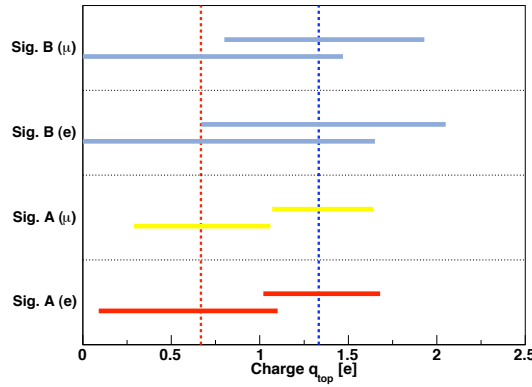


Figure 10.8: Limits at 95% CL on the electromagnetic charge quantum number of the top-quark for the two signal scenarios separated by the electron and muon stream. The limits on Signal A are determined for $\mathcal{L} = 10 \text{ fb}^{-1}$, for Signal B $\mathcal{L} = 1 \text{ fb}^{-1}$ is considered.

Limits on q_{top} – Neglecting the $t\bar{t}$ + Jets Contribution (Signal A)

If the $t\bar{t}$ -Jets background can be removed efficiently, i.e. the separation of photons that are radiated off top-quarks from photons that are decay products of π^0 -mesons is improved, then the distinction of the two charge scenarios according to Signal A can already succeed at $\mathcal{L} = 10 \text{ fb}^{-1}$ for the muon event stream. Both coupling scenarios are clearly separated at 95% CL.

Regarding the electron stream it is obvious that a distinction at $\mathcal{L} = 10 \text{ fb}^{-1}$ is not yet possible because there is a small overlap between both scenarios. This is expected because not only the number of non-top and fake photons but also the number of non- $t\bar{t}$ background events is higher than for the muon event stream. This is also the reason why the intervals are more widely spread in case of the electron event stream for both values of q_{top} .

Limits on q_{top} – Considering the $t\bar{t}$ + Jets Events as Signal (Signal B)

For the so far more realistic Signal B the distinction between the two scenarios is not possible at $\mathcal{L} = 1 \text{ fb}^{-1}$, neither for the electron nor for the muon event stream. But the latter gives tighter intervals. Even if it is not favoured by the indirect measurements [51] an exclusion of $|q_{top}| = 2/3$ seems be possible with only $\mathcal{L} = 1 \text{ fb}^{-1}$ in the muon event stream at 95% CL.

In order to address the significance of the present $t\bar{t}\gamma$ -analysis the limits are extrapolated to a higher integrated luminosity to determine when the Signal B becomes sensitive to a separation between the two scenarios. As the expected numbers of observed events are sufficiently high for $\mathcal{L} \geq 5 \text{ fb}^{-1}$ the Poisson distributions of the Feldman and Cousins method are approximated by Gaussian functions. This is an acceptable approximation because it can already be observed in Figure 10.5 and Figure 10.6 that the Poisson distributions begin to converge against Gaussian curves. Thus, the problem of extraordinary computing time of the Feldman and Cousins method with large event numbers is overcome.

In Figure 10.9 the limits on q_{top} as a function of the integrated luminosity for the Signal B are revealed. They are calculated not only for 95% CL but also for 99% CL. The results of the muon event stream are again better than of the electron event stream. Not only the value of the integrated luminosity at which the two bands do not overlap anymore is lower, but also the limits on both potential values of q_{top} are tighter at any value of the integrated luminosity. If only a separation at 95% CL is envisaged then up to only $\mathcal{L} \approx 10 \text{ fb}^{-1}$ is sufficient to distinguish between the two charge values in both lepton event streams. The stricter limit at 99% CL needs $\mathcal{L} \approx 20 \text{ fb}^{-1}$ for the electron event stream and $\mathcal{L} \approx 10 \text{ fb}^{-1}$ for the muon event stream in order to favour either $|q_{top}| = 2/3$ or $|q_{top}| = 4/3$.

In summary, it can be concluded that the distinction between the two values of the electromagnetic charge of the top-quark is possible. It becomes obvious that the present $t\bar{t}\gamma$ -analysis is not meant for a “first day physics program“ but several 10 fb^{-1} of LHC data should provide enough events to reliably perform a first direct measurement of q_{top} .

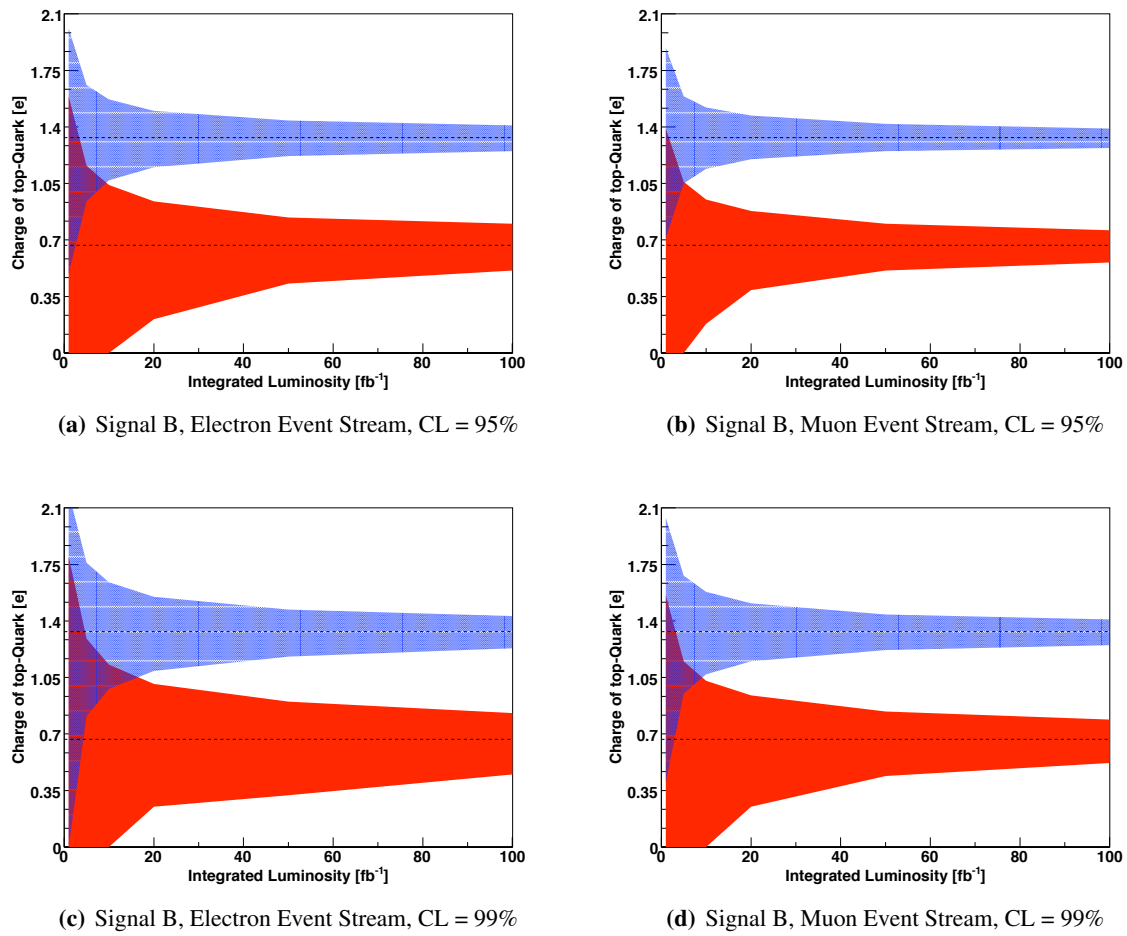


Figure 10.9: Limits on the electromagnetic charge of the top-quark for $q_{top} = 2/3$ and $|q_{top}| = 4/3$ versus integrated luminosity for the Signal B in both events streams at 95% CL (upper row) and 99% CL (lower row).

