INCLUSIVE SINGLE LEPTON TRIGGER STUDIES FOR TOP PHYSICS AT THE CMS EXPERIMENT

AFIQ AIZUDDIN ANUAR

DISSERTATION SUBMITTED IN FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE
(EXCEPT MATHEMATICS & SCIENCE PHILOSOPHY)

DEPARTMENT OF PHYSICS
UNIVERSITY OF MALAYA
KUALA LUMPUR

2015

UNIVERSITI MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: AFIQ AIZUDDIN ANUAR (I.C./Passport No.: 900918-08-6367)
Matric No.: SGR 120107

Name of Degree:
MASTER OF SCIENCE (EXCEPT MATHEMATICS & SCIENCE PHILOSOPHY)

Title of Dissertation ("this Work"):
INCLUSIVE SINGLE LEPTON TRIGGER STUDIES FOR TOP PHYSICS AT THE CMS EXPERIMENT

Field of Study: High Energy Physics

I do solemnly and sincerely declare that:

1. I am the sole author/writer of this Work;
2. This Work is original;
3. Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the tittle of the Work and its authorship have been acknowledged in this Work;
4. I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
5. I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
6. I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness' Signature Date

Name:
Designation:

Abstract

Due to the overwhelming rate of data delivered by the Large Hadron Collider, a preselection system is necessary to reduce it to a size manageable by our available computing resources. This is done by the trigger system which performs a fast filtering of events to be saved by running a version of the event reconstruction algorithm optimized for the fast online environment. In this thesis, the inclusive single lepton triggers are studied and optimized for CMS Run 2 data-taking, which focuses on events containing at least one isolated muon or electron, within the context of physics involving top quarks.

Abstrak

Disebabkan kadar data yang sangat besar yang diperolehi daripada Pelanggar Elektron Besar, suatu sistem pra-pemilihan adalah penting untuk mengurangkan saiznya kepada saiz yang boleh diterima oleh sumber-sumber komputeran yang ada. Tugas ini dilakukan oleh sistem pencetus yang melakukan pemilihan secara pantas dengan melaksanakan suatu versi kepada algoritma pembentukan semula yang telah dioptimumkan untuk keadaan diatas talian yang pantas. Didalam tesis ini, pencetus lepton tunggal inklusif telah dikaji dan dioptimumkan untuk pengambilan data di CMS Run 2, yang memfokuskan kepada peristiwa-peristiwa yang mengandungi sekurang-kurangnya satu muon atau elektron terasing, didalam konteks fizik yang melibatkan top kuark.

Acknowledgements

I was indebted to many towards the production of this work. While it would be impossible to thank them all, there are some people whom are of particular significance that I nonetheless wish to mention.

First and foremost, the NCPP managing directors, Wan Ahmad Tajuddin Wan Abdullah and Zainol Abidin Ibrahim for provision of resources and facilities, without which much of this work would not have been possible.

Secondly, my supervisor, Jyothsna Rani Komaragiri, for giving me direction when I needed it the most and guiding me along the way. Thanks to you I got to know some of the most wonderful people to work with and more, a field that is becoming more and more a calling with each passing day.

Third, my CMS collaborators, for many direct and indirect contributions in producing this work. Among them, I would like to especially mention Javier Fernandez, Nadjieh Jafari and Matteo Sani for their many helpful input. Next, my gratitude goes to Andrey Popov for being such an amazing collaborator and source of guidance, encouragement and depression alike. Last but not least, I would like to offer my heartfelt gratitude to Stephanie Beauceron, by far the best superior I have ever had the pleasure to work under. Her kindness and sincerity had helped me far beyond the scope of this work; if there was someone I would thank for bringing me to where I am today, that person would be her.

Fourth, my colleagues and friends, for being a source of amusement from time to time. My special thanks go to Siew Yan Hoh, Nur Zulaiha Jomhari and Nurfikri Norjoharuddeen for being the closest to me, and therefore the source I derived the most amusement from.

Lastly, I thank my family, particularly my parents, maternal grandmother (*Tokpah*), uncles and aunts for their continuous support...

This work is dedicated to all of you.

Contents

List of figures

List of tables

List of acronyms

p_T	Transverse momentum
CMS	Compact Muon Solenoid
CMSSW	CMS SoftWare
ECAL	Electromagnetic calorimeter
GSF	Gaussian Sum Filter
HCAL	Hadronic calorimeter
HLT	High Level Trigger
KF	Kalman Filter
MC	Monte Carlo
PF	Particle Flow
PU	Pile-up
QCD	Quantum Chromodynamics
SF	Scale factor

Chapter 1

Introduction

"In any moment of decision, the best thing you can do is the right thing."

— Theodore Roosevelt, 1858 – 1919

To date, the Standard Model (SM) [1, 2], the theory describing the interactions between all known fundamental particles, is the most successful theory developed by mankind. In order to verify its validity, many large experiments have been built since decades ago. The most recent one to join the hunt, the Large Hadron Collider (LHC), boasting the highest energy achievable to date, are tasked with many crucial tests, including finding the hints on territories uncharted by the SM. Due to the extreme rarity of the phenomena sought by the LHC, they could only be discovered by sifting through vast amounts of data, far more than what could be processed by our computing resources. Thus the need for a system to filter the data keeping only the most interesting events, what we have come to call the trigger system. This thesis focuses on

the optimization of the single lepton triggers at the Compact Muon Solenoid (CMS) experiment, one of two general purpose experiments at the LHC.

1.1 Project Statement

1.1.1 Inclusive Single Lepton Triggers Optimization

The trigger paths studied in this project are the inclusive single lepton triggers. Inclusive means that only the existence of one lepton of the desired type is needed to fire the paths regardless of any other object present in the same event, as long as the lepton passes some identification cuts to be described in their respective chapters. Lepton within the context of this thesis is restricted only to muons and electrons among the six types in the Standard Model [1]. This is because three of them, the neutrinos, interact very weakly with the remaining particle types (to the point that they could pass through the whole planet without interacting) and therefore escape the detector undetected. Among the charged leptons, the heaviest tau has a short lifetime and therefore decay before reaching the detector, thus requiring a more sophisticated reconstruction technique which is not within the scope of this project. This leaves only muons and electrons as the directly detectable leptons and consequently, they are the ones with dedicated trigger paths studied in this project.

1.1.2 Project Motivation and Objective

The primary focus of this project is the optimization of the single lepton triggers, which within this context means that the cuts used in the triggers are set so

as to minimize the background contribution for a given signal efficiency. The necessity for doing so lies in the fact that every event passing the triggers contribute to the overall rate of the path. If a given trigger path primarily record background events, this is both wasteful from the computing and storage resources point of view and contradictory to the purpose of the trigger system, i. e. to record events of physics interest.

In optimizing the single lepton trigger paths, the reconstruction algorithms of muon and electrons are studied and the cuts in the form of identification variables are retuned to better suit the harsher conditions in 2015 Run 2 data taking. The optimization focuses on the second part of 25 ns bunch crossing scenario, with 1.4×10^{34} cm^{-2} s^{-1} instantaneous luminosity and 40 average pile-up interactions (PU). This choice is made because this is the harshest data-taking scenario in 2015 and therefore the most challenging environment for the HLT. The project focuses on the unprescaled single lepton paths with lowest p_T threshold which is usually used to trigger events involving leptonically decaying W boson that occur in a wide rage of physically interesting final states at the CMS experiment, such as vector boson, top quarks and Higgs boson through the associated production modes. As muons and electrons have different physical properties, their reconstruction in CMS are done in very different ways. Consequently the optimization procedure is also different for these two paths, with all the relevant details discussed in their respective chapters.

In summary, the objectives of this project are:

1. To optimize the single muon trigger by tightening the selection criteria used in stand-alone and global muon reconstruction

2. To optimize the single electron trigger by tightening the selection criteria used in electron identification such as cluster shape and variables based on calorimeter energy deposit

1.1.3 Thesis Outline

The thesis is organized into several chapters as follows. Chapter 2 describes the CMS detector; focusing in particular on the parts relevant for this project. Chapter 3 describes muon reconstruction at trigger level and single muon trigger optimization. Chapter 4 describes the electron reconstruction at trigger level, single electron trigger optimization leading toward the creation of a new working point and studies associated with it. Chapter 5 describes the data-driven measurement of muon leg efficiencies of the single muon cross triggers, done on 7 TeV CMS data. Finally, the entire project is summarized in Chapter 6.

Chapter 2

Experimental Background

"Elementary particles are terribly boring, which is one reason why we're so interested in them."

— S. Weinberg, 1933 – Present

2.1 The Large Hadron Collider (LHC)

The LHC was installed following the dismantling of the previous flagship accelerator operated by European Organization for Nuclear Research (CERN), the Large Electron Positron (LEP) collider [3]. Occupying the original 27 km tunnel used for LEP, it was designed to collide proton beams of 7 TeV energy, for a center of mass-energy of 14 TeV. It could also be used to collide lead nuclei in the so-called heavy ion collisions. Located at different points along the LHC ring are the four major experiments, built to test various aspects of the SM:

- CMS (Compact Muon Solenoid): A general purpose detector that discovered the Higgs boson in 2012 and will focus on precise measurement of Higgs properties and physics beyond the SM in Run 2, starting in 2015
- ATLAS (A Toroidal LHC ApparatuS): A general purpose detector like CMS but built with a different design emphasis, serving also as a mutual cross-check of CMS results
- LHCb (Large Hadron Collider beauty): A single-arm forward spectrometer dedicated for b-physics and related studies
- ALICE (A Large Ion Collider Experiment): Focusing on heavy ion collisions to study the properties of quark-gluon plasma

2.2 The Compact Muon Solenoid (CMS) Experiment

As a general purpose detector, the ability to detect and identify many different kinds of objects in a wide angular coverage is very important to CMS [4]. To fulfil this requirement, the detector is divided into multiple components arranged in a concentric layered structure that was optimized according to their functionalities. In order to obtain precise measurements of position and momentum of the objects, the exact orientations and positions of these components must be known to a high accuracy, both in the absolute sense and relative to other components. To facilitate this process, these parameters are expressed as points and directions in a coordinate system known as the CMS coordinate system.

2.2.1 CMS Coordinate System

CMS coordinate system takes as origin the nominal interaction point inside the detector. The x-axis is pointing toward the center of the LHC ring from this origin, while the y-axis points vertically upward. The z-direction on the other hand is defined along the beam direction towards the Jura mountains from the LHC Point 5 where the CMS detector is installed.

With this coordinate system, the azimuthal angle ϕ is measured from the x-axis within the transverse x-y plane. The polar angle θ is measured with respect to the z-axis. On the other hand, pseudorapidity η is defined as $\eta = -ln(tan(\theta/2))$. Another important quantity that is commonly used is the transverse momentum p_T, which is the component of the momentum in the transverse plane.

2.2.2 CMS Detector Components

The CMS detector is a massive machine spanning a length of 21.6 m and diameter 14.6 m at a total weight of 12500 tons. Figure 2.1 shows a transverse slice revealing the concentric layer structure of the detector. In the outermost layer sits the muon system which consists of aluminium drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC); these components work in tandem to provide accurate measurements of muon position and momenta with a high reconstruction efficiency. The muon system components are divided into 4 layers of detector component, which are called 'stations'. This is done so that the position of hits in each stations could be connected together to form a trajectory in the track reconstruction procedure, from which the muon

position and momentum could be measured. Interspersing these layers are iron yokes used to saturate the magnetic field from inside the detector.

Covered by the muon system is the superconducting solenoid responsible for the 3.8 T magnetic field permeating the detector. The solenoid spans 13 m in length with an inner diameter of 5.9 m. The high momentum resolution of charged particles is achieved thanks to the powerful magnetic field provided by the solenoid.

Going deeper inside, we have the calorimeter system fit inside the solenoid. The outer layer is the hadronic calorimeter (Hcal) consisting of interleaving brass and scintillating plates, dedicated to the measurement of energy of hadronic particles and missing transverse energy. It is further divided based on the η regions into Hcal-Barrel (HB) and Hcal-Endcap (HE) respectively, providing a total coverage up to $|\eta| < 3$. The energy resolution (measured in GeV) of the HB is [6]:

$$\frac{\sigma(E)}{E} = \frac{84.7\%}{\sqrt{E}} \oplus (7.4\%) \tag{2.1}$$

The energy resolution of the HE is found to be similar as above. Additionally, the absolute energy scale of the Hcal has been checked by comparing the results from muon beam tests and cosmic muons [7]. In the very forward region is the Hcal-Forward calorimeter which provides full geometry coverage (up to $|\eta| < 5$) for the measurement of missing transverse energy.

Inside the Hcal is the electromagnetic calorimeter (Ecal) dedicated to measuring the energy of electromagnetic particles such as electrons and photons.

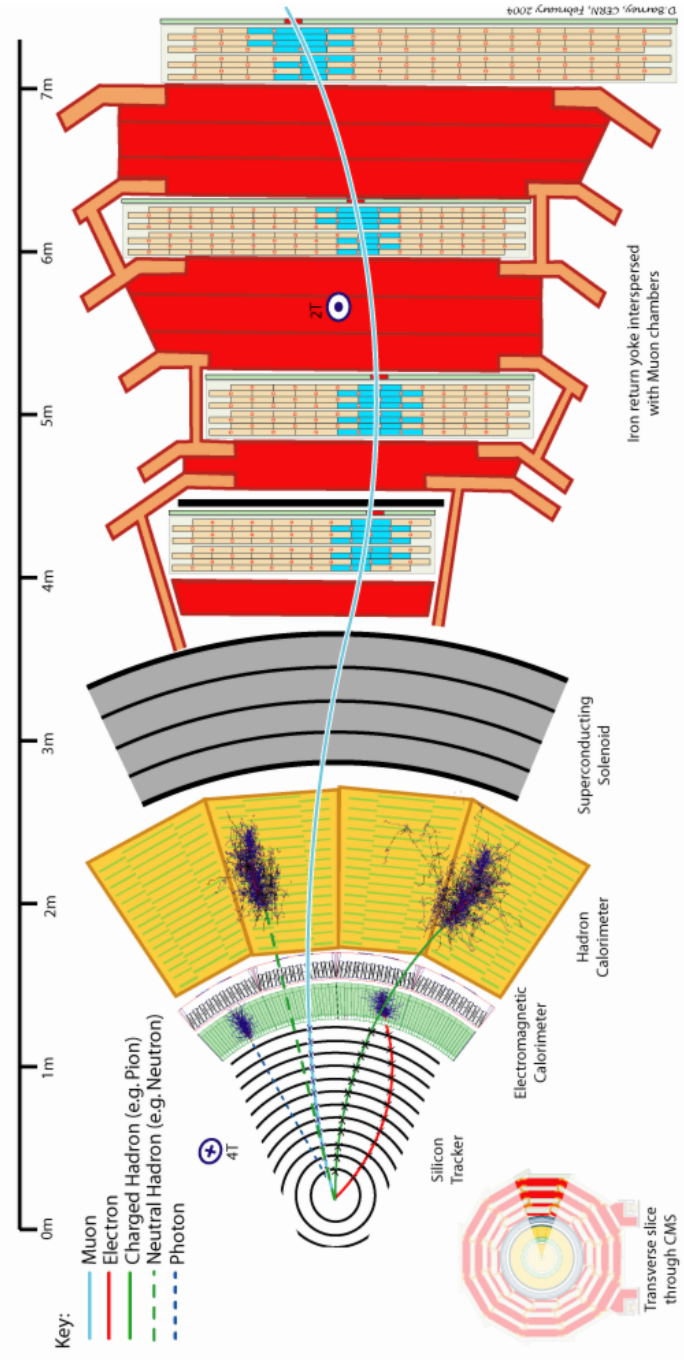


Figure 2.1: Transverse slice of the CMS detector [5]. From the outside we have the muon system interspersed by return yokes covering the superconducting solenoid providing the 3.8 T magnetic field, which in turns cover the calorimeter system. The innermost layer of the detector is occupied by the Tracker.