Chapter 11

Summary and Concluding Remarks

THE present thesis has been done in the framework of the CMS experiment at the Large Hadron Collider, the next generation proton-proton collider at the European laboratory for particle physics, CERN. It is concerned with two topics. The first one deals with an aspect on the development and construction of the silicon microstrip tracking detector at the CMS experiment, while the second one covers a Monte Carlo analysis in the realm of physics with top-quark pairs.

Study on Silicon Microstrip Module Defects

The mass production of the silicon microstrip detector modules for the CMS silicon tracker has achieved a yield of 97.2%. These module which have not passed the qualification procedure have been sorted out and examined by special Module Debug Centres. In these centres an independent examination of the problem has taken place, and the module has been repaired if possible. This centralized procedure has allowed to install a global book-keeping and to detect systematic insufficiencies in the mass production chain.

For the modules of the tracker endcaps the Module Debug Centre has been located at Aachen. Here, 225 modules of the tracker endcaps have been investigated in total. In this thesis the debug procedures are explained and the results of the investigations are discussed. As many modules have shown similar problems, seven fundamental defect classes have been introduced. The signatures of each class are described, and the methods and tests to determine if a module belongs to a certain class are illustrated. Finally, the number of modules which are associated to a certain defect class are counted.

The defect class comprising modules suffering from inappropriate handling has appeared to be the most frequent one, affecting about one third of all examined modules. This high defect rate has been observed and published to the tracker collaboration, so that an extension of that problem has been able to be prevented. The overall statistic has revealed that no systematic source of defects has been present in the mass production chain. In the end, 126 modules have been recovered by either directly repairing or initiating a repair action. From the remaining ones, 41 modules have been able to be de-assembled. The quality of their sensors has been still good enough to be reintegrated into a new module.

A centralized Module Debug Centre has proven to be very efficient to examine defects during a mass production, without blocking the production and to recover modules contemporarily. The efficiency would be even increased if the Debug Centre is equipped with more facilities and technical expertise to repair all kinds of module defects. This should be discussed for future projects if a Debug Centre must be installed again. A possible solution would be not to nominate an existing laboratory as Debug Centre but to found a new one and to gather experts for each type of defect at one central place.

Study on Photon Identification in Semileptonic $t\bar{t}$ -Events

The semileptonic (electron and muon) $t\bar{t}$ -decay channel has been used to select events containing a photon that is radiated off a top-quark. The signal data-set consists of a Pythia $t\bar{t}$ -sample that is enriched with high-energetic photons that are radiated off top-quarks by final state radiation ($\sigma \approx 218$ fb). All selection criteria are elaborated and optimized with respect to that sample. For the physical background samples $t\bar{t} + \text{Jets}$, $W + \text{Jets}$, $Z + \text{Jets}$, di-boson, $\gamma + \text{Jets}$ and QCD-multijet events are considered.

After an event preselection to impose the basic signature of the $t\bar{t}$ -decay chain on the events, the lepton identification is performed. A selection purity of 92.7% for the electron identification and 97.3% for the muon identification is achieved. Afterwards, all data-sets are split into an electron and a muon event stream if either an isolated electron or muon has been identified. The photon identification is optimized such that photons are retained that are radiated off top-quarks. The selection purity for these top photons is 90.9% for the electron event stream and 92.7% for the muon event stream. On average, 1.023 (electron event stream) and 1.016 (muon event stream) photons per event are identified. It is shown that the photons that remain and are not radiated off top-quarks are irreducible background photons because they are first and foremost decay photons of π^0 -mesons. Hence, the present photon identification is actually selecting real photons rather than fake ones.

In addition to the lepton and photon identification the kinematic reconstruction of the $t\bar{t}$ -fraction of the event signature is implemented. Cuts on the single objects themselves as well as on the combinatorial event solutions are applied to reconstruct the semileptonic $t\bar{t}$ -decay. The final selection efficiency for signal events is 1.1%, while signal-to-background ratios of 1/46 for the electron event stream and 1/31 for the muon event stream can be achieved. These ratios are relatively low because an inclusive $t\bar{t} + \text{Jets}$ data-set is included in the background samples. Its semileptonic fraction provides the same signature as the signal events. In contrast, the non- $t\bar{t}$ background only slightly degrades the ratio. It is estimated to be only of the order of the signal events for the $W + \text{Jets}$ events, and even lower up to several orders of magnitude for the other non- $t\bar{t}$ data-sets.

In the end, the photon energy spectra are discussed for two cases. The first one (Signal A) excludes the $t\bar{t} + \text{Jets}$ events and the signal events are only compared to the non- $t\bar{t}$ background samples. For the present analysis this is a very daring choice unless a high-efficient method is incorporated into the analysis that would actually reject the photons from π^0 -decays and so the $t\bar{t} + \text{Jets}$ events. The second case (Signal B) replaces the signal events with the $t\bar{t} + \text{Jets}$ sample. It is possible to compare the latter to the non- $t\bar{t}$ events because they contain an admixture of photons that are radiated off top-quarks which is about 10% (15%) for the electron (muon) event stream. This is the more realistic scenario because it rather resembles the future real event data than the Signal A for the present analysis.

The long-term goal of the present analysis is to compare the proton-proton collision data to the photon spectra to measure the coupling of the photon to the top-quark. In order to assess a potential to distinguish between different coupling scenarios two different photon spectra are compared assuming that the top-quark has either an electromagnetic charge quantum number of $2/3$ or $-4/3$. A dedicated matrix element generator has been used to create the different spectra. Their ratio is transformed into an nearly energy-independent scaling function. This is applied to the reconstructed photons that are radiated off top-quarks individually for the Signal A and Signal B case. It can be shown that it is possible to distinguish between these two charge scenarios already at an integrated luminosity of 10 fb^{-1} for the Signal A case especially for the cleaner muon event stream at a confidence level of 95%.

For the Signal B the situation is different because of the lower admixture of photons radiated off top-quarks. The mere separation at a confidence level of 95% succeeds already with an integrated luminosity of $5 - 10 \text{ fb}^{-1}$. In addition, for an integrated luminosity of approximately 20 fb^{-1} (10 fb^{-1}) in the electron (muon) event stream both values of the electromagnetic charge do not overlap anymore and can be distinguished already at a confidence level of 99%.

The extrapolation of the present analysis to the point when it becomes sensitive to either of the two charge values must be carefully discussed. At first, effects of a systematic study must clearly be included. Furthermore, the analyzed data-sets do not contain the effect of multiple proton-proton interactions. But this becomes even more important if the integrated luminosity is approaching the values at which the analysis starts to distinguish between a top-quark with an electric charge of $2/3$ or $-4/3$. This aspect must be taken into account when a precise estimate on the desired integrated luminosity should be given. Furthermore, it must be tested to which extent the extrapolation from the Standard Model spectra to the alternative spectra via the matrix element generator and the scaling procedure is valid. An implementation of the generator output itself into a proper detector simulation and reconstruction would be advisable to get even more reliable results.

For the photon identification the correlations between the photon and the top-quarks are not exploited so far. This aspect can be checked as soon as a larger number of signal events becomes available. This allows for taking more sophisticated photon identification methods like neural networks into consideration as well.

Driven by the large number of photons from π^0 -mesons it is worthwhile to repeat the analysis not for the semileptonic but for the dileptonic $t\bar{t}$ -decay channel. In principal, the event signature should be more well-defined to at least partially compensate the lower event rate. The $t\bar{t}$ -decay reveals a cleaner signature and the rate of irreducible background photons should be decreased because the mean number of jets and so π^0 -mesons is expected to be lower.

Nevertheless, it can be demonstrated that the signature based method that is developed in the present thesis is able to reconstruct $t\bar{t}$ -events that have undergone a complete detector simulation and contain photons that are radiated off top-quarks. The first investigation of different physical models for the coupling of photons to top-quarks at the CMS experiment is successfully studied. It becomes evident that the analysis is not feasible for the beginning of the data-taking period of the LHC but it is not an only-hypothetical analysis as well. In the very first chapter it has been mentioned that one of the design goals of the LHC and the CMS detector is to measure properties of particles. The present analysis provides a new strategy to directly measure the electromagnetic properties of the top-quark.

Appendix A

Background Statistics

As mentioned in the main part of the thesis the statistical results are presented in a summarized way for each main data-set. After listing which data-sets are considered for the background events, detailed tables are written down to give detailed figures on the event selection for all data-sets.

A.1 The Data-Sets

For the sake of reproducibility the following list gives a full composition of background data-samples which have been analyzed from the Spring 07 data production. All information are read from the central Database-Server (DBS)¹.

- $t\bar{t}$ + Jets
/tt0j/CMSSW_1_3_5-Spring07-2033/GEN-SIM-DIGI-RECO
/tt1j/CMSSW_1_3_5-Spring07-2034/GEN-SIM-DIGI-RECO
/tt2j/CMSSW_1_3_5-Spring07-2035/GEN-SIM-DIGI-RECO
/tt3j/CMSSW_1_3_5-Spring07-2036/GEN-SIM-DIGI-RECO
/tt4j/CMSSW_1_3_5-Spring07-2169/GEN-SIM-DIGI-RECO

- W + Jets
/W0j/CMSSW_1_3_1-Spring07-1550/GEN-SIM-DIGI-RECO
/W1j/CMSSW_1_3_1-Spring07-1551/GEN-SIM-DIGI-RECO
/W2j/CMSSW_1_3_1-Spring07-1552/GEN-SIM-DIGI-RECO
/W3j/CMSSW_1_3_1-Spring07-1553/GEN-SIM-DIGI-RECO
/W4j/CMSSW_1_3_1-Spring07-1554/GEN-SIM-DIGI-RECO
/W5j/CMSSW_1_3_5-Spring07-2042/GEN-SIM-DIGI-RECO
/W6/CMSSW_1_3_5-Spring07-2093/GEN-SIM-DIGI-RECO

- Z + Jets
/Z0j/CMSSW_1_3_1-Spring07-1555/GEN-SIM-DIGI-RECO
/Z1j/CMSSW_1_3_1-Spring07-1556/GEN-SIM-DIGI-RECO
/Z2j/CMSSW_1_3_1-Spring07-1557/GEN-SIM-DIGI-RECO
/Z3j/CMSSW_1_3_1-Spring07-1558/GEN-SIM-DIGI-RECO

¹see webpage: <https://cmsweb.cern.ch/dbs.discovery/>

/Z4j/CMSSW_1_3_1-Spring07-1559/GEN-SIM-DIGI-RECO
 /Z5j/CMSSW_1_3_5-Spring07-2040/GEN-SIM-DIGI-RECO
 /Z6j/CMSSW_1_3_5-Spring07-2041/GEN-SIM-DIGI-RECO

- Di-Boson

/WW_incl/CMSSW_1_3_4-Spring07-1898/GEN-SIM-DIGI-RECO
 /WZ_incl/CMSSW_1_3_1-Spring07-1153/GEN-SIM-DIGI-RECO
 /ZZ_incl/CMSSW_1_3_1-Spring07-1152/GEN-SIM-DIGI-RECO
 /Wgamma/CMSSW_1_3_1-Spring07-1346/GEN-SIM-DIGI-RECO

- γ + Jets

/PhotonJetsBkg_pt_0_15/CMSSW_1_3_4-Spring07-1904/GEN-SIM-DIGI-RECO
 /PhotonJetsBkg_pt_15_20/CMSSW_1_3_4-Spring07-1899/GEN-SIM-DIGI-RECO
 /PhotonJetsBkg_pt_20_30/CMSSW_1_3_4-Spring07-1905/GEN-SIM-DIGI-RECO
 /PhotonJetsBkg_pt_30_50/CMSSW_1_3_4-Spring07-1946/GEN-SIM-DIGI-RECO
 /PhotonJetsBkg_pt_50_80/CMSSW_1_3_4-Spring07-1906/GEN-SIM-DIGI-RECO
 /PhotonJetsBkg_pt_80_120/CMSSW_1_3_4-Spring07-1907/GEN-SIM-DIGI-RECO
 /PhotonJetsBkg_pt_120_170/CMSSW_1_3_4-Spring07-1908/GEN-SIM-DIGI-RECO
 /PhotonJetsBkg_pt_170_300/CMSSW_1_3_4-Spring07-1909/GEN-SIM-DIGI-RECO
 /PhotonJetsBkg_pt_300_500/CMSSW_1_3_4-Spring07-1945/GEN-SIM-DIGI-RECO
 /PhotonJetsBkg_pt_500_7000/CMSSW_1_3_4-Spring07-2000/GEN-SIM-DIGI-RECO

- QCD-multijets

/QCD_pt_0_15/CMSSW_1_3_1-Spring07-1565/GEN-SIM-DIGI-RECO
 /QCD_pt_15_20/CMSSW_1_3_1-Spring07-1566/GEN-SIM-DIGI-RECO
 /QCD_pt_20_30/CMSSW_1_3_1-Spring07-1567/GEN-SIM-DIGI-RECO
 /QCD_pt_30_50/CMSSW_1_3_1-Spring07-1568/GEN-SIM-DIGI-RECO
 /QCD_pt_50_80/CMSSW_1_3_1-Spring07-1569/GEN-SIM-DIGI-RECO
 /QCD_pt_80_120/CMSSW_1_3_1-Spring07-1570/GEN-SIM-DIGI-RECO
 /QCD_pt_120_170/CMSSW_1_3_4-Spring07-1656/GEN-SIM-DIGI-RECO
 /QCD_pt_170_230/CMSSW_1_3_4-Spring07-1657/GEN-SIM-DIGI-RECO
 /QCD_pt_230_300/CMSSW_1_3_4-Spring07-1658/GEN-SIM-DIGI-RECO
 /QCD_pt_300_380/CMSSW_1_3_4-Spring07-1659/GEN-SIM-DIGI-RECO
 /QCD_pt_380_470/CMSSW_1_3_4-Spring07-1660/GEN-SIM-DIGI-RECO
 /QCD_pt_470_600/CMSSW_1_3_4-Spring07-1661/GEN-SIM-DIGI-RECO
 /QCD_pt_600_800/CMSSW_1_3_4-Spring07-1662/GEN-SIM-DIGI-RECO
 /QCD_pt_800_1000/CMSSW_1_3_4-Spring07-1663/GEN-SIM-DIGI-RECO
 /QCD_Pt_1000_1400/CMSSW_1_3_1-Spring07-1303/GEN-SIM-DIGI-RECO
 /QCD_Pt_1400_1800/CMSSW_1_3_1-Spring07-1301/GEN-SIM-DIGI-RECO
 /QCD_Pt_1800_2200/CMSSW_1_3_1-Spring07-1298/GEN-SIM-DIGI-RECO
 /QCD_Pt_2200_2600/CMSSW_1_3_1-Spring07-1296/GEN-SIM-DIGI-RECO
 /QCD_Pt_2600_3000/CMSSW_1_3_4-Spring07-2066-bis/GEN-SIM-DIGI-RECO
 /QCD_Pt_3000_3500/CMSSW_1_3_1-Spring07-1293/GEN-SIM-DIGI-RECO
 /QCD_Pt_3500_inf/CMSSW_1_3_1-Spring07-1290/GEN-SIM-DIGI-RECO