Accurate measurement of the energy was done by making use of the scintillating lead tungstate ($PbWO_4$) crystals that boast the advantages of short radiation length (0.89 cm) and Moliere radius (2.3 cm), resulting in a compact calorimeter with excellent close cluster separation [8]. The Ecal is also separated into regions of EB and EE providing similar $|\eta|$ coverage as the Hcal. The energy resolution of the Ecal is:

$$\frac{\sigma(E)}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\% \tag{2.2}$$

At the front of the endcap region, a preshower detector is installed consisting of two planes of silicon sensors interleaved by lead blocks. This preshower detector serves to reject the close-together diphoton background from the π^0 decay from the single photon signals.

Finally, the innermost part of the detector, covering the beam pipe is the Tracker, responsible for reconstructing the tracks of charged particles and vertex position measurement. It consists of silicon pixel and strip detectors with a coverage of $|\eta| < 2.5$ that record the hits left by charged particles as they pass through each layer of the components. These hits are then fed into the track reconstruction algorithm (discussed in detail in Chapter 3 and Chapter 4) which outputs the tracks with which the primary and secondary vertices positions are determined.

2.3 Trigger System in the CMS Experiment

At the LHC, proton bunches cross each other at the designated interaction points forty million times per second, corresponding to a frequency of 40 MHz. Considering the typical number of inelastic proton collision or as it is more commonly called, an event, in each crossing, this translates to roughly 1 MB of data for every event with O(10^9) Hz event rate. As the experiments could only handle an upward of O(10^3) MB/s of data storage rate each (for 2015 Run 2), they are vastly overwhelmed by the huge rate of the incoming data stream. Thus the necessity to record only a small fraction of these events for later analysis. However, the decision on which events are to be saved has to be made quickly as unsaved events will be overwritten and lost forever as soon as the next event enter the data stream. Therefore the trigger system is expected to perform a selection of the entire data keeping only the most interesting events within a very short time frame compatible with the collision rate.

At the CMS experiment the trigger system is split into two stages; the Level 1 (L1) seeding stage and the High Level Trigger (HLT) stage. This is because the rejection power expected of the trigger system is far too large to be achieved by any single selection step if a high efficiency of events of physics interest is to be maintained. The L1 stage performs a preliminary selection to reduce the event rate before feeding it to the HLT stage, where a more sophisticated selection is performed and where the decision to save the data is made.

2.3.1 Level 1 (L1) Seeding

The L1 system is based on custom-made electronics that performs a fast preselection on the data based on coarse information provided by the calorimeters

and muon system. This stage reduces the event rate to roughly 100 kHz for further processing in the HLT step. Each L1 seed is designed to trigger on common object types: electrons, muons, photons, jets and others, making use of the combined inputs provided by the subdetectors. Track reconstruction is not done at this level due to the more complex reconstruction needed and therefore increased timing, which could not be accommodated within the system. A detailed discussion of the L1 system is available elsewhere and will only be briefly touched upon in later parts of this thesis [9].

2.3.2 High Level Trigger (HLT)

The events that passed the L1 seeding stage are sent to the HLT stage where they are further processed. The HLT is a software-based algorithm running on commercial electronics [10]. Due to the more relaxed interval available to process each event in this stage, the HLT could afford to compute more complex quantities to identify the objects with, making use of more refined information provided by the subdetectors. Object reconstruction at HLT is similar to what is done offline with similar performance, the key difference being that some simplifications are made to the algorithm in order to have a complete decision within the allowed time frame; such as track reconstruction being done only on a region compatible with the L1 seed (as opposed to the entire Tracker for offline). Events passing this stage are then saved into tape for offline reconstruction and further analysis. Detailed studies in Run 1 has shown that the trigger system is performing well, keeping a high signal efficiency while keeping a manageable rate [11].

In regards to HLT, several terms are commonly used and is defined here. The first of them is menu, which refers to a list containing all modules and

paths used in the HLT system. Secondly, a path is a sequence of modules that perform the computation and filtering on a specific (combination of) object(s); for instance the path `HLT_Jet100` is a series of modules that aim to trigger on events containing at least a jet with p_T above 100 GeV. Then, a module refers to a software (or part of one) that performs a specific function; usually computing some object identification variable from some input (which either comes from the L1 seeding stage or previous HLT modules) or filtering on the variables produced by previous modules or less commonly, both. An event is considered to pass the HLT stage if it passes at least a path in the entire HLT menu. However, in some cases such as control paths with loose cuts, where the event rate is too high, a prescale could be assigned to reject some of the passing events. Prescales are assigned in the form of integer N which means only 1/N of the events passing the prescaled path are actually recorded. By default all paths are unprescaled, which means they accept all the events passing them.

HLT path names are of the form `HLT_ObjXX(_YY)`, with “HLT" denoting that the path is part of the larger HLT master table, “Obj" denotes the object type being triggered on, “XX" referring to the minimum p_T threshold of the object and “YY" referring to any additional selection criteria and/or some specific label, such as $|\eta|$ restriction on the candidates or that the path is using some non-standard algorithm, for instance a specialized version of the track builder used in the single electron paths.

2.4 CMS Event Data Model (EDM)

Before a physics analysis could be performed, the data provided by the detector has to be combined to reconstruct the high level objects such as electron or

muons which could then be used in the analysis. At CMS, the processing steps to reconstruct the physics objects are done centrally, leading to a splitting of the data into three tiers with varying level of size and detail [12].

The first data tier is called *RAW*, which contains the raw detector information such as hits in a subdetector element, energy deposits and others. Additionally it also contains the results of L1 and HLT stage processing and possibly the high level quantities calculated in the HLT selection steps. The average size of the *RAW* dataset is 1.5 MB/event. As this project deals primarily with HLT path optimization, this data tier is the one that is used the most.

The second tier, derived from the *RAW* dataset is called *RECO*, after the fact that this data tier contains mainly reconstructed objects. The reconstruction algorithm starts by doing detector-specific processing, where the detector calibration constants are applied and the *RAW* information are unpacked and decoded. Then the cluster and hits within the detectors are reconstructed. This is then followed by the tracking step. Due to the lack of time constraints, it is done over the entire detector, unlike in the HLT step. After the tracks are available the vertices are reconstructed and finally the particles are identified as electrons, muons, jets and others based on their physics signatures by the Particle Flow (PF) algorithm [13]. Most of the information available in the *RAW* dataset are then dropped, keeping only the links to the *RECO* information, reducing the average size to 0.25 MB/event.

Although the *RECO* dataset is much smaller than its *RAW* parent, it is still rather large for full copies of it to be stored in multiple computing centers around the world. Additionally, while more compact, the *RECO* still contains information not commonly used in physics analysis, making its presence in the dataset often unnecessary. As such, another data tier called *AOD* (standing

for Analysis Object Data) is introduced, which is a subset of the *RECO* dataset, keeping only information of the physics objects of interest to physics analyses, such as tracks with associated hits, vertices and high level objects such as electrons and muons, as well as the links to the corresponding *RECO* information. This compression leads to a size of 0.05 MB/event, which is compact enough for the dataset to be fully stored in most centers. For Run 2, due to the increased amount of data to be available due to the increased luminosity and PU, a more compact data format called miniAOD is introduced, which is as small as 0.005 MB/event [14].

2.4.1 Monte Carlo Samples

As the project aims to optimize the single lepton triggers for use in Run 2, for which the dataset has yet to be available, the samples used are primarily simulated using the Monte-Carlo method. The simulation for both signal and background processes was done using the CMS software (CMSSW) version `6_2_5` using the `PYTHIA` [15] generator which provides an unbiased simulation of the pp collisions. The CMS detector was simulated using the `GEANT4` [16] toolkit. PU events are simulated by overlaying the QCD background events on top of the signal events, the number of which increases as a function of luminosity. All the samples used for this study were centrally produced by CMS for use in trigger-related studies, each dataset containing O(1M) events.

Chapter 3

Single Muon Trigger Optimization

"Who ordered that?"

— I. I. Rabi, 1898 – 1988

Being the particle the experiment is named after, the ability to reconstruct muons with a high efficiency is of particular importance to the CMS experiment. The single muon trigger makes use of the muon reconstruction algorithm, designed with this goal in mind, to select events containing one isolated muon, typically the result of a W boson decay. This chapter describes the optimization of the trigger to maximize the efficiency of the recorded top pair events while keeping the rate to a minimum, in preparation for the CMS Run 2.

The specific trigger path studied in this optimization was `HLT_IsoMu24`. As the path name implies, it selects events containing one isolated muon with transverse momentum (p_T) above 24 GeV. It was the inclusive single muon path with the lowest p_T threshold that was unprescaled, which meant that no events passing this trigger were discarded.

3.1 Muon Trigger Overview

Muon reconstruction done in single muon trigger, as with all other muon triggers, can be divided into three levels [17]. The L1 trigger is a hardware-based trigger that provides the initial input to the subsequent software-based reconstruction levels. Using this seed, the muon tracks are reconstructed at level 2 (L2) using the information collected by the muon system. At level 3 (L3), the tracker tracks are reconstructed using the silicon tracker information and matched with the output of L2 reconstruction.

As the single muon trigger is designed to detect isolated muons, it also checks for the isolation of the muon candidates. Within the path, this is done after the final fit of L3 reconstruction, where the tracker isolation is computed using the tracks around the global muon track. It is also possible to check for the calorimetric isolation using the energy deposit in the calorimeter around the muon track, which could be done after the L2 reconstruction is complete. This however is not done in the specific trigger path being studied.

3.2 L2 Reconstruction and Optimization

The L2 muon reconstruction makes use of the information collected by the muon system. It converts the CSC, DT and RPC measurements into seeds to be fit by the track reconstruction algorithm. Duplicate tracks are then cleaned out of the candidate list and the surviving tracks are filtered based on momentum and track quality parameters. A separate track collection is then produced which contains the tracks constraint to the primary vertex.

3.2.1 L1 Seeding

The first step of the muon reconstruction is the production of L1 muon seed, which is an estimate of p_T and global hit position at the second muon station. From this information, an initial seed state, defined as the momentum and direction of the object, is created. The momentum estimation is done under the constraint that its transverse component is compatible the p_T estimated by the L1, while the direction is taken to be the same as the global hit position vector. Although the muon p_T is necessarily underestimated due to energy loss into the detector material before reaching the muon system, it has the advantage of speed and the loose filtering done at this stage ensures that the efficiency is close to 100% with respect to the final muon candidates. Additionally, this bias is corrected by the final fit in L2 and L3 reconstructions, ensuring that the accuracy of the measurements are not compromised in the final output.

3.2.2 L2 Stand-Alone Muon Reconstruction

L2 reconstruction starts by reconstructing the segments and hits inside the individual muon chambers. In the CSCs, this is measured in the form of two-dimensional point, one of which is measured by the wires, while the other is obtained through a Gatti function fit on the charge distribution inside the strips. Up to six such points, one from each CSC plane, are then fitted to build a three-dimensional track segment. On the other hand, in the DTs, the track segment is reconstructed by fitting the hits in individual drift cells. The RPCs produce three-dimensional points instead of track segments but they are also used as input for the track reconstruction algorithm and are collectively referred to as reconstructed hits.

The reconstructed segments and estimated track parameters from the L1 seed are used as input for track reconstruction algorithm, which is based on the Kalman Filter (KF) technique [18, 19]. Starting with the seed-estimated track parameters propagated to the innermost reachable muon system layer, the next compatible layer is searched for by fitting the track segment and selecting on χ^2 basis. This is iterated to search for the next in an outward direction, with only the measurements with incremental χ^2 less than 1000 being considered. A final cut of χ^2 less than 25 is then imposed to determine if the track parameters are to be updated with the information from the best measurement in the fit. The track reconstruction is then done again in the opposite direction, taking as input the track parameters obtained in the previous step. The iterative χ^2 cut in this step is 100 and the track parameters are updated with each measurement if it also passes an additional cut of χ^2 less than 25. The fitting-smoothing step is iterated on the newest available track a number of times (3 in the default configuration) to remove the possible biases coming from the lack of rescaling of errors between the forward and backward fitting or the seed parameters.

The output of the fitting step is a collection of tracks spanning the muon system. As each track is fitted independently, it is possible that the same hits are assigned to multiple tracks. The duplicate-cleaning procedure is performed using this criteria:

- The track with more hits is kept if the hit difference between two tracks is larger than 4
- If more than 95% of the hits are shared, the track with higher p_T is kept if its p_T is higher than 7 GeV and the other lower than 3.5 GeV
- In all other cases, the track with smaller normalized χ^2 is kept

Finally, another track collection is produced by copying from the current collection. The track parameters of this collection are then constrained by the beam spot position. The position errors used in this constrained are 5.3cm in the z-direction and 0.1cm in the $x - y$ plane, a deliberate overestimation of the design uncertainty as to have a looser cut. Tracks failing the constraint are removed from this collection and both collections are output to the next reconstruction step.

3.2.3 L2 Parameter Optimization

The information used in the L2 reconstruction provides us with a number of variables to filter the muon candidates with. The ones defined within the `HLTMuonL2PreFilter` module, along with the type of cut (upper or lower bound) and their default values, to be used by the trigger are:

- `MinNstations`: Number of muon stations that registered at least a hit, (0, 2, 0)
- `MinNhits`: Number of valid hits within the muon system, (0, 1, 0)
- `MinNchambers`: Number of CSC or DT chambers that registered at least a hit, (0, 0, 0)
- `MaxDz`: Longitudinal distance of the muon candidate to the beam spot in cm (9999.0)
- `MaxDr`: Transverse distance of the muon candidate to the beam spot in cm (9999.0)
- `MinDxySig`: Significance (ratio of the uncertainty and the measured value) of the transverse distance of the muon candidate from the beam spot (-1.0)

- `MinPt`: Transverse momentum of the muon candidate, whose default value depends on the main objective of the muon trigger in question (16.0)

Table 3.1 summarizes the setup for the study. Note that the background was not estimated from the data as indicated for the L2 parameter optimization, for reasons to be explained later. The optimization was done in N - 1 manner, meaning that the effects of tightening one variable is investigated on the signal efficiency, with all others being kept constant. This was done using the OpenHLT tool, which provided direct access to the filter parameters to be varied [20]. Only the geometry-dependent variables are looked at in this study, with the contributions from other regions subtracted out. The results are summarized in Figure 3.1.