A.2 Tables on the Event Selection of the Background Data-Sets

From Table A.1 to Table A.6 the event selection results of the above listed data-sets are presented. At first they are grouped by a more general class of event type, and then divided into several sub-classes giving all detailed information about the number of events that have been selected after each of the six steps of the $t\bar{t}\gamma$ -analysis. None of the entries is scaled to a certain integrated luminosity, but giving the raw information about the absolute and unscaled event numbers. Tab. A.1
Tab. A.6

The following abbreviations are used to describe at which point in the entire chain of the analysis the numbers are extracted

- **Initial**
Number of events before any selection criterion or cut is applied.

- **Preselection**
Number of events after the event preselection when a rough event signature has been demanded (chapter 6). Chp. 6

- **Lepton ID**
Number of events after the identification of leptons has been taken place. From now onwards a data-set is split into an electron and a muon event stream (chapter 7.2). Chp. 7.2

- **Photon ID**
Number of events after applying the photon identification cuts on the sample (chapter 7.3). Chp. 7.3

- **Correlations**
Number of events when the sample has been cleaned for the correlations between photons, electrons and jets (chapter 7.4). Chp. 7.4

- **Selection**
Number of events after checking the data-set for tighter kinematic cuts on the single reconstructed objects also including b-tagging (first half of chapter 8). Chp. 8

- **Kinematics**
Number of events that have passed the more complete event reconstruction when composite objects have been found (second half of chapter 8). Chp. 8

Statistics on $t\bar{t}$ + Jets Events

	TT + N Jets				
	N = 0	N = 1	N = 2	N = 3	N = 4
Initial	492549	431218	170287	63003	97561
Preselection	67095	95887	52742	24812	51618
Lepton-ID	16427/4099	21396/10334	10557/7185	4572/3806	8885/10345
Photon-ID	716/119	1029/379	515/247	214/145	491/384
Correlations	219/81	498/311	309/209	147/129	336/334
Selection	54/25	153/102	77/53	35/43	95/114
Kinematics	35/15	49/31	18/9	5/9	8/11

Table A.1: Final selection statistics on $t\bar{t}$ + Jets events. After the lepton identification the event samples are split into an electron and a muon event stream. Therefore, there are two numbers in a cell from that step onwards, the first representing the electron and the second the muon stream.

Statistics on W + Jets Events

	W + N Jets						
	N = 0	N = 1	N = 2	N = 3	N = 4	N = 5	N = 6
Initial	195933	191966	180937	68199	72585	55209	19619
Preselection	3	27	213	1573	6001	9833	6861
Lepton-ID	1/0	17/1	113/12	900/73	2675/851	3504/2027	1824/1718
Photon-ID	0/-	4/0	14/0	37/7	123/26	148/52	75/39
Correlations	-	2/-	1/-	13/6	60/21	104/43	52/37
Selection	-	0/-	0/-	0/0	0/0	1/0	0/0
Kinematics	-	-/-	-/-	-/-	-/-	0/-	-/-

Table A.2: Final selection statistics on W + Jets events. After the lepton identification the event samples are split into an electron and a muon event stream. Therefore, there are two numbers in a cell from that step onwards, the first representing the electron and the second the muon stream.

Statistics on Z + Jets Events

	Z + N Jets						
	N = 0	N = 1	N = 2	N = 3	N = 4	N = 5	N = 6
Initial	43814	47869	46560	47913	21402	4779	4185
Preselection	20	58	1654	4800	4236	1487	2158
Lepton-ID	8/0	29/1	739/9	2229/108	1787/313	540/205	618/457
Photon-ID	1/-	8/0	95/1	221/16	169/16	56/9	42/9
Correlations	0/-	1/-	5/1	43/9	63/13	27/9	25/9
Selection	-	0/-	0/0	0/0	1/0	0/0	2/0
Kinematics	-	-	-	-	0/-	-	0/-

Table A.3: Final selection statistics on Z + Jets events. After the lepton identification the event samples are split into an electron and a muon event stream. Therefore, there are two numbers in a cell from that step onwards, the first representing the electron and the second the muon stream.

Statistics on Di-Boson Events

	Di-Boson			
	WW	WZ	ZZ	W γ
Initial	478864	195000	59478	48535
Preselection	3492	1782	1065	245
Lepton-ID	857/154	419/39	267/13	44/4
Photon-ID	38/7	30/2	27/2	24/2
Correlations	12/6	14/2	4/1	8/2
Selection	0/1	0/0	0/0	0/0
Kinematics	-/0	-	-	-

Table A.4: Final selection statistics on di-boson events. After the lepton identification the event samples are split into an electron and a muon event stream. Therefore, there are two numbers in a cell from that step onwards, the first representing the electron and the second the muon stream.

Statistics on γ + Jets and QCD-Multijet Events

The γ + Jets and QCD-multijet data-sets are divided into several p_T -bins representing the momentum transfer from the two-parton initial state to the two-parton final state ($2 \rightarrow 2$ -process). The highest-energetic object in such an event can only have this transverse momentum. The separation of the sample is indicated by the second row in the header of the tables. The physical unit of the intervals is GeV/c.

It must be noted that the first four sub-samples already vanish after the event preselection. As these data-sets have the largest cross-section of all γ + Jets and QCD-multijet sub-samples it is debatable if sufficient events have been analyzed in order cover the entire phase space for these events. But from the physical point of view the kinematic event reconstruction of the analysis demands for objects like a W-boson or a top-quark. This means, that the energy content of these types of events is in general too low to contribute to the events that remain after the event selection ($p_T \leq 50 \text{ GeV/c} \approx \frac{1}{2} M_W$). Therefore, the potential problem of too little events can be ignored.

	γ + Jets				
	0-15	15-20	20-30	30-50	50-80
Initial	49990	38384	71293	38488	47538
Preselection	0	0	0	0	13
Lepton-ID	-	-	-	-	0
Photon-ID	-	-	-	-	-
Correlations	-	-	-	-	-
Selection	-	-	-	-	-
Kinematics	-	-	-	-	-

	γ + Jets(continued)				
	80-120	120-170	170-300	300-500	500-7000
Initial	49880	44431	43970	46107	46899
Preselection	279	944	2189	4382	5792
Lepton-ID	9/0	25/0	76/4	143/9	186/25
Photon-ID	0/-	1/-	9/0	9/4	10/4
Correlations	-	0/-	7/-	4/4	6/4
Selection	-	-	0/-	0/0	1/1
Kinematics	-	-	-	-	0/0

Table A.5: Final selection statistics on γ + Jets events. After the lepton identification the event samples are split into an electron and a muon event stream. Therefore, there are two numbers in a cell from that step onwards, the first representing the electron and the second the muon stream.

QCD-Multijet						
	0-15	15-20	20-30	30-50	50-80	80-120
Initial	142991	172936	141549	188961	1388153	584134
Preselection	0	0	0	0	294	1737
Lepton-ID	-	-	-	-	1/0	5/2
Photon-ID	-	-	-	-	0/-	0/0
Correlations	-	-	-	-	-	-
Selection	-	-	-	-	-	-
Kinematics	-	-	-	-	-	-

QCD-Multijet (continued)						
	120-170	170-230	230-300	300-380	380-470	470-600
Initial	395110	406244	421259	393250	188817	156107
Preselection	5433	12866	21127	26499	15188	14378
Lepton-ID	9/1	22/2	17/2	28/3	18/2	20/4
Photon-ID	2/0	1/0	0/0	0/0	1/0	1/0
Correlations	0/-	0/-	-	-	1/-	1/-
Selection	-	-	-	-	0/-	0/-
Kinematics	-	-	-	-	-	-

QCD-Multijet (continued)						
	600-800	800-1000	1000-1400	1400-1800	1800-2200	2200-2600
Initial	186115	177392	25166	20000	15000	19432
Preselection	19390	20233	3040	2623	2114	2703
Lepton-ID	15/4	18/5	4/1	6/2	1/2	4/3
Photon-ID	2/0	1/0	0/0	0/0	0/0	0/0
Correlations	0/-	0/-	-	-	-	-
Selection	-	-	-	-	-	-
Kinematics	-	-	-	-	-	-

QCD-Multijet (continued)			
	2600-3000	3000-3500	3500-7000
Initial	31929	19000	19790
Preselection	4551	2717	2716
Lepton-ID	6/1	8/3	7/1
Photon-ID	0/0	1/0	0/0
Correlations	-	0/-	-
Selection	-	-	-
Kinematics	-	-	-

Table A.6: Final selection statistics on QCD-multijet events. After the lepton identification the event samples are split into an electron and a muon event stream. Therefore, there are two numbers in a cell from that step onwards, the first representing the electron and the second the muon stream.

Appendix B

Example of a CMSSW Configuration File

THIS chapter gives an overview how a configuration file is implemented in the CMSSW-project. The configuration files are based on a configuration syntax which is explained on the CMS TWiki-Pages¹. The configuration file presented below has been used to steer the lepton identification. It is a common example giving reference about the main structure of these files. There are commenting lines starting with “#” to explain what the blocks are meaning.

```
process LeptonID =
{

# CMSSW parameters for textual output (precisely defined in ''MessageLogger.cfi'')
# and to define the input files

    include "FWCore/MessageLogger/data/MessageLogger.cfi"

    untracked PSet options = { untracked bool wantSummary = true }

    source = PoolSource
    {
        untracked vstring fileNames =
        {
            "file:/Daten/@FILENAME@.root"
        }
        untracked int32 maxEvents = -1
    }

# Modules to apply tighter kinematical cuts on electrons and muons

    module newSelectedLayer1TopElectrons = TopElectronSelector
    {
        InputTag src = selectedLayer1TopElectrons
        string cut   = "pt > 30.0"
    }

    module newSelectedLayer1TopMuons = TopMuonSelector
    {
        InputTag src = selectedLayer1TopMuons
        string cut   = "pt > 30.0"
    }

# Main Lepton Identification Module: Definition of basic parameters,
#                                     particle input collections and cut values

    module LeptonIDFilterModule = LeptonID_Filter
    {
        untracked string datapath                = "/CMS_Analyse/Results"
        untracked string LeptonID_Filter_OutFile = "LeptonIDFilter_Results"
```

¹The CMS Offline Workbook: <https://twiki.cern.ch/twiki/bin/view/CMS/WorkBook>

```

untracked bool    tt_gamma_data          = false
untracked bool    Signal_Dataset         = @SIGNAL@

untracked vint32  cuts_to_apply = {1,1,1,0}    # cuts on CaloIso, TrackIso, HoE, HoP
InputTag MCData    = VtxSmeared
InputTag Electrons = newSelectedLayer1TopElectrons
InputTag Muons     = newSelectedLayer1TopMuons

untracked double max_dE_match = 0.1
untracked double max_dR_match = 0.05

untracked double electron_calor_iso_cut = 3.5
untracked double electron_track_iso_cut = 6.0
untracked double electron_hoe_cut      = 0.02
untracked double electron_eop_cut      = 0.75

untracked double muon_calor_iso_cut = 5.5
untracked double muon_track_iso_cut = 5.0
}

# Modules to cut on the number of electrons, muons and leptons

module CountElectrons = TopLeptonCountFilter
{
    InputTag electronSource = LeptonIDFilterModule:selectedElectrons
    InputTag muonSource     = LeptonIDFilterModule:selectedMuons
    InputTag tauSource      = selectedLayer1TopTaus
    bool        countElectrons = true
    bool        countMuons    = false
    bool        countTaus     = false
    uint32      minNumber     = 1
    uint32      maxNumber     = 1
}

module CountMuons = TopLeptonCountFilter
{
    InputTag electronSource = LeptonIDFilterModule:selectedElectrons
    InputTag muonSource     = LeptonIDFilterModule:selectedMuons
    InputTag tauSource      = selectedLayer1TopTaus
    bool        countElectrons = false
    bool        countMuons    = true
    bool        countTaus     = false
    uint32      minNumber     = 1
    uint32      maxNumber     = 1
}

module CountLeptons = TopLeptonCountFilter
{
    InputTag electronSource = LeptonIDFilterModule:selectedElectrons
    InputTag muonSource     = LeptonIDFilterModule:selectedMuons
    InputTag tauSource      = selectedLayer1TopTaus
    bool        countElectrons = true
    bool        countMuons    = true
    bool        countTaus     = false
    uint32      minNumber     = 1
    uint32      maxNumber     = 1
}

# Definition of the time-order in which the modules are to run

sequence LeptonSequence = { newSelectedLayer1TopElectrons, newSelectedLayer1TopMuons,
                             LeptonIDFilterModule, CountLeptons }

sequence ElectronStream = { LeptonSequence, CountElectrons }
sequence MuonStream     = { LeptonSequence, CountMuons }

path p1 = { ElectronStream }
path p2 = { MuonStream }

# Data output modules, also defining which data collections shall be kept and dropped

```

```

module out_PhotonIDModuleElectronStream = PoolOutputModule
{
    untracked string fileName = "LeptonID_ElectronEvents_Data.root"

    untracked PSet SelectEvents =
    {
        vstring SelectEvents = { "p1" }
    }

    untracked vstring outputCommands =
    {
        keep "*",

        "drop TopElectrons_allLayer1TopElectrons_*_TQAF",
        "drop TopElectrons_selectedLayer1TopElectrons_*_TQAF",
        "drop TopElectrons_newSelectedLayer1TopElectrons_*_TQAF",
        "drop TopMuons_allLayer1TopElectrons_*_TQAF",
        "drop TopMuons_selectedLayer1TopElectrons_*_TQAF"
    }
}

module out_PhotonIDModuleMuonStream = PoolOutputModule
{
    untracked string fileName = "LeptonID_MuonEvents_Data.root"

    untracked PSet SelectEvents =
    {
        vstring SelectEvents = { "p2" }
    }

    untracked vstring outputCommands =
    {
        "keep *",

        "drop TopElectrons_allLayer1TopElectrons_*_TQAF",
        "drop TopElectrons_selectedLayer1TopElectrons_*_TQAF",
        "drop TopMuons_allLayer1TopElectrons_*_TQAF",
        "drop TopMuons_selectedLayer1TopElectrons_*_TQAF"
    }
}

# Final sequence to secure that the data are written to hard disk

sequence s = {out_PhotonIDModuleElectronStream & out_PhotonIDModuleMuonStream}

endpath e = {s}
}

```


Appendix C

Mathematical Definitions

THIS section contains the mathematical definitions for the selection efficiency and selection purity as they have been used throughout this thesis. In order to facilitate the following definitions some symbols are defined. The symbol N denotes any number of countable measure which is not closer specified so far. In contrast, the symbols S and B stand for a number of signal and background events. Instead of events the numbers of a certain particle or class of particle can be meant as well.