In the default configuration of the path, no variables are tightly filtered on tightly. Additionally, since the first three variables are dependent on detector geometry, the default values are set based on η regions. Their upper bounds are given as $|\eta| = (0.9, 1.5, 2.1)$ and for the purposes of this section, these regions will be identified as Barrel, Transition and Endcap respectively.

From the `MinNstations` graph, it could be seen that the efficiency drops significantly if hits from more than 2 stations are required to register a hit for a muon candidate to be accepted. This applies in particular to the Transition

Table 3.1: Setup and samples for the single muon trigger optimization.

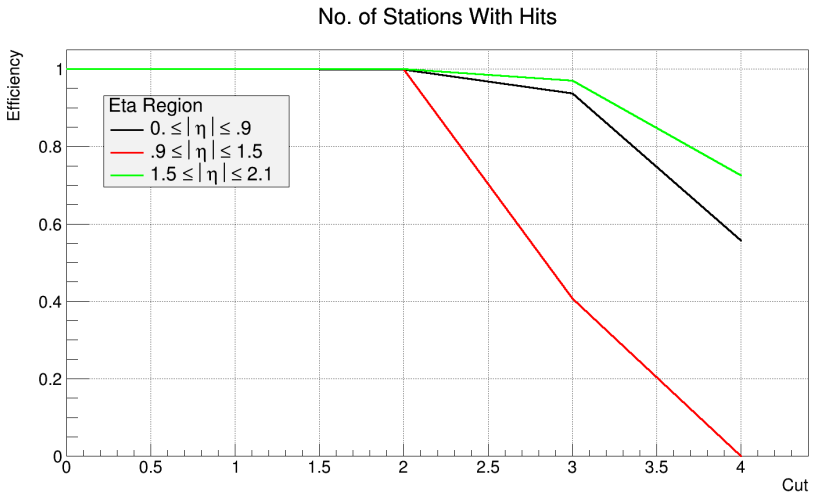
Setup	Information
CMSSW	`CMSSW_7_0_0_pre13`
Menu	`/dev/CMSSW_7_0_0/HLT/V91`
Signal MC	13 TeV `Fall13` $t\bar{t} \rightarrow \mu + 4j, \lvert\eta_\mu\rvert \leq 2.1$
Data	HLT Physics Parked Run 207884 Lumi Section 2 - 106, 108 - 182

region, where due to detector geometry, the tracks tend to miss one or more stations. It is also for the same reason that a default cut is imposed only in this region, as quality of the fit would be compromised if the hit information is provided only by one station. While by default there is no cut imposed on the Barrel and Endcap regions, we observe the same trend due to the implicit requirements imposed by the reconstruction steps. As such it was decided that no modification to this cut should be made, for it was already at the optimal point.

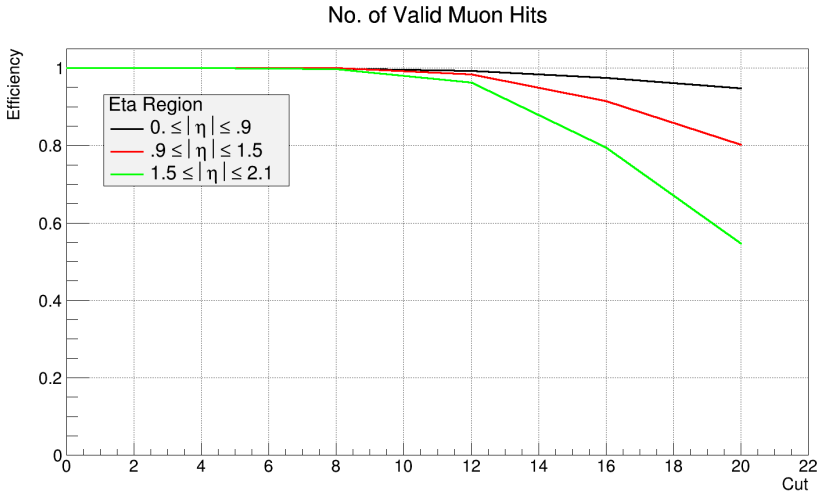
The `MinNhits` and `MinNchambers` graphs showed a similar trend. While the cuts imposed by the default filters were very loose, they were implicitly imposed within the reconstruction step itself. The minimum number of hits are restricted by the fact that only tracks with a number of hits above a certain threshold (which was observed to be 8) would pass the χ^2 requirement of the final fit. The number of CSC or DT chambers registering a hit could not be less than one as the reconstruction algorithm by design accepts only candidates with at least two measurements, one of which coming from either CSC or DT. In both variables tightening the cut reduced the efficiency significantly, hinting at the fact that implicit cuts imposed in the reconstruction steps were already sufficient in dealing with the background. No gain was expected from varying the parameters without a large efficiency loss and therefore the background was not estimated.

3.3 L3 Reconstruction and Optimization

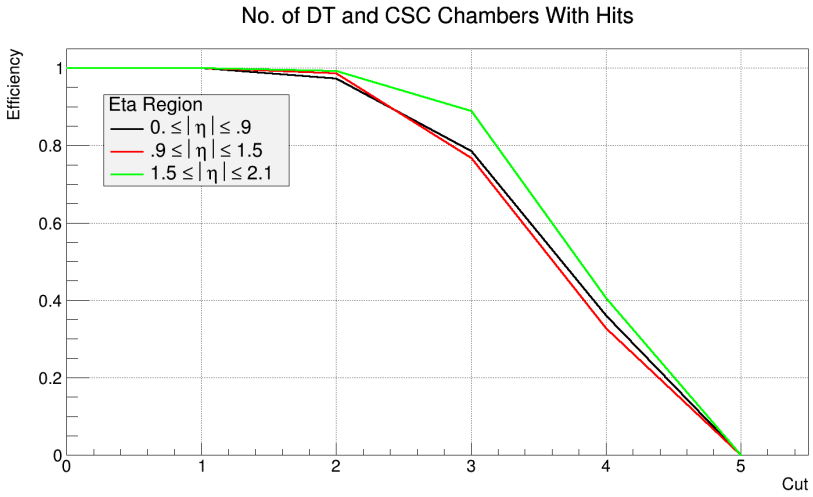
As multiple scattering effects dominate the momentum resolution of L2 reconstruction [21], it is necessary to improve it by combining with the information



(a) MinNstations



(b) MinNhit



(c) MinNchamb

Figure 3.1: Signal efficiency of the signal muon trigger as functions of the variables used in L2 reconstruction filtering.

obtained from the silicon tracker. However, since the full tracker reconstruction is very resource intensive, only a small region is reconstructed at the HLT, selected based on compatibility with the L2 muon found in the previous stage. The L3 reconstruction proceeds in a similar flow as the L2, as they are based on the same algorithm, using silicon tracker information and its matching to the L2 muon candidates.

3.3.1 Cascade Seeding Algorithm

Using the output of L2 reconstruction, two types of seed are produced for use in the L3 reconstruction. The state-based seed is produced by propagating the L2 muon candidate to the outer layer of the tracker after rescaling its error to find a compatible tracker module. The estimated state of this module is then used to create the seed.

On the other hand, the hit-based seed is produced by combining the hits in the tracker layers to estimate its position and direction. This is done in two directions; outside-in and inside-out. The outside-in seeding starts from a region in the outer tracker layer compatible with the propagated L2 muon candidate. Compatible inner layers are then searched for and updated using the Kalman filtering algorithm, similar to what is done in L2 reconstruction. Good seeds are then selected using the constraint to the beam spot position. The inside-out seeding is done in the opposite way. After defining a tracker region around the L2 muon candidate, pixel hit pairs and triplets are searched for inside it starting from the innermost tracker layer. These hits are then fit together to produce the seed and it is kept if compatible with the L2 muon candidate.

The L3 reconstruction algorithm does not produce seeds of all types for each candidate. In order to minimize timing, the fastest seeding algorithm, the outside-in state-based, is run first. The reconstruction algorithm will try to reconstruct the muon using this seed. If the reconstruction is successful, the other algorithms are not run. If the reconstruction could not be done with this seed, the second type is used, the inside-out state-based and only when this seed also fails will the slowest seeding algorithm, the inside-out hit-based is run. Due to the cascading structure of the seeding process, this algorithm is called the Cascade algorithm.

3.3.2 L3 Global Muon Reconstruction

As mentioned earlier, the track reconstruction in the tracker is done in a flow similar as the L2 reconstruction. After the forward (defined as the direction extending away from the seed) fitting of the hits, a second iteration is done backward. While the same algorithm is used also in the offline reconstruction [21], at HLT only one hit is selected per tracker layer in order to reduce timing. Unlike the L2 reconstruction however, the error matrix of the forward fit is rescaled by a factor of 100 before being fed into the backward fit. Additionally, due to the increased availability of seed types from the Cascade algorithm, the forward and backward fits are done both in inside-out and outside-in direction, using the appropriate seed. A cleaning procedure is then performed to ensure that the tracks do not contain duplicated hits.

Although the track reconstruction at HLT is done only on a tracker region compatible with the L2 muon candidate, usually there are still multiple tracker tracks being compatible with the muon candidate. In order to select the best one to be combined into a global muon, a track matching procedure is performed

on the collection of tracks based on their relative momentum and position with respect to the L2 muon candidate. A final track spanning the entire CMS detector is then built by fitting the tracker and muon tracks together, giving rise to a global muon. A filter is then applied to ensure that the quality of the candidate is consistent with what would be expected of a signal muon.

3.3.3 L3 Parameter Optimization

The variables used in the filtering step reflect the fact that the L3 reconstruction step uses both the tracker and muon system information. The `HLTMuonL3PreFilter` module contains the definition of the variables and the cut to be performed. They are:

- `MaxNormalizedChi2`: The normalized χ^2 of the final global muon fit (20.0)
- `MinNhits`: Number of valid tracker hits (0)
- `MaxDXYBeamSpot`: Transverse distance of the global muon candidate to the beam spot in cm (0.1)
- `MaxDr`: Impact parameter of the global muon candidate (2.0)
- `MinDxySig`: Significance of the transverse distance of the global muon candidate from the beam spot (-1.0)
- `MaxDz`: Longitudinal distance of the global muon candidate to the beam spot in cm (9999.0)
- `MaxPtDifference`: Difference in p_T measured by the muon system and silicon tracker (9999.0)
- `MinNmuonhits`: Number of valid hits in muon chamber (0)

The L3 parameter optimization was done in N - 1 manner using the OpenHLT tool, just like the L2 parameter optimization. The setup used was also the same, as summarized in Table 3.1. The background was estimated from the 'parked' data, which is a portion of data recorded with minimal triggering requirements [22]. In both cases the efficiency is plotted as functions of the variables in order to study the variation of signal and background. The results are summarized in Figure 3.2 and Figure 3.3.

As could be seen in Figure 3.2, varying the cut points had a very small effect on signal and background efficiencies. For `MaxNormalizedChi2`, this is understood as being the effect of the track matching step of the L3 reconstruction, where the tracker track that best matched the L2 muon candidate is chosen on χ^2 basis. A similar argument is used to explain the distribution of `MinNhits`, as the best fits tend to be from the tracks with higher number of hits. `MaxDXYBeamSpot` and `MaxDr` on the other hand were understood by looking from the physical perspective; isolated muons are produced only in interactions involving heavy particles such as the vector boson or the top quark. The low values of impact parameters and the transverse distance from the beam spot are natural considering the high energies and short lifetimes of these interactions. This study did not choose to tighten the cuts harder than the studied range as these cuts are also dependent on detector alignment and other conditions which are rarely ideal, therefore cuts that are too tight are not desirable at trigger level.

The other half of the variables as shown in Figure 3.3 told a different story. These variables, due to them being not or differently influenced by the physics behind the interactions that produce the muons, are more affected by the variations in the cut thresholds. The variation of signal and background efficiency

as a function of `MinDxySig` are roughly similar, which meant that there is no gain in tightening the cut. Similar trends could also be seen in `MaxDz` and `MaxPtDifference`, as these variables do not carry much information on the characteristic of the event. The `MinNmuonhits` could be optimized, owing to the fact that the background contain punch-through kaons or pions, or real muons resulting from the decays-in-flight of these particles. The optimal threshold of 14 was proposed, which offered around 10% background suppression at the cost of less than 5% signal.

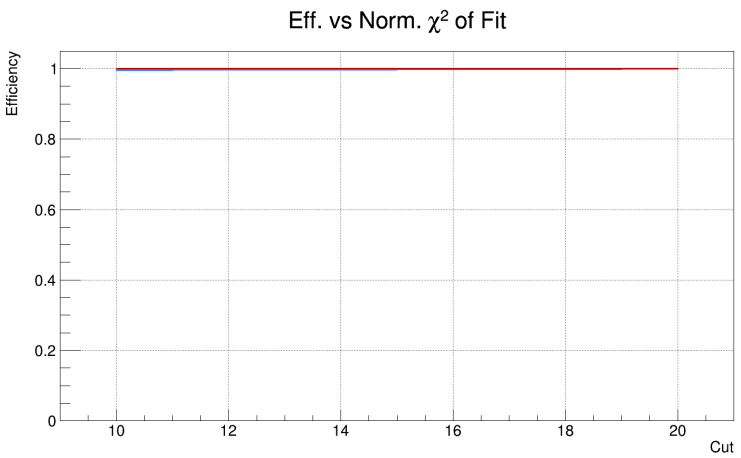
3.4 L3 Isolation Optimization

Isolation is a measure of activity, which is usually defined as the sum of p_T or energy deposit, around the object of interest. The version that was studied was the detector-based relative isolation (default cut 0.15), which is defined as:

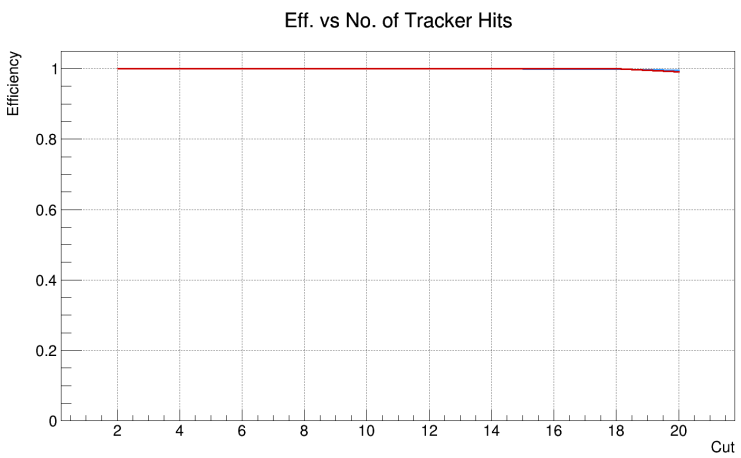
$$\text{RelIso}_{\text{Det}} = \frac{\Sigma p_T(\text{Trk}) + max(0., \Sigma E_T(\text{CaloTowers}) - \text{EffArea} \cdot \langle\rho\rangle}{p_T(\mu)}, \quad (3.1)$$

$$\text{EffArea} = k/a,\ \langle\rho\rangle\,(N_{vtx}) = aN_{vtx} + b, \langle\text{Iso}\rangle\,(N_{vtx}) = kN_{vtx} + j \quad (3.2)$$

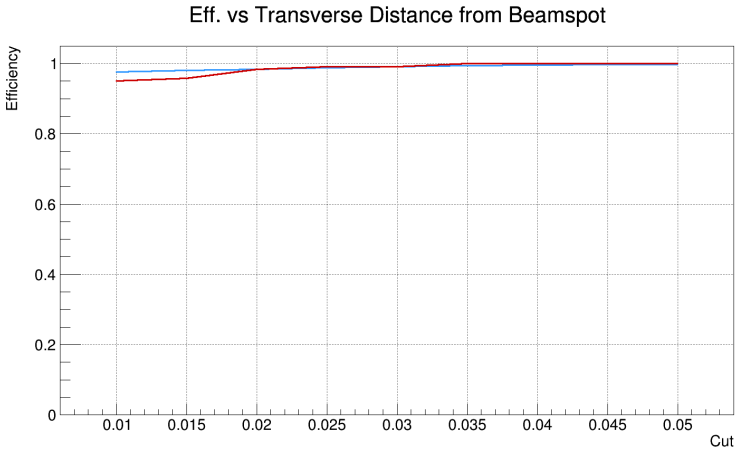
Due to the updates in single muon paths, the study was conducted on the an updated version of the path which made use of an iterative tracking algorithm during the reconstruction steps, using `CMSSW_7_1_0_pre9`. The samples used for this study were the same ones used for the previous studies. The optimal threshold of 0.12 was proposed, which provided a 5.4% background suppression at a less than 1% signal cost. The efficiencies over the entire studied range was shown in Figure 3.4.



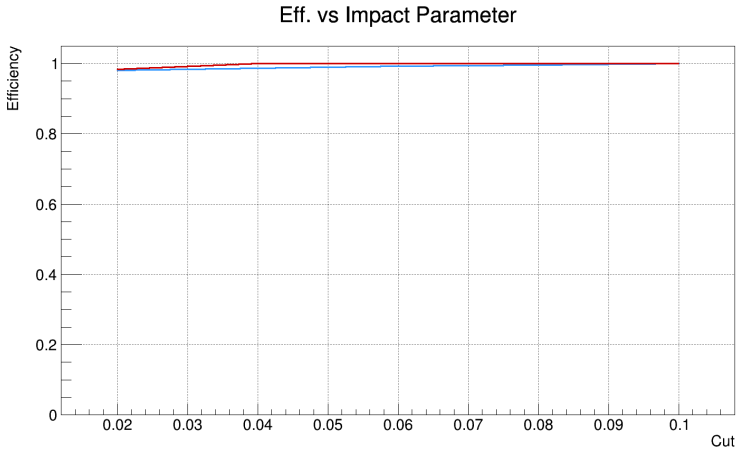
(a) MaxNormalizedChi2



(b) MinNhits

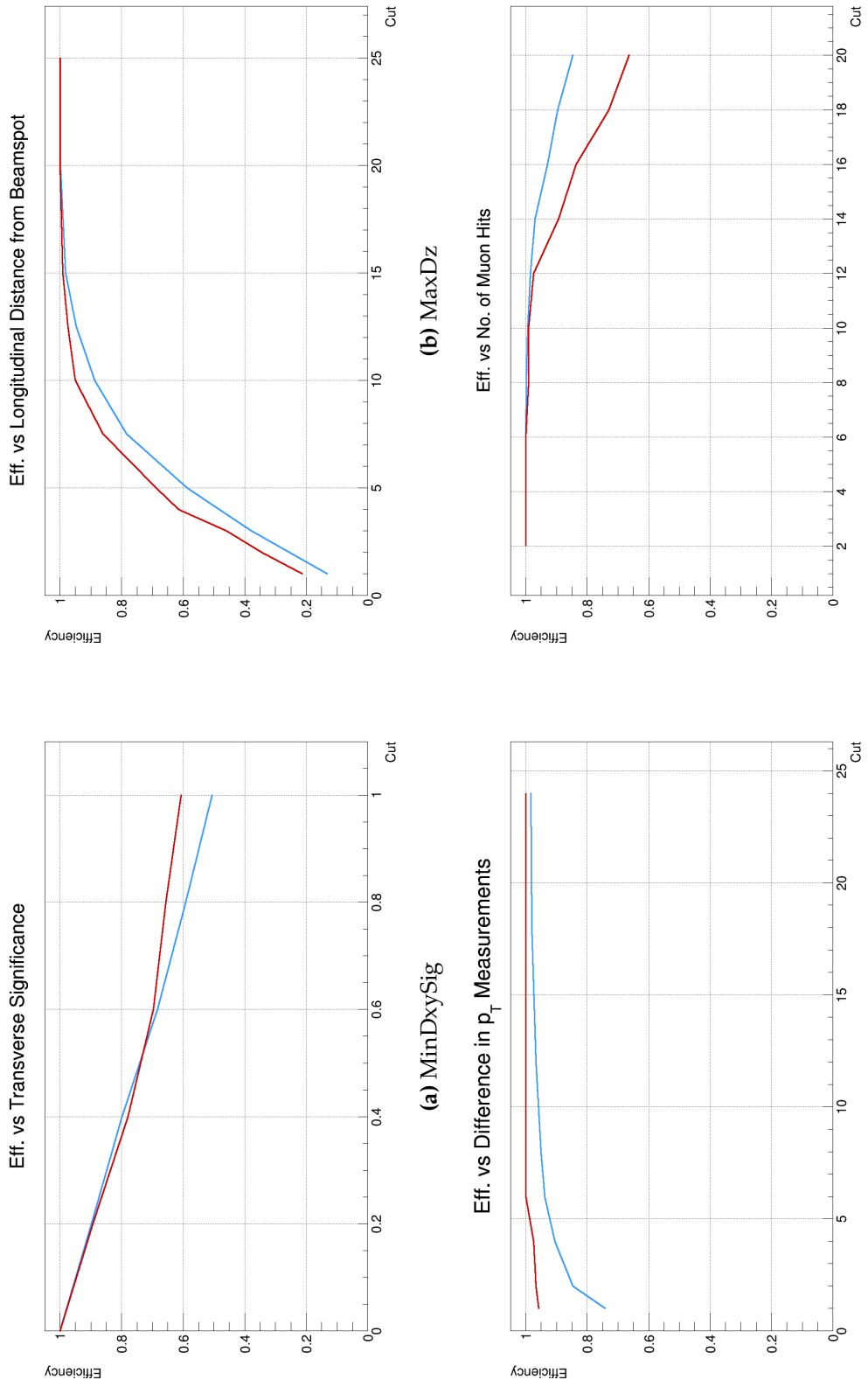


(c) MaxDXYBeamSpot



(d) MaxDr

Figure 3.2: First four of efficiency vs. cut value graphs for L3 filter parameters. The blue line represents the signal distribution while the red line represents the estimated background distribution.



(a) MinDxySig **(b)** MaxDz **(c)** MaxptDifference **(d)** MinNmuonhits

Figure 3.3: Second four of efficiency vs. cut value graphs for L3 filter parameters. The blue line represents the signal distribution while the red line represents the estimated background distribution.

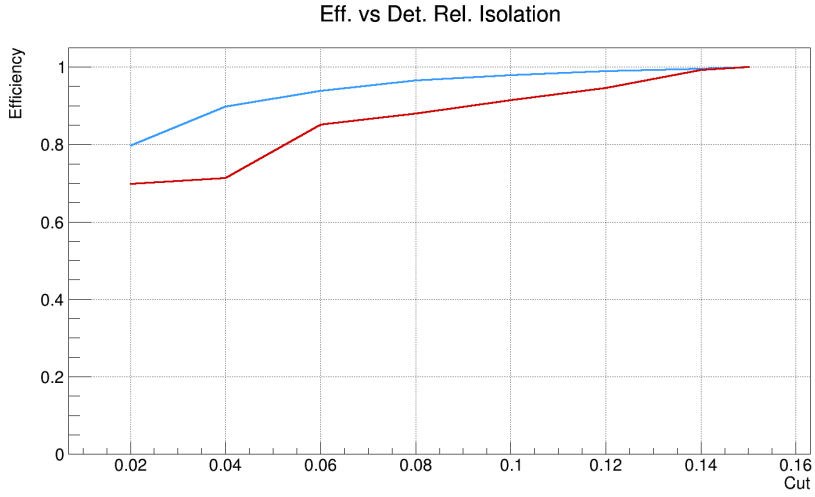


Figure 3.4: Efficiency vs. detector-based relative isolation. The blue line represents the signal distribution while the red line represents the estimated background distribution. An optimized cut of 0.12 was proposed which provided a 10% background rejection at 2% signal loss.

Chapter 4

Single Electron Trigger Optimization

"There is one simplification at least. Electrons behave…in exactly the same way as photons; they are both screwy, but in exactly in the same way…"

— R. P. Feynman, 1918 – 1988

As the only other lepton that is directly detectable in the detector, electrons play almost as major a role as muons to CMS physics program. This is because leptons provide a handle to discriminate the few events of physics interest from the overwhelming number of QCD events produced in the pp collisions. Unlike muons however, electrons are not as easy to reconstruct, leading to a more complex reconstruction algorithm.

This chapter describes the optimization of the single electron trigger in preparation for CMS Run 2. Just like its muon sibling, this trigger is designed

to record events involving the electronic decay of the W boson. Additional studies to control the rates and in some cases, improve the acceptance of this trigger are also described.

4.1 Electron Trigger Overview

Electron reconstruction in the trigger paths are divided into two level; L1 seeding and HLT reconstruction. As with muon trigger the L1 seeding step is hardware-based, taking the energy deposit within the calorimeter as the initial input to start off the reconstruction chain. This will then be sent to the HLT reconstruction step for more elaborate quantities to be computed, that the electron candidates may be reconstructed and identified. One thing worth mentioning here, although it will not be discussed further in the chapter, is that due to their similar footprints inside the detector, electrons and photons are reconstructed using largely similar algorithms, with the only crucial difference being that for photons no associated track is reconstructed inside the Tracker due to it being uncharged.

4.2 Electron Reconstruction at HLT

The HLT reconstruction step starts by clustering the Ecal crystals referenced in the seeding step into a group of crystals called a supercluster [23]. Following this step, the Hcal tower directly behind the supercluster is made from the energy deposit into the Hcal. After that, for paths that require the electron candidate to be isolated such as the single electron path, the isolation of electron candidate is computed separately in Ecal and Hcal. Finally the tracks associated

with the electron candidates are built using a dedicated algorithm called the Gaussian Sum Filter (GSF) algorithm.

4.2.1 Ecal Clustering and Hcal Tower Creation

In the seeding step, crystals that registered energy deposits are recorded and passed to the clustering step. These crystals are then grouped together into a so-called supercluster, centered around the crystal with the highest deposit. As electrons readily radiates photons away in the form of bremsstrahlung, clustering is necessary to fully capture their initial energy. The clustering algorithm used depends on the electron candidate's η, with the "Hybrid" algorithm used for the barrel region ($|\eta| < 1.4791$) and the "Multi5 $\times$ 5" algorithm endcap ($|\eta| > 1.4791$). In summary, the Hybrid algorithm groups crystals within a $\Delta\phi < 0.3$ rad window around the seed crystal in a domino fashion, while the Multi5 $\times$ 5 algorithm does so by collecting the energy deposit in 5 $\times$ 5 crystal matrices and grouping those within the same $\Delta\phi$ window as in the barrel case into a supercluster.

In the updated versions of the trigger, a different clustering algorithm aiming to reconstruct the individual particle showers are used instead of the algorithms described above. As this clustering algorithm is part of the full PF reconstruction algorithm [13], it is referred to as the PF clustering algorithm. The clustering is done by grouping together around a seed crystal all neighboring ones that registered energy deposit at least 2σ above the electronic noise threshold, which is taken to be 0.23 GeV for barrel and 0.6 GeV (or 0.15 GeV transverse energy) for endcap. Although this algorithm offers no boost in identification performance as compared to the old one [24], it allows for a neater way of computing the isolation sum to be described in the next section,

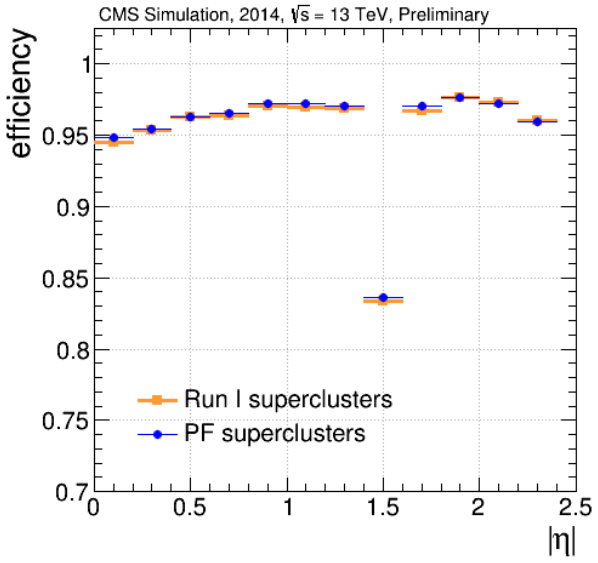
on top of making it possible to share the energy of one crystal between multiple clusters. Additionally, it offers significant improvements in energy resolution, as shown in Figure 4.1.

After the clustering step, the Hcal tower directly behind the supercluster is reconstructed. The energy deposit into the Hcal provides another variable with which an electron can be identified, which will be described in more detail in Section 4.3.