- **Selection Efficiency**

If N_0 denotes the number of events before a cut is applied, and N is the number after that cut, then the *selection efficiency* ε is defined by

$$\varepsilon = \frac{N}{N_0}. \quad (\text{C.1})$$

- **Selection Purity**

The *selection purity* P for a signal is defined by

$$P = \frac{S}{S + B}. \quad (\text{C.2})$$

The selection purity for B can be calculated when exchanging $S \leftrightarrow B$.

- **Error on Selection Efficiency and Selection Purity**

The error on the selection efficiency and selection purity are calculated by the error propagation. Variables with a preceding Δ denote the error of that value.

$$\Delta\varepsilon = \sqrt{(|\frac{1}{N_0}| \cdot \Delta N)^2 + (|-\frac{N}{N_0^2}| \cdot \Delta N_0)^2} \quad (\text{C.3})$$

$$\Delta P = \frac{1}{(S + B)^2} \cdot \sqrt{(B^2 \cdot \Delta S^2 + S^2 \cdot \Delta B^2)} \quad (\text{C.4})$$

$$\Delta(\varepsilon \cdot P) = \sqrt{(|P| \cdot \Delta\varepsilon)^2 + (|\varepsilon| \cdot \Delta P)^2} \quad (\text{C.5})$$

List of Figures

2.1	Geographical Position of the LHC	6
2.2	CMS Detector	7
3.1	Silicon Tracking Detector	12
3.2	Silicon Sensor Cross Section	13
3.3	Silicon Microstrip Module	14
3.4	Frontend Hybrid of a TEC Ring-6 Module	15
3.5	ARC Test Setups	17
3.6	Distribution of Opens and Regular Channels	18
3.7	Single Channel Defects	19
4.1	IV-curves without and with IV-breakdown	23
4.2	Repairing of Crack in Carbon Frame	25
4.3	Gantry Precision Distributions	26
4.4	Pipeline Defects	28
4.5	Module Bonds and Pushed Bonds	29
4.6	Final TEC Module Debug Centre Statistics	31
5.1	Coupling Scenarios	36
5.2	Feynman Diagrams of $t\bar{t}\gamma$ Processes	37
5.3	Generator Validation Distributions	40
5.4	Feynman Graph for the Estimation of Interference Effects	41
5.5	Estimation of Interference Effects	42
5.6	Relative Cross Sections before Event Selection	43
5.7	Event Simulation Flow Chart	44
5.8	Material Budget of the Silicon Tracker	46
5.9	Scheme of the Island Algorithm	47
5.10	Scheme of the Hybrid Algorithm	48
5.11	Multiplicity of Reconstructed Objects	50
5.12	Flow Chart of the Analysis	51
6.1	Candidate Multiplicity after Preselection Cuts	55

6.2	Candidate Energy/Momenta Distribution after Preselection Cuts	56
7.1	Lepton p_T - Distributions	62
7.2	Electron ID Criteria - Signal Electron Events	63
7.3	Muon ID Criteria - Signal Muon Events	65
7.4	Number of Identified Leptons	67
7.5	Correlation between Fake Photon and Generator Top Photon Candidates	69
7.6	Photon Generator Matching - Signal Electron Events, Top Photons	70
7.7	Photon Generator Matching - Signal Electron Events, Non-Top Photons	71
7.8	Photon Generator Matching - Signal Electron Events, Fake Photons	71
7.9	Photon Generator Matching - Signal Muon Events, Top Photons	72
7.10	Photon Generator Matching - Signal Muon Events, Non-Top Photons	72
7.11	Photon Generator Matching - Signal Muon Events, Fake Photons	73
7.12	Photon Object Creation Scheme	74
7.13	Photon ID Efficiency Correlation - Signal Electron Events	78
7.14	Photon ID Correlation Matrices - Signal Electron Events	79
7.15	Photon ID Criteria - Signal Electron Events	80
7.16	Photon ID Efficiency Correlation - Signal Muon Events	81
7.17	Photon ID Correlation Matrices - Signal Muon Events	82
7.18	Photon ID Criteria - Signal Muon Events	83
7.19	Photon Generator Matching - Signal Electron Events, Fake Photons after Photon ID . . .	85
7.20	Photon Generator Matching - Muon Events, Fake Photons after Photon ID	86
7.21	Photon Electron Correlation	86
7.22	E/p of Electrons Matching to fake Photons	87
7.23	Photon-Jet Correlations, Signal Electron Events	89
7.24	Photon-Jet Correlations, Signal Muon Events	90
8.1	Distribution of b-tag Discriminator Values	94
8.2	Jet p_T Distributions, Electron Event Stream	96
8.3	Jet p_T Distributions, Muon Event Stream	97
8.4	MET Distributions	98
8.5	Reconstructed Masses of Hadronic W-Bosons and Top-Quarks	99
8.6	Correlation between Photons and Reconstructed Top Quarks	100
9.1	Relative Cross-Sections Scaled by Selection Efficiency	105
10.1	Standard Model Spectra	110
10.2	$t\bar{t}\gamma$ Vertex	111
10.3	Baur Photon Energy Spectra and Scaling Function	112
10.4	Photon Spectra for $q_{top} = 2/3$ and $q_{top} = -4/3$	114
10.5	Confidence Belts $q_{top} = 2/3$	115

10.6	Confidence Belts $q_{top} = -4/3$	116
10.7	Scaling Function for Different Charge Quantum Numbers	117
10.8	Limits on the Electromagnetic Charge of the Top-Quark	117
10.9	Limits on the Electromagnetic Charge of the Top-Quark versus Lumionosity	119

List of Tables

1.1	Standard Model of Particle Physics: Fermions	1
1.2	Standard Model of Particle Physics: Gauge Bosons	2
2.1	Key Parameters of the CMS Sub-Detectors.	9
3.1	Measure for Finding Faulty Channels	17
4.1	Modules with IV-Problems	24
4.2	Modules with Hybrid-Problems	24
4.3	Modules with Broken Carbon Frame	26
4.4	Grading of Misaligned Modules	27
4.5	TEC Module Grading	27
4.6	Regrading of Grade C Modules	28
4.7	Number of Modules with Pushed Bonds	30
5.1	Background Data-Sets	43
6.1	Preselection Statistics	57
7.1	Lepton Numbers after Lepton ID	66
7.2	Results of the Lepton ID	66
7.3	Photon Candidate Numbers before Photon Identification	69
7.4	Results of the Photon ID	84
7.5	Photon Numbers after Photon ID	84
7.6	Results of the Photon ID after Particle Correlations	89
7.7	Photon Numbers after Particle Correlations	91
8.1	Efficiencies Event Solutions	99
8.2	Final Event Statistics	101
9.1	Final Event Selection Efficiency	104
9.2	Final S/B Ratio	106
10.1	Limits on the Signal Events	115

A.1	Selection Statistics on $t\bar{t}$ + Jets Events	128
A.2	Selection Statistics on W + Jets Events	128
A.3	Selection Statistics on Z + Jets Events	129
A.4	Selection Statistics on Di-Boson Events	129
A.5	Selection Statistics on γ + Jets Events	130
A.6	Selection Statistics on QCD-Multijet Events	131