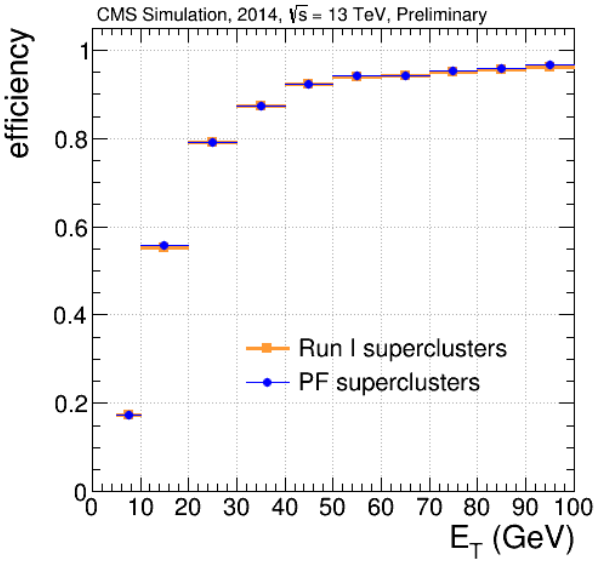
4.2.2 Track Reconstruction

Following the supercluster creation, tracks within the tracker region are reconstructed and a compatible one is associated to the supercluster as the electron track. Unlike the standard track reconstruction which uses the standard KF technique, electron tracks are reconstructed using the Gaussian Sum Filter (GSF) algorithm due to the former being inadequate to accurately approximate their highly non-Gaussian energy loss behavior [18, 25]. The higher accuracy is achieved by approximating the energy loss using a weighted combination of multiple trajectory components with their helix parameters having Gaussian uncertainties, which leads to improved momentum and angular resolutions, as shown in Figure 4.2 [24]. The shapes of the distributions are not affected as they are determined by the underlying physics: energy losses leads to the selected track having lower momentum compared to the simulated electron momentum, leading to a skew-symmetric distribution while the symmetry in the angular resolution is due to the fact that the detector is isotropic in ϕ.

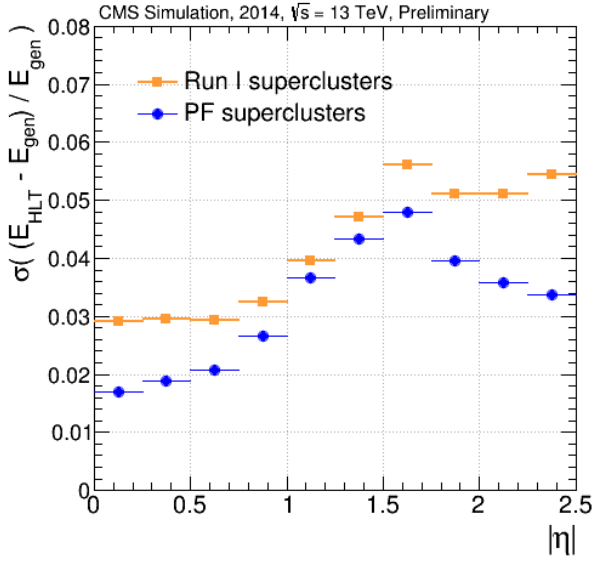
Within the single electron path, the track reconstruction is done only in tracker regions compatible with the supercluster. This significantly reduces



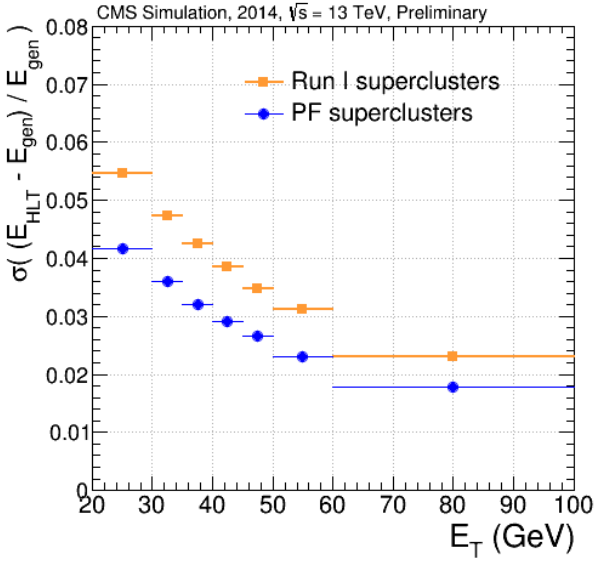
(a) Efficiency in η



(b) Efficiency in E_T



(c) Energy resolution in η



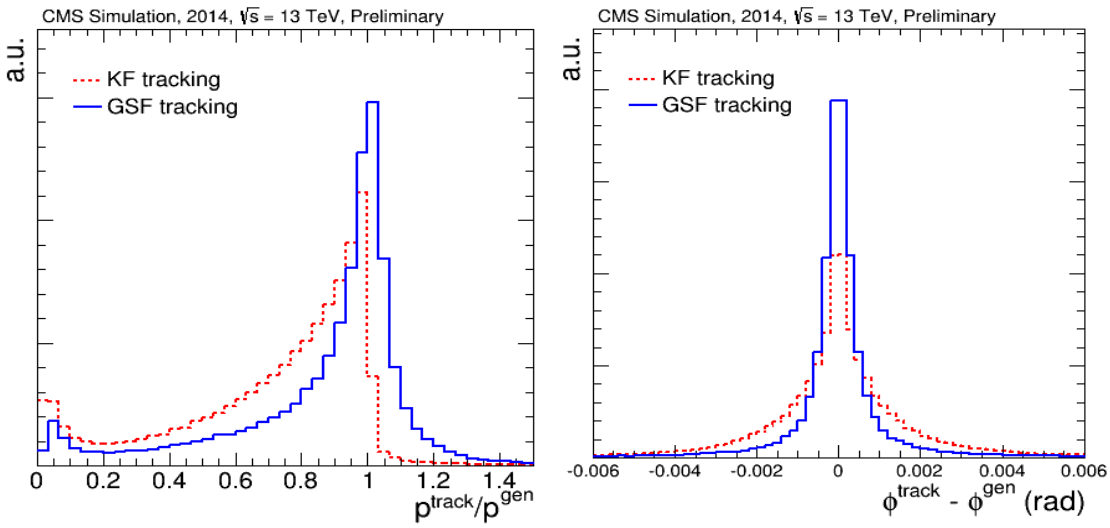
(d) Energy resolution in E_T

Figure 4.1: Comparison between Run 1 and Run 2 clustering algorithms. Note that while the PF clustering algorithm maintains the same efficiency as the one used in Run 1, it offers a significant improvement in energy resolution and other aspects of the electron reconstruction algorithm, making it the algorithm of choice in Run 2.

the timing of the path, as track reconstruction, in particular the slower GSF algorithm, is very resource intensive. Besides, this does not cause a significant drop in performance as the identification filters applied prior to the track reconstruction ensures that most of the surviving candidates are prompt electrons, which must have a track pointing toward the supercluster.

4.3 Optimization of Single Electron Identification

Optimization of the single electron trigger was done in a largely similar manner as single muon trigger; by minimizing the background efficiency (and therefore the rate, of which background processes are the main contribution) for a given signal efficiency. However, as the electron trigger is more complex, the OpenHLT tool was found to be rather slow for the task, as using it requires running the trigger modules over simulated events for every varied cut point.



(a) Momentum resolution

(b) Angular resolution

Figure 4.2: Comparison of momentum and angular resolution of the two track reconstruction algorithms

Instead of starting from the default cut points as implemented in Run I, the electron identification phase space was fully opened such that the signal efficiency is 100%. Distributions of electron identification variables are then drawn separately for signal and background. The cut points are then set according to the usual optimization procedure; minimizing the background for a given signal efficiency. This is done following the order of the identification filters within the single electron trigger, which was chosen to minimize the timing of the path. Table 4.1 summarizes the setup of the study.

Since the main focus of this study is the efficiencies and not the event count, both signal and background distributions are normalized to unit area so that the shape comparison between the two can be straightforwardly interpreted in terms of efficiencies. As a general rule, for the histograms and graphs to be shown in the sections to follow, the color blue is used to denote the signal and the color red is used to denote the background.

It is worth noting that in optimizing the identification cuts, there is another concern that needs to be taken into account. As the identification variables are computed using the detector input, they are sensitive to detector alignment issues, which in the real case rarely, if ever, ideal. It is for this reason that the concept of 'cut safety' is introduced, which alludes to the idea that the cut

Table 4.1: Setup and samples for the single electron trigger optimization.

Setup	Information
CMSSW	`CMSSW_7_2_1_patch2`
Menu	`/dev/CMSSW_7_2_1/HLT/V113`
Path	`HLT_Ele32_eta2p1_Gsf`
Signal MC	13 TeV `Fall13` $W \rightarrow e\nu, p_T \geq 30GeV, \lvert\eta_e\rvert \leq 2.1$
Background MC	13 TeV `Fall13` Dijet QCD $\hat{p}_T$ bins 30 - 170 GeV

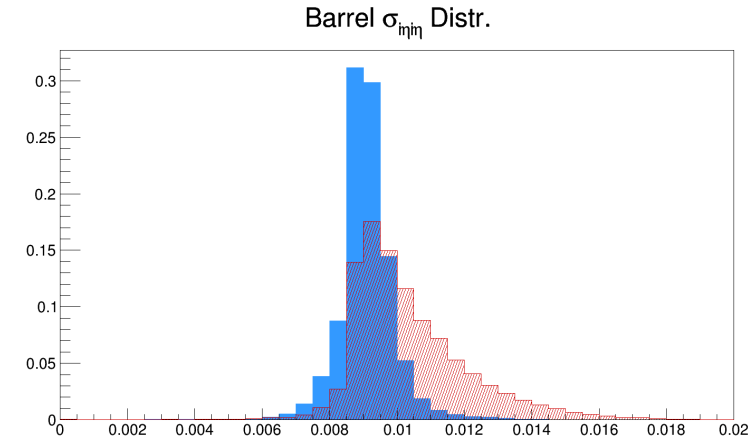
should be within the 'plateau' of the signal efficiency curve, i. e., it should be within the region where the signal efficiency slowly levels off to unity. By doing so one ensures that signal efficiencies are stable against shifts in variable distributions due to alignment adjustments throughout the data-taking, thus minimizing the systematic uncertainties due to trigger efficiency fluctuations.

4.3.1 Cluster Shape: $\sigma_{i\eta i\eta}$

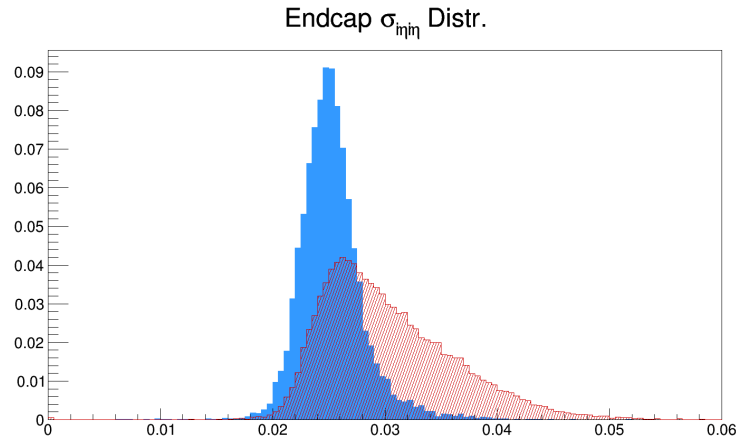
From the supercluster, an identification variable called the cluster shape is computed, which is the weighted η width of the supercluster centering on the crystal with the highest energy deposit, $\sigma_{i\eta i\eta}$, and is given by:

$$\sigma^2_{i\eta i\eta} = \frac{\sum_i^{5\times 5} w_i (0.0175 n_i + \eta_{seed} - \bar{\eta}_{5\times 5})^2}{\sum_i^{5\times 5} w_i} \tag{4.1}$$

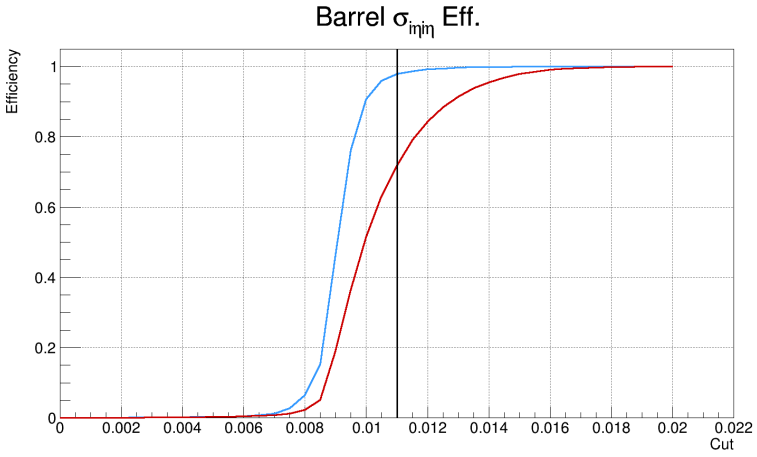
The rejection power of this variable comes from the fact that electron energy deposit is typically narrow; it appears as a focused shower of light within the supercluster. As $\sigma_{i\eta i\eta}$ is a measure of how spread out the energy deposit is within the supercluster, it is a powerful handle to discriminate between electrons and other types of deposit such as hadronic particles within a jet. Figure 4.3 shows the signal and background distributions of the cluster shape variable, in barrel ($\eta < 1.479$) and endcap ($1.479 < \eta < 2.1$) regions. Also shown are the efficiency vs $\sigma_{i\eta i\eta}$ graphs in the two regions, from which the optimized cut points are set.



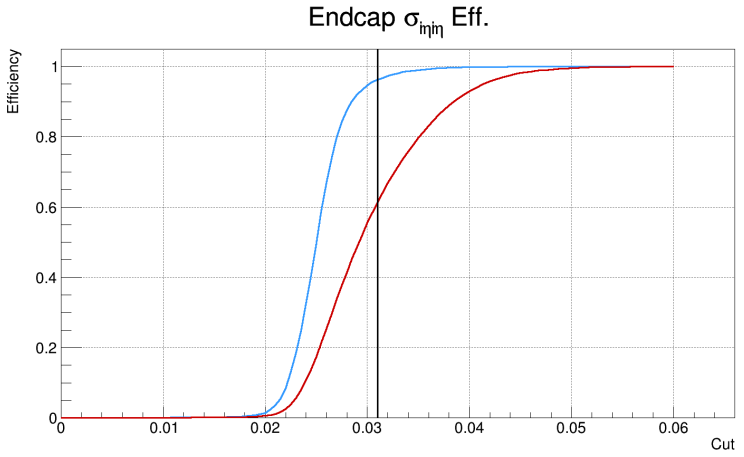
(a) Barrel $\sigma_{i\eta i\eta}$ distribution



(b) Endcap $\sigma_{i\eta i\eta}$ distribution



(c) Barrel $\sigma_{i\eta i\eta}$ efficiency



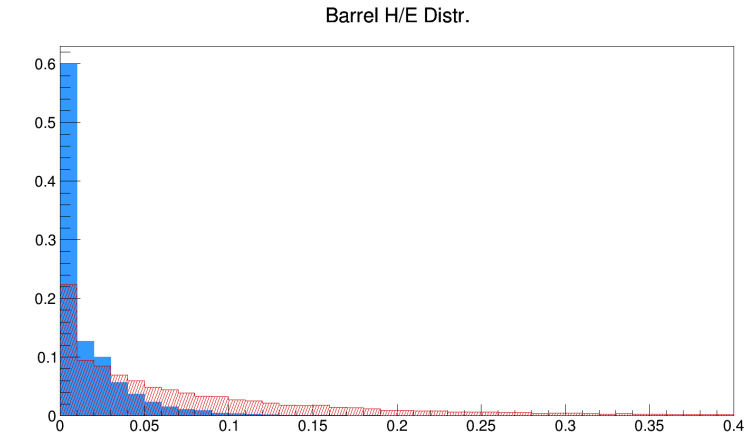
(d) Endcap $\sigma_{i\eta i\eta}$ efficiency

Figure 4.3: Signal (blue) and background (red) distribution and efficiencies of the cluster shape variable, $\sigma_{i\eta i\eta}$. For the barrel region, the optimized cut point was chosen to be 0.011, which provided a 28.1% background rejection at 97.9% signal efficiency. For the endcap region, the optimized cut point was chosen to be 0.031, which provided a 38.8% background rejection at 96.3% signal efficiency. The vertical lines in the efficiency graphs denote the optimized cut point and the corresponding signal and background efficiencies.