References

- [1] **D. Gross et al.**
Asymptotically Free Gauge Theories
Physical Review Letters D 8, 3633 - 3652, November 1973.
- [2] **S. Weinberg**
A Model for Leptons
Physical Review Letters D 19, 1264 - 1266, November 1967.
- [3] **Particle Data Group**
The Review of Particle Physics
Physics Letters B667, 1 (2008)
→ <http://pdg.lbl.gov>.
- [4] **G. Altarelli et al.**
Precision Electroweak Tests of the Standard Model
arXiv:hep-ph/0404165v1
- [5] **S. Forte**
Tests of the Standard Model at HERA
arXiv:hep-ph/9910397v1
- [6] **The LHC study group**
The Large Hadron Collider, Conceptual Design
CERN/AC95-05, October 1995.
- [7] **M. H. Zöller**
Private Communication, February 2008.
- [8] **Website CMS Outreach**
→ <http://cmsdoc.cern.ch/cms/outreach/html/CMSdocuments/DetectorDrawings/DetectorDrawings.html>.
- [9] **ATLAS Collaboration**
Technical Proposal, Technical report
CERN/LHCC 94-43 LHCC/P2, 1994.
- [10] **The CMS Collaboration**
Technical Proposal, Technical report
CERN/LHCC 94-38 LHCC/P1, 1994.
- [11] **The LHCb Collaboration**
Technical Proposal, Technical report
CERN/LHCC 98-4 LHCC/P4, 1998.

-
- [12] **The ALICE Collaboration**
Technical Proposal, Technical report
CERN/LHCC 95-71 LHCC/P3, 1995.
- [13] **The CMS Collaboration**
CMS Physics Technical Design Report, Addendum on High Density QCD with Heavy Ions
Journal of Physics G: Nuclear and Particle Physics 34 (2007) 2307-2455, October 2007.
- [14] **The CMS Collaboration**
The Tracker Project, Technical Design Report
CERN/LHCC 1998-006, April 1998.
- [15] **The CMS Collaboration**
Addendum to the CMS Tracker TDR
CERN/LHCC 2000-016, February 2000.
- [16] **The CMS Collaboration**
The Electromagnetic Calorimeter, Technical Design Report
CERN/LHCC 1997-033, December 1998.
- [17] **The CMS Collaboration**
The Hadron Calorimeter, Technical Design Report
CERN/LHCC 1997-031, June 1997.
- [18] **The CMS Collaboration**
The Magnet Project, Technical Design Report
CERN/LHCC 1997-010, December 1997.
- [19] **The CMS Collaboration**
The Muon Project, Technical Design Report
CERN/LHCC 1997-032, December 1997.
- [20] **The CMS Collaboration**
The Trigger and Data Acquisition Project Volume 1, The Level-1 Trigger, Technical Design Report
CERN/LHCC 2000-038, December 2006.
- [21] **The CMS Collaboration**
The Trigger and Data Acquisition Project Volume 2, Data Acquisition and High-Level Trigger, Technical Design Report
CERN/LHCC 2002-026, December 2002.
- [22] **The CMS Collaboration**
Detector Performance and Software, Physics Technical Design Report, Volume 1
CERN/LHCC 2006-001, February 2006.
- [23] **The CMS Collaboration**
The CMS experiment at the CERN LHC
To be published in the Journal of Instrumentation (JINST).
- [24] **G. Lutz**
Semiconductor Radiation Detectors, Device Physics
Springer-Verlag Berlin Heidelberg, 1999.

- [25] **T. Franke**
Development and Evaluation of a Test System for the Quality Assurance during the Mass Production of Silicon Microstrip Detector Modules for the CMS Experiment
Dissertation RWTH Aachen, July 2005.
- [26] **L. Borrello et al.**
Sensor Design for the CMS Silicon Strip Tracker
CMS Note 2003/020 (2003).
- [27] **W.R. Leo**
Techniques for Nuclear and Particle Physics Experiments
Springer-Verlag Berlin Heidelberg, 1994.
- [28] **L. Jones**
APV25-S1, User Guide Version 2.2, September 2001
→ http://www.te.rl.ac.uk/med/projects/High_Energy_Physics/CMS/APV25-S1/pdf/User_Guide_2.2.pdf.
- [29] **P. Murray**
APVMUX User Guide version 1.0, Mai 2000
→ http://www.te.rl.ac.uk/med/projects/High_Energy_Physics/CMS/APVMUXPLL/pdf/UserGuide.pdf.
- [30] **C. Foudas et al.**
The CMS Tracker Readout Front End Driver
IEEE Transactions on Nuclear Science, Volume 52, No. 6, December 2005.
- [31] **P. Placidi et al.**
CMS Tracker PLL Reference Manual, Version 2.0, Juli 2000
→ <http://cmstrackercontrol.web.cern.ch/CMSTrackerControl/documents/PauloMoreira/TrackerPLLManual.pdf>.
- [32] **G. Magazzu et al.**
DCU2 User Guide, Oktober 2001
→ http://cmstrackercontrol.web.cern.ch/CMSTrackerControl/documents/Magazzu/DCU2_User_Manual%20v2.12.pdf.
- [33] **Philips Semiconductors**
The I²C-Bus Specifications
Version 2.1, January 2000.
- [34] **M. Axer**
Development of a Test System for the Quality Assurance of Silicon Microstrip Detectors for the Inner Tracking System of the CMS Experiment
Dissertation RWTH Aachen, December 2003.
- [35] **M. Poettgens**
Development and evaluation of test stations for the quality assurance of the silicon micro-strip detector modules for the CMS experiment
Dissertation RWTH Aachen, November 2007.
- [36] **M. Axer et. al.**
The Qualification of Silicon Microstrip Detector Modules for the CMS Inner Tracking Detector
CMS Note 2006/141, October 2006.

-
- [37] **CMS Tracker Database**
CMS Tracker Database Browser
→ <http://cmsdoc.cern.ch/~cmstrkdb/>.
- [38] **A. Linn**
Construction of the CMS Tracker Endcaps and an Impact Study on Defects
Dissertation RWTH Aachen, November 2008.
- [39] **Th. Hermanns**
Aufbau eines Systems für Kühlttests zur Qualitätsüberwachung von CMS Silizium-Modulen
Diploma Thesis (in German) RWTH Aachen, January 2004.
- [40] **S. Costa**
CMS Tracker Module Failure Report User Interface
→ <http://cms.ct.infn.it/failures/>.
- [41] **W. Braunschweig**
Private Communication, December 2005.
- [42] **L. Fiore et al.**
Assembly of the Inner Microstrip Tracker Modules
CMS Note 2008/004, July 2007.
- [43] **Th. Bergauer**
Module Production Yield
Presentation TEC Meeting, December 2005.
- [44] **I. Tomalin**
Problem due to HIPS & Pinholes
Presentation CMS General Tracker Meeting, January 2002.
- [45] **B. Reinhold**
Automatic Bonding test of Front End Hybrids for the CMS Tracker
Diploma Thesis RWTH Aachen, September 2004.
- [46] **A. Honma**
Private Communication, November 2005.
- [47] **R. Bremer**
Integration of the end cap TEC+ of the CMS silicon strip tracker
Dissertation RWTH Aachen, April 2008.
- [48] **The CDF Collaboration**
Evidence for Top Quark Production in $\bar{p}p$ Collisions at $\sqrt{s}=1.8\text{ TeV}$
Physical Review Letters, Volume 73, Number 2, July 1994.
- [49] **The Tevatron Electroweak Working Group**
Combination of CDF and D0 Results on the Mass of the Top Quark
arXiv:hep-ex/0808.1089v1, August 2008.
- [50] **D. Chang et al.**
Alternative Interpretation of the Fermilab Tevatron top events
Physical Review Letters D 59 091503, March 1999.