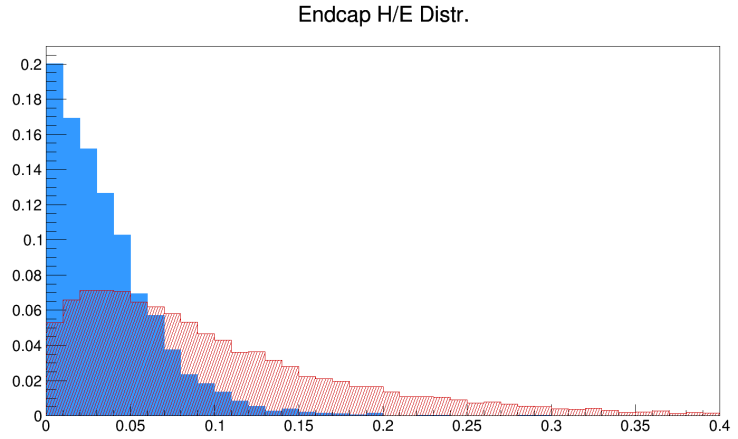
4.3.2 Hadronic Energy Variables: H/E and H - 0.01E

Due to its small mass, bremsstrahlung radiation is a significant channel through which an electron travelling through detector material can lose its energy, more so than the usual ionization mode shared by other commonly detected particles. It is for this reason that electron energy deposits are usually fully contained in the Ecal, a fact that can be exploited to provide us other variables to identify them with; using the deposit in the Hcal tower directly behind the supercluster. The first variable defined to exploit the fact that true electrons are expected to have only a small energy leak into the Hcal is H/E, which is a ratio between the Hcal energy deposit behind the supercluster, H and the supercluster energy, E, which was used in Run I single electron trigger. Figure 4.4 shows the distributions and efficiencies of the variable.

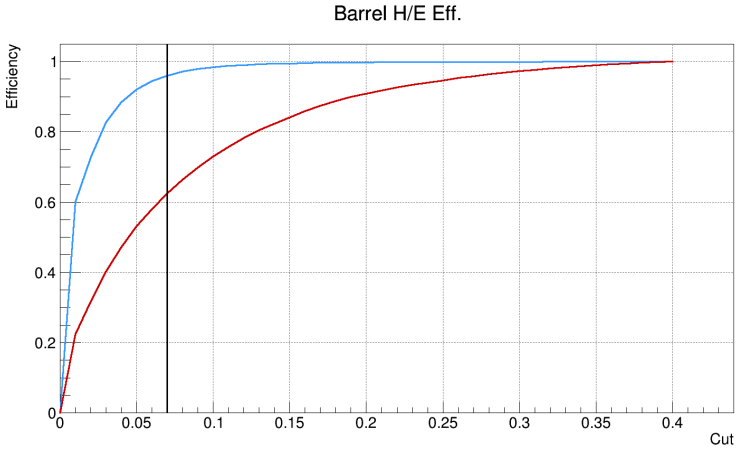
In principle, any combination of H and E that highlights the fact that electron energy leaking into the Hcal should be small can serve as a discriminatory variable in the same way H/E does. Therefore, it is worthwhile to explore different combinations and their rejection powers. One combination that was found to perform better was H - 0.01E. To explain the superiority of this combination, it is worth noting the fact that there are two main contributions to the H term; the noise in the event and from the electron energy itself. While the first contribution averages to a constant between events, the second one is proportional to the electron energy; highly energetic electrons are more probable to 'punch-through' the Ecal into the Hcal. The cut is then set to separately account for these contributions in a linear form, with the factor 0.01 chosen to maximize the rejection power in the energy range relevant for signal and background processes studied. Figure 4.5 shows the distributions and efficiencies of the variable.



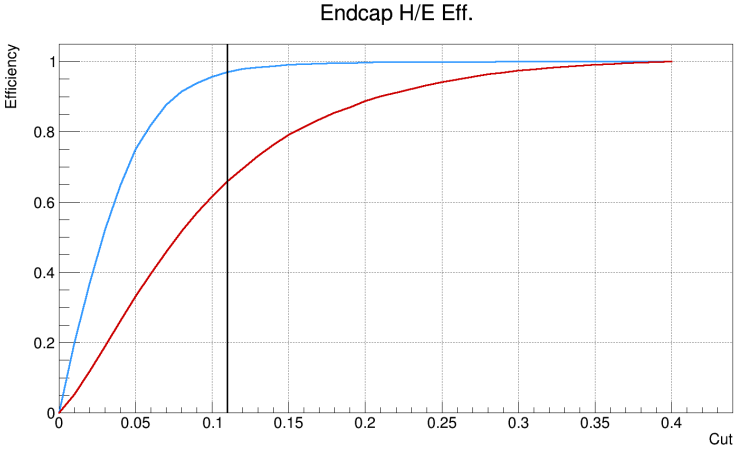
(a) Barrel H/E distribution



(b) Endcap H/E distribution

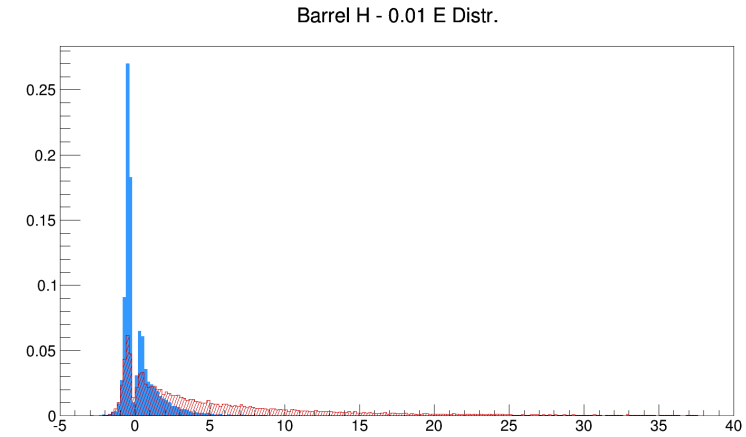


(c) Barrel H/E efficiency

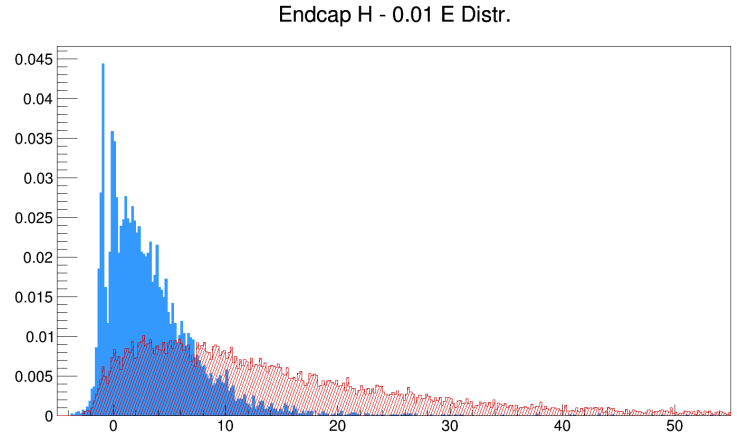


(d) Endcap H/E efficiency

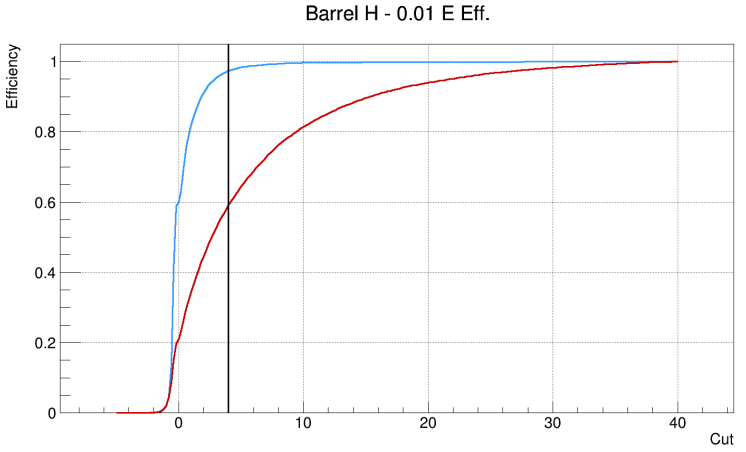
Figure 4.4: Signal (blue) and background (red) distribution and efficiencies of the ratio between hadronic and electromagnetic energies, H/E. For the barrel region, the optimized cut was chosen to be 0.07, which provided a 37.5% background rejection at 96.0% signal efficiency. For the endcap region, the optimized cut was chosen to be 0.11, which provided a 34.1% background rejection at 97.0% signal efficiency. The vertical lines in the efficiency graphs denote the optimized cut point and the corresponding signal and background efficiencies.



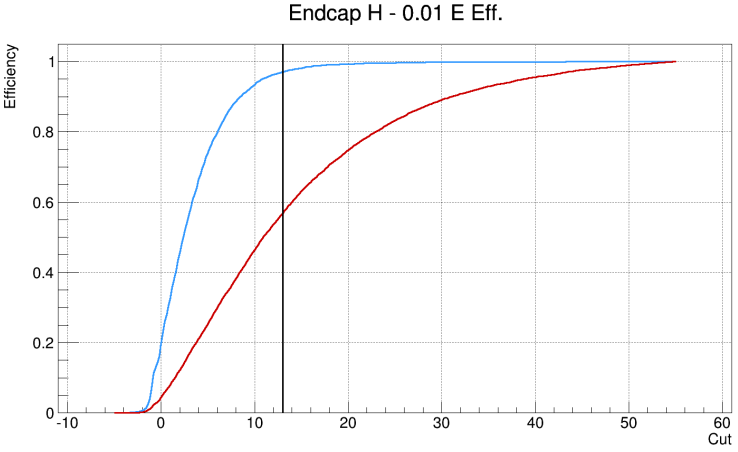
(a) Barrel H - 0.01E distribution



(b) Endcap H - 0.01E distribution



(c) Barrel H - 0.01E efficiency



(d) Endcap H - 0.01E efficiency

Figure 4.5: Signal (blue) and background (red) distribution and efficiencies of the H - 0.01E variable. For the barrel region, the optimized cut was chosen to be 4.0 GeV, which provided a 40.9% background rejection at 97.3% signal efficiency. For the endcap region, the optimized cut was chosen to be 13.0 GeV, which provided a 43.2% background rejection at 97.0% signal efficiency. The vertical lines in the efficiency graphs denote the optimized cut point and the corresponding signal and background efficiencies.

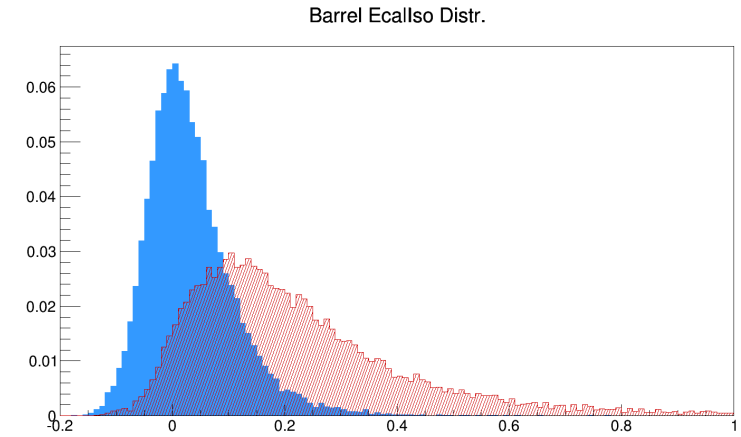
4.3.3 Relative Calorimeter Isolation: EcalIso and HcalIso

As discussed in the previous chapter, isolation is a powerful tool to discriminate between the prompt and background electrons. In the single electron path, instead of computing the combined isolation using the input of all relevant detector components as in the single muon trigger, the isolation in Ecal, Hcal and tracker are computed and filtered on separately. The algorithm used in the single electron is the PF cluster based isolation, which means that the input used in the isolation sum is provided by the clusters as defined by the PF algorithm instead of the detector-based clusters, towers etc. The actual quantity being cut on is called the relative isolation, which, similar to the single muon trigger, means that it is the ratio between the isolation sum and the transverse energy of the electron candidate.

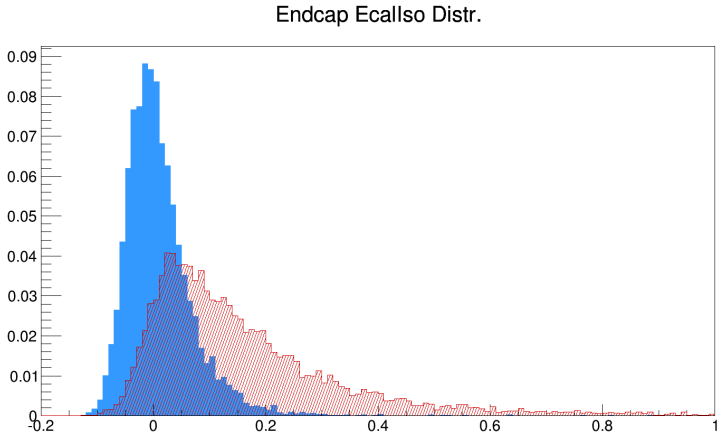
First, the Ecal isolation is computed within a cone of $\Delta R < 0.3$ around the electron candidate. Then the Hcal isolation is computed within the same cone size. The third type of isolation being cut on in single electron trigger, the track isolation, is not computed at this stage as it requires input from the track reconstruction step which is run later in the path. Figure 4.6 and Figure 4.7 show the distributions and efficiencies of the relative Ecal and Hcal isolation respectively.

4.3.4 Track Identification Variables: 1/E - 1/P, Fit χ^2, $\Delta\eta$ and $\Delta\phi$

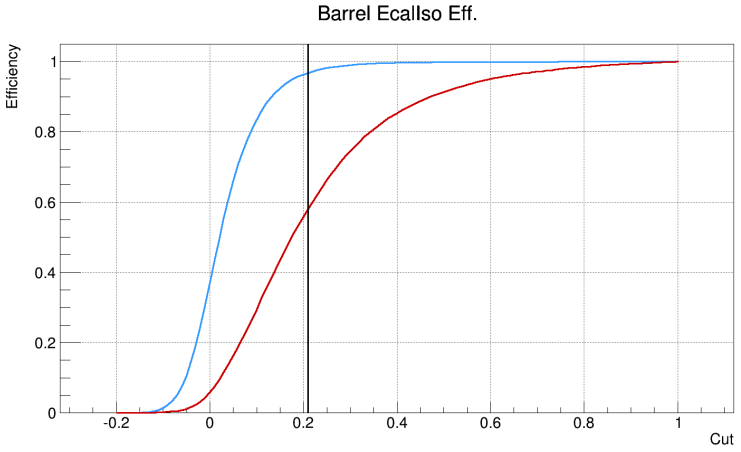
The track reconstruction step provided us with many quantities from which identification variables can be computed. There are three variables used in the single electron trigger; 1/E - 1/P, $\Delta\eta$ and $\Delta\phi$. The first one, 1/E - 1/P, is the difference between the inverse of supercluster energy E and the inverse of track



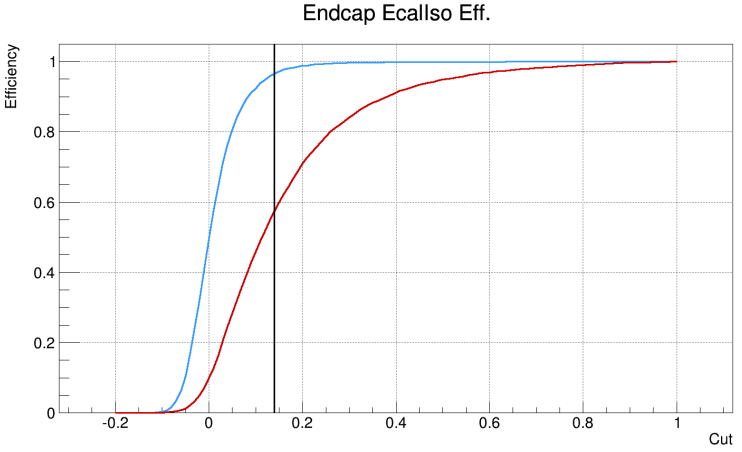
(a) Barrel EcalIso distribution



(b) Endcap EcalIso distribution

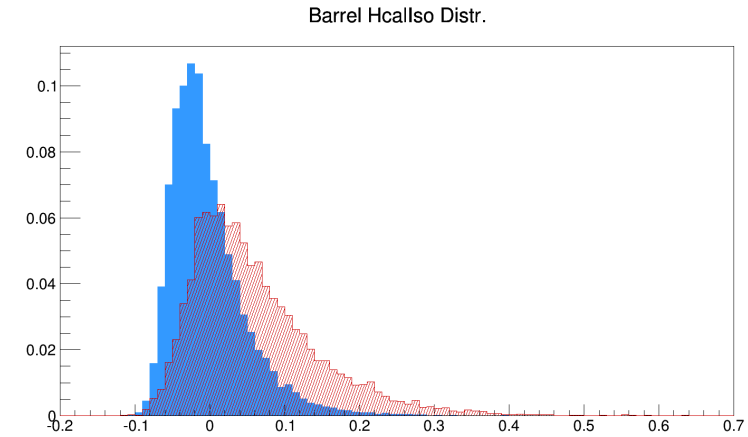


(c) Barrel EcalIso efficiency

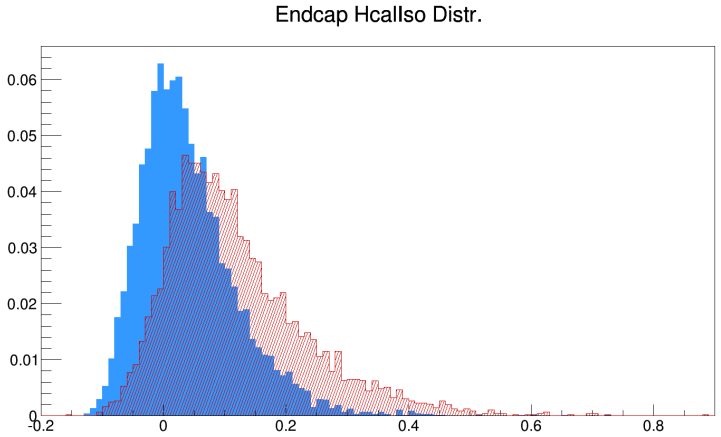


(d) Endcap EcalIso efficiency

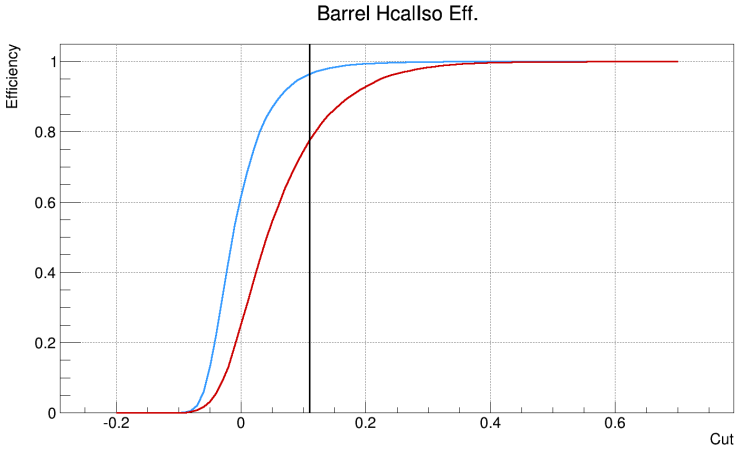
Figure 4.6: Signal (blue) and background (red) distribution and efficiencies of the relative Ecal isolation. For the barrel region, the optimized cut was chosen to be 0.21, which provided a 41.1% background rejection at 96.8% signal efficiency. For the endcap region, the optimized cut was chosen to be 0.14, which provided a 42.6% background rejection at 96.5% signal efficiency. The vertical lines in the efficiency graphs denote the optimized cut point and the corresponding signal and background efficiencies.



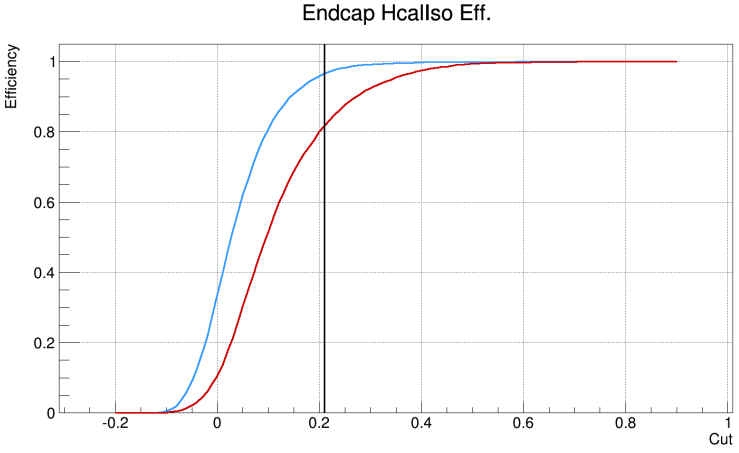
(a) Barrel HcalIso distribution



(b) Endcap HcalIso distribution



(c) Barrel HcalIso efficiency



(d) Endcap HcalIso efficiency

Figure 4.7: Signal (blue) and background (red) distribution and efficiencies of the relative Hcal isolation. For the barrel region, the optimized cut was chosen to be 0.11, which provided a 22.4% background rejection at 96.5% signal efficiency. For the endcap region, the optimized cut was chosen to be 0.21, which provided a 18.3% background rejection at 96.6% signal efficiency. The vertical lines in the efficiency graphs denote the optimized cut point and the corresponding signal and background efficiencies.

momentum P. This variable is motivated by the fact that electrons are light particles and are therefore very relativistic within the p_T regime considered by the single electron trigger. As such, the mass contribution towards electron energy is very small and the variable is therefore expected to peak near zero for signal electrons, ensuring compatibility between supercluster and track measurements. This was indeed what was observed, as shown in Figure 4.8.

The second variable, fit χ^2, is the normalized χ^2 of the track fitting step. As this variable is a measure of the track quality rather than any physical properties of the electron, the discrimination power of this variable is rather weak compared to the other variables. Nevertheless, it provided a still-usable rejection of the background and was therefore included in the list of identification variables to be used in the trigger. The distributions and efficiencies of this variable are shown in Figure 4.9.

As the track reconstruction step is done independently of the clustering step, taking entirely different set of inputs, the output of the step is naturally also separate from the latter, with the only constraint being that the track reconstruction step is done only in a region compatible with the supercluster. However, as this region can contain multiple tracks, of which only one could be from the electron, it is very beneficial to introduce identification variables to ensure the compatibility of the reconstructed track with the supercluster. These variables are the $\Delta\eta$ and $\Delta\phi$, which are simply the absolute difference of track and supercluster η and ϕ respectively. Requiring the angular windows to be small amounts to requiring the track to point in the direction of the supercluster, increasing the likelihood that the particle leaving that track being the one depositing its energy into the supercluster. As expected, these variables

provide a good discrimination between signal and background, as could be seen in Figure 4.10 and Figure 4.11.

4.3.5 Relative Tracker Isolation: TrackIso

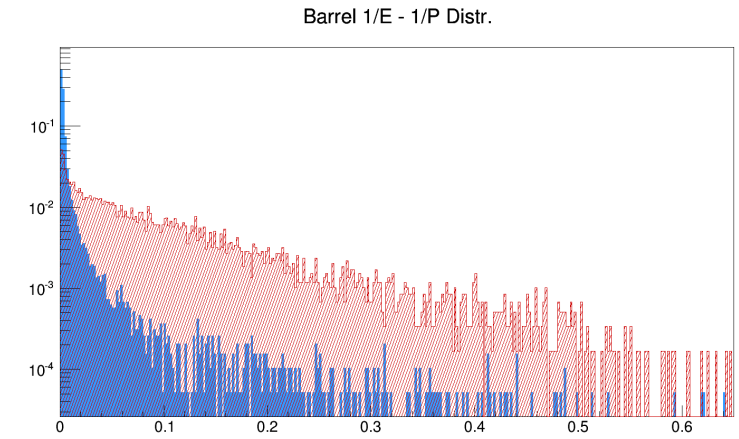
The final identification variable used in the single electron trigger is the track isolation, taking the input of all tracks around a cone of $\Delta R < 0.3$ around the electron track. This variable is cut on last in the trigger because in order to obtain the isolation sum, all the tracks within the cone have to be reconstructed and this is a very time-consuming step. Nevertheless it is a powerful variable rejecting a significant portion of the background, as shown in Figure 4.12.

4.3.6 Optimized Working Point: Single Electron WP75

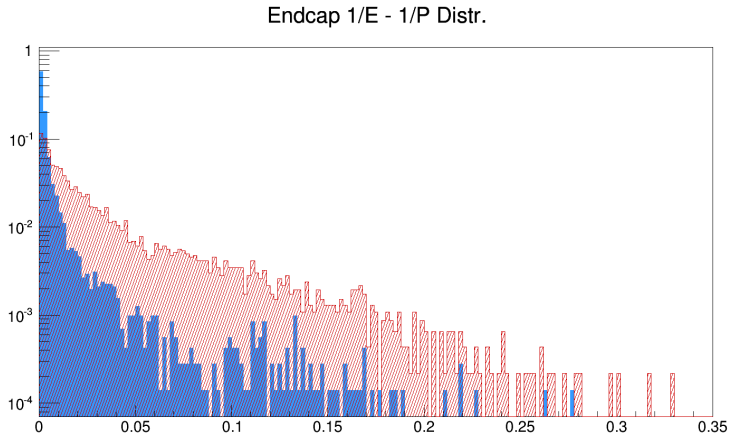
All the optimized cuts described above are combined into a set called a working point (WP) to denote a particular selection at trigger level. As the total signal efficiency of this WP is 78.4% and 75.8% in barrel and endcap regions respectively, this WP is referred to as WP75 in the full trigger menu. Table 4.2 summarizes the cuts and efficiencies of the WP75.

4.4 Rate Estimation

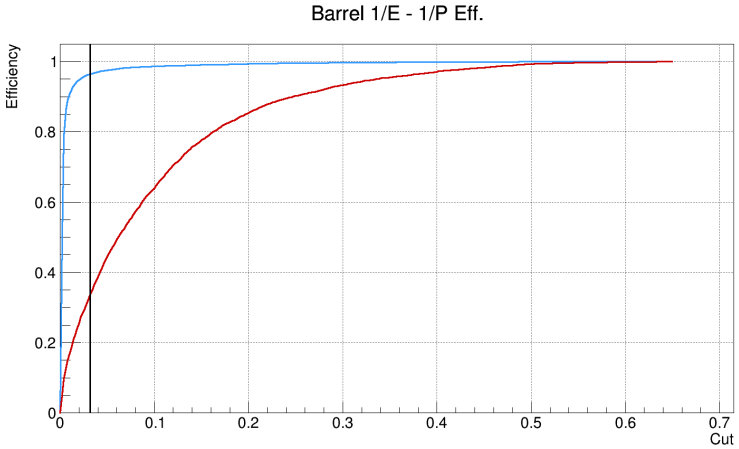
In Table 4.2, the total rate for the WP75 set is referred to. This rate was calculated using what was called the 'math method' [26], where the rate is estimated directly from efficiencies of simulated common process passing the trigger, rather than the 'scaling method' where the data was used to estimate the