-
- [51] **The D0 Collaboration**
Experimental Discrimination between charge $2e/3$ top quark and charge $4e/3$ exotic quark production scenarios
arXiv:hep-ex/0608044v2, January 2008.
- [52] **T. Sjöstrand et al.**
PYTHIA 6.3 Physics and Manual
arXiv:hep-ph/0308153, August 2003.
- [53] **M. Mangano et al.**
ALPGEN, a generator for hard multiparton processes in hadronic collisions
arXiv:hep-ph/0206293, December 2006.
- [54] **J. Allison et al.**
Geant4 developments and applications
IEEE Transactions on Nuclear Science 53 No. 1 (2006) 270-278, February 2006.
- [55] **R. Brun et al.**
ROOT - An Object Oriented Data Analysis Framework
Nuclear Instruments and Methods in Physics Research A 389 (1997) 81-86, September 1996
See also $\rightarrow$ <http://root.cern.ch/>.
- [56] **D. Bonacorsi**
Computing Operations at CMS Facilities
Presentation on the International Conference on Computing in High Energy and Nuclear Physics, Victoria BC, Canada, September 2007.
- [57] **W. Adam et al.**
Track Reconstruction in the CMS Tracker
CMS Note 2006/041, December 2005.
- [58] **E. Meschi et al.**
Electron Reconstruction in the CMS Electromagnetic Calorimeter
CMS Note 2001/034, June 2001.
- [59] **S. Baffioni et al.**
Electron Reconstruction in CMS
CMS Note 2006/040, February 2006.
- [60] **A. Heister et al.**
Measurements of Jets with the CMS Detector at the LHC
CMS Note 2006/036, February 2006.
- [61] **S.V. Chekanov**
Jet algorithms: a minireview
arXiv:hep-ph/0211298v1, November 2002.
- [62] **G. Blazey et al.**
Run II Jet Physics
arXiv:hep-ex/0005012v2, May 2000.
- [63] **S. Ellis et al.**
Successive Combination Jet Algorithm For Hadron Collisions
Physical Review Letters D 48 (1993) 3160-3166, May 1993.

- [64] **V. Konopliyanikov et al.**
Jet Calibration using γ + Jet Events in the CMS Detector
CMS AN 2006/042, February 2006.
- [65] **G. Bruno et al.**
Local Reconstruction in the Muon Detectors
CMS Note 2002/043, December 2002.
- [66] **CMS Remote Analysis Builder**
→ <http://cmsdoc.cern.ch/cms/cc/wm/www/Crab>.
- [67] **S. Lowette et al.**
The Top Quark Analysis Framework
CMS Internal Note 2007/068, December 2007.
- [68] **A. Rizzi et al.**
Track impact parameter based b-tagging with CMS
CMS Note 2006/019, January 2006.
- [69] **The CMS Collaboration**
Physics Performance, Physics Technical Design Report, Volume 2
Journal of Physics G: Nuclear and Particle Physics 34 (2007) 995-1579, April 2007.
- [70] **J. D'Hondt et al.**
Fitting of Event Topologies with External Kinematic Constraints in CMS
CMS Note 2006/023, January 2006.
- [71] **G. Hammand**
LR-based background suppression for $t\bar{t}$ events
Presentation TQAF Meeting, 13. June 2007.
- [72] **U. Baur et al.**
Probing electroweak top quark couplings at hadron colliders
Physical Review Letters D 71 (2005) 054013, March 1993.
- [73] **U. Baur**
Private Communication, December 2008.
- [74] **G. Feldman et al.**
Unified approach to the classical statistical analysis of small signals
Physical Review Letters D 57 (1998), 3873 - 3889, April 1998.

Acknowledgment

This last page is reserved for all people who accompanied and encouraged me during the last years.

- Prof. Dr. Achim Stahl and Prof. Dr. Joachim Mnich offered me the opportunity to work not only in the field of detector construction, but also to develop some ideas about top-physics for this thesis. Even at the time when I had to change from the first from the second topic they still gave me the freedom to complete my job for the CMS module production. Therefore, I must thank them not only for their scientific support but also for their patience.
- Thanks a lot to Dr. Oliver Pooth not only for having always his door widely open, but especially for proof-reading this thesis. I suppose, that this must have been a hard job, but your feedback has been highly appreciated.
- Dr. Heiko Geenen and Dr. Marc Henning Zöller, thank you for your helpful support with respect to all kinds of statistical affairs for the top-quark analysis. In addition, I will really miss our frequent discussions about the various aspects of elementary particle physics. You helped me a lot in getting prepared for the final exam.
- Manuel Giffels and Daiske Tornier, thanks a lot for patiently answering my questions about programming and computing. I never tried to count how often you explained me what my source code did instead of what it was intended to do.
- Then I must thank the members of the ARC system project. It has been a pleasure to be a member of that project, but especially I would like to mention Dr. Torsten Franke and Dr. Michael Pöttgens. You both taught me a lot about the functionality of silicon microstrip detector modules and patiently answered so many questions. Apart from that, the atmosphere in the “Tanzsaal” will remain unmatched.
- My thank also goes to the Aachen-IIIB top-group. The lively discussions about all aspects on my analysis were very welcomed. In this respect, I must mention explicitly Dr. Stefan Kasselmann and Daiske Tornier who assisted me with my first steps in the field of top-quarks.
- Thanks to all other not explicitly mentioned colleagues who worked or are still working at the III. Physikalisches Institut B. You always generated a pleasant atmosphere. I really enjoyed working for this institute.
- The last but one special gratitude is dedicated to Stefan Kasselmann, Alexander Linn and Daiske Tornier. To my mind we occupied the most exciting office on the yellow floor.
- But all would be nothing without the everlasting support of my parents.
Danke Mama, danke Papa, dass Ihr mich zur dieser Doktorarbeit ermutigt, mich während der gesamten Zeit immer vorbehaltlos unterstützt und mir so viele Aufgaben des Alltages abgenommen habt. Jede Beschreibung kann der gesamten Sache eigentlich nicht wirklich gerecht werden. Daher also einfach noch einmal: DANKE!

Lebenslauf/Curriculum Vitae

Persönliche Daten

Name Thomas Hermanns
Geburtsdatum 11. Dezember 1978
Geburtsort Erkelenz

Schulbildung

1985 – 1989 Luise-Hensel-Grundschule Erkelenz
1989 – 1998 Cornelius-Burgh-Gymnasium Erkelenz
Juni 1998 Abschluss mit der allgemeinen Fachhochschulreife

Wehrdienst

1998 – 1999 Marineoperationsschule Bremerhaven und 1. Zerstörergeschwader Kiel

Studium

1999 – 2004 Studium der Physik an der RWTH Aachen
Januar 2003 Beginn der Diplomarbeit
Titel: Aufbau eines Systems für Kühltests zur Qualitätsüberwachung von CMS-Silizium-Modulen
März 2004 Abschluss als Diplom-Physiker
Schwerpunkt: Experimentelle Elementarteilchenphysik
Nebenfach: Astronomie und Astrophysik

Promotion und Berufsweg

März 2004 Beginn der Promotion, RWTH Aachen, III. Physikalisches Institut B
März 2004 –
Oktober 2008 Wissenschaftlicher Angestellter, RWTH Aachen, III. Physikalisches Institut B
09. September 2008 Einreichen des Promotionsgesuchs
seit November 2008 Fellow bei der European Organization for Nuclear Research (CERN), Genf, Schweiz
15. Dezember 2008 Mündliche Doktorprüfung