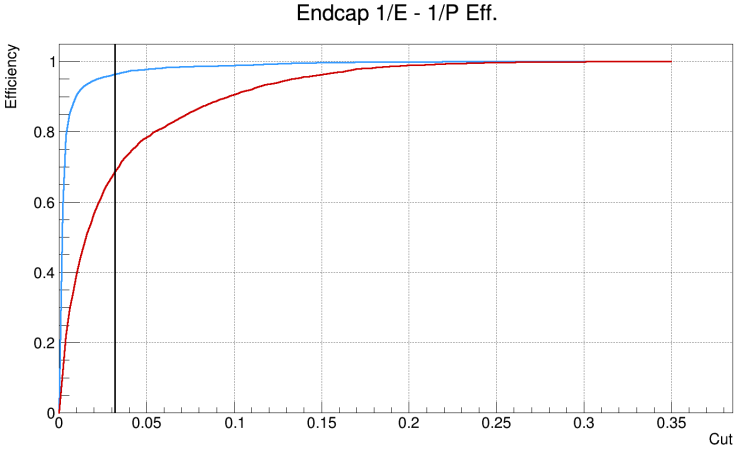
(a) Barrel 1/E - 1/P distribution



(b) Endcap 1/E - 1/P distribution

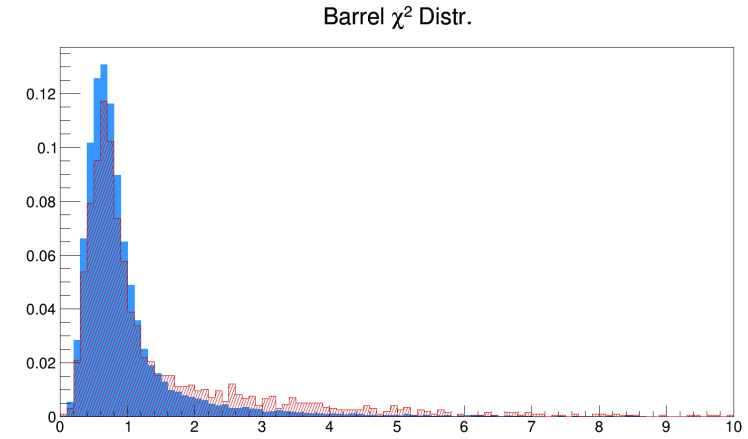


(c) Barrel 1/E - 1/P efficiency

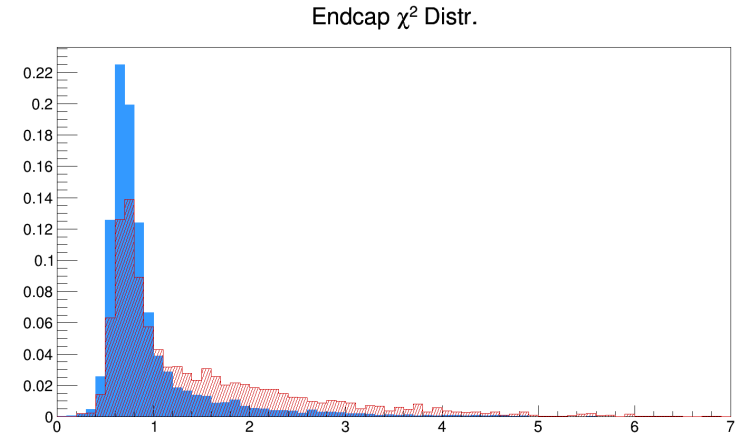


(d) Endcap 1/E - 1/P efficiency

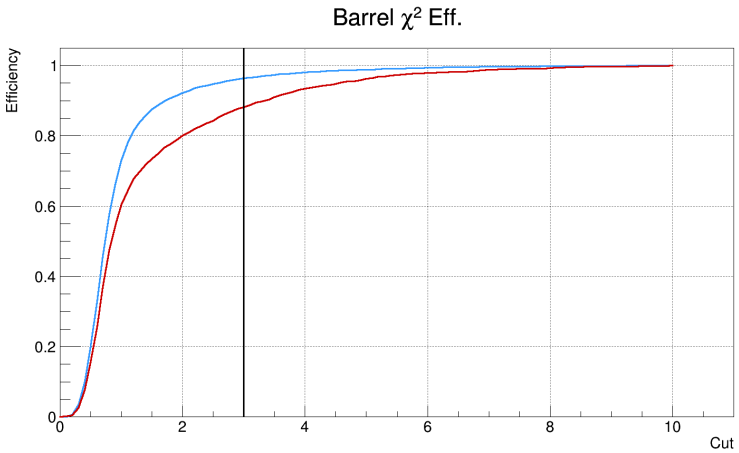
Figure 4.8: Signal (blue) and background (red) distribution and efficiencies of the 1/E - 1/P variable. For the barrel region, the optimized cut was chosen to be 0.032, which provided a 66.4% background rejection at 96.4% signal efficiency. For the endcap region, the same cut point was chosen, which provided a 31.4% background rejection at 96.4% signal efficiency. The vertical lines in the efficiency graphs denote the optimized cut point and the corresponding signal and background efficiencies.



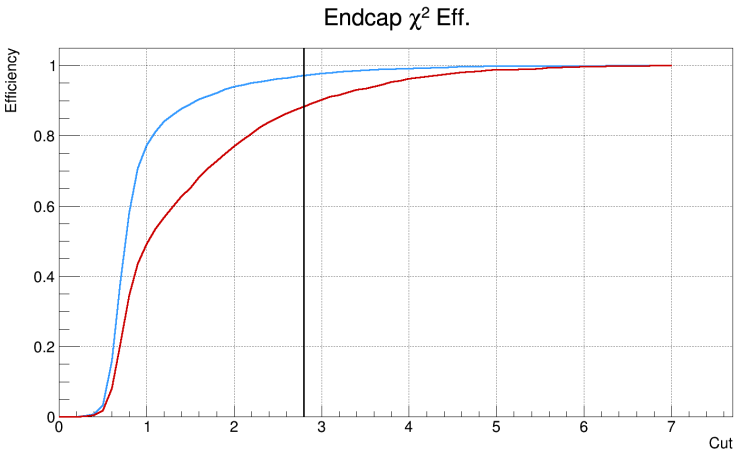
(a) Barrel fit χ^2 distribution



(b) Endcap fit χ^2 distribution

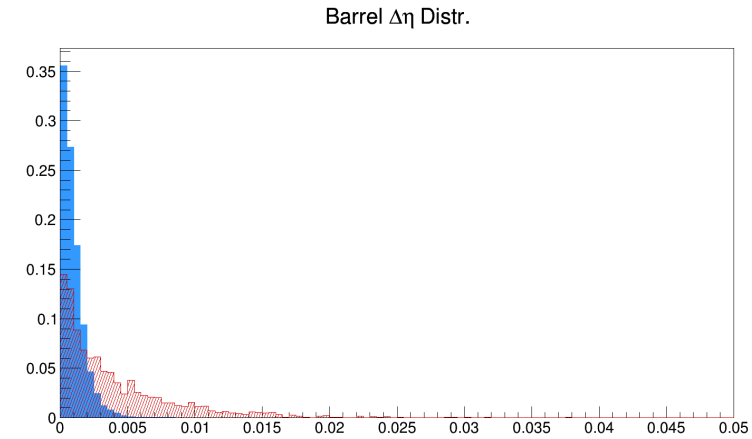


(c) Barrel fit χ^2 efficiency

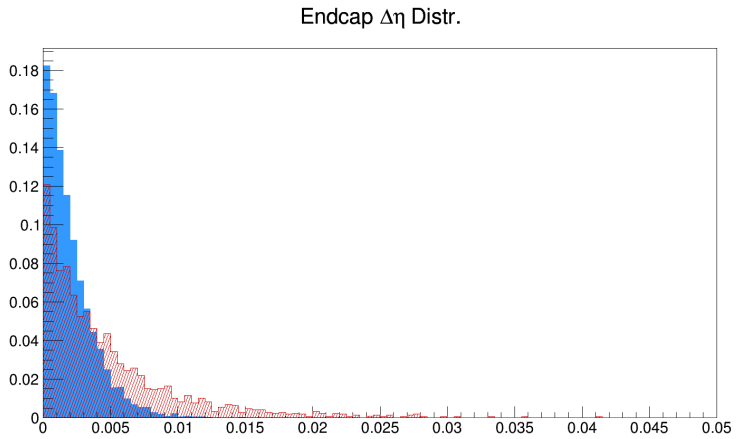


(d) Endcap fit χ^2 efficiency

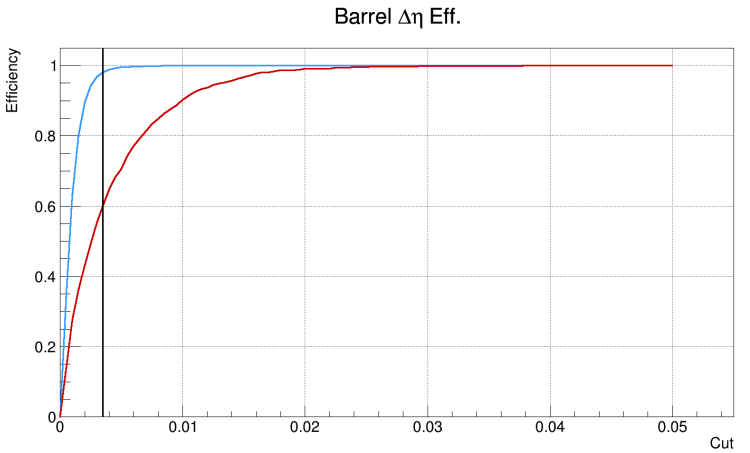
Figure 4.9: Signal (blue) and background (red) distribution and efficiencies of the track fit χ^2. For the barrel region, the optimized cut was chosen to be 3.0, which provided a 11.9% background rejection at 96.4% signal efficiency. For the endcap region, the optimized cut was chosen to be 2.8, which provided a 11.7% background rejection at 97.2% signal efficiency. The vertical lines in the efficiency graphs denote the optimized cut point and the corresponding signal and background efficiencies.



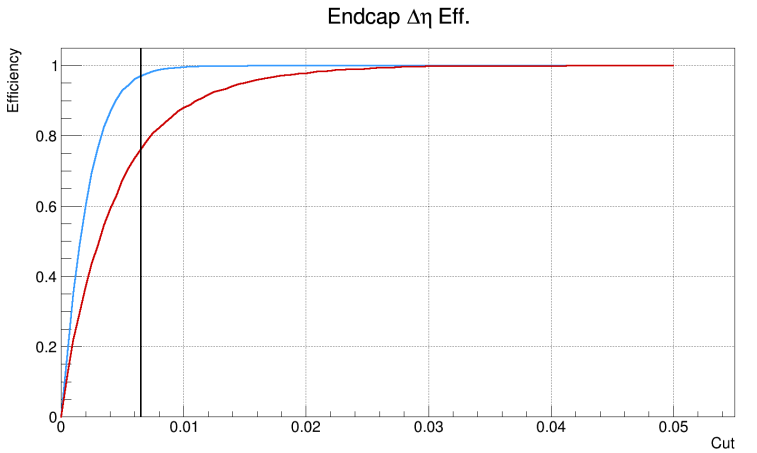
(a) Barrel $\Delta\eta$ distribution



(b) Endcap $\Delta\eta$ distribution

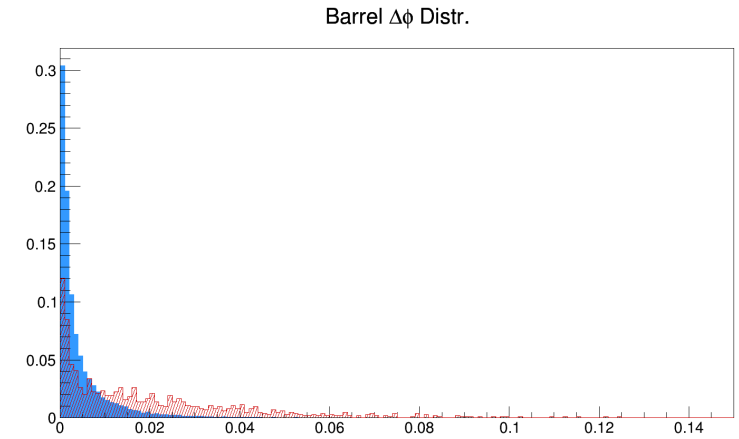


(c) Barrel $\Delta\eta$ efficiency

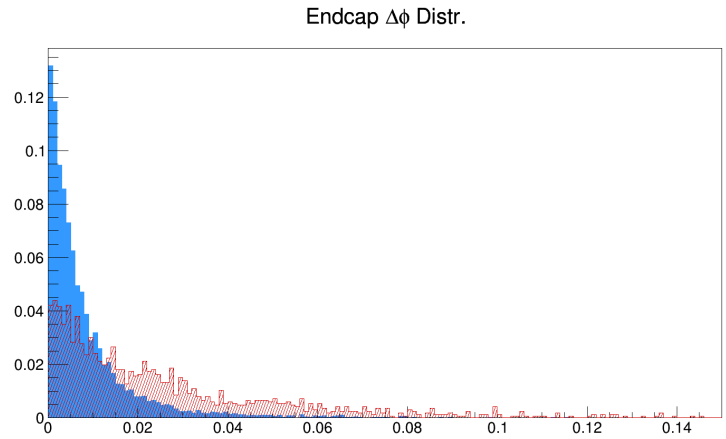


(d) Endcap $\Delta\eta$ efficiency

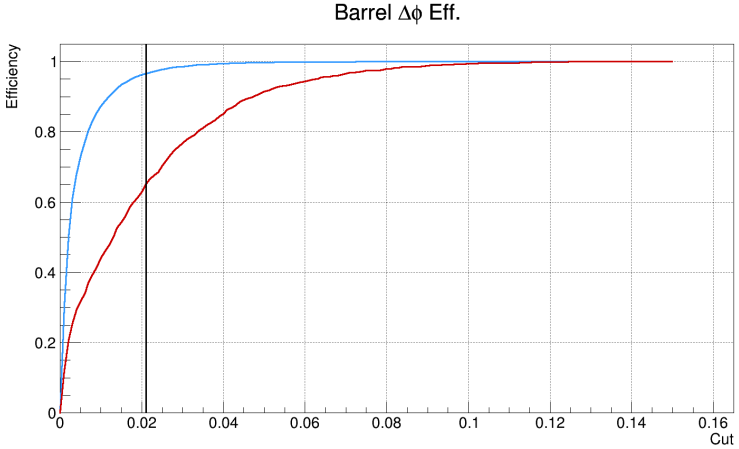
Figure 4.10: Signal (blue) and background (red) distribution and efficiencies of the $\Delta\eta$ variable. For the barrel region, the optimized cut was chosen to be 0.0035, which provided a 39.9% background rejection at 98.0% signal efficiency. For the endcap region, the optimized cut was chosen to be 0.0065, which provided a 23.9% background rejection at 97.1% signal efficiency. The vertical lines in the efficiency graphs denote the optimized cut point and the corresponding signal and background efficiencies.



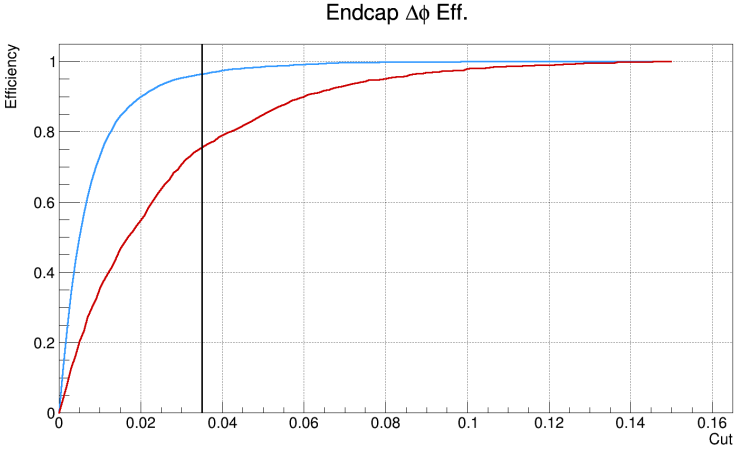
(a) Barrel $\Delta\phi$ distribution



(b) Endcap $\Delta\phi$ distribution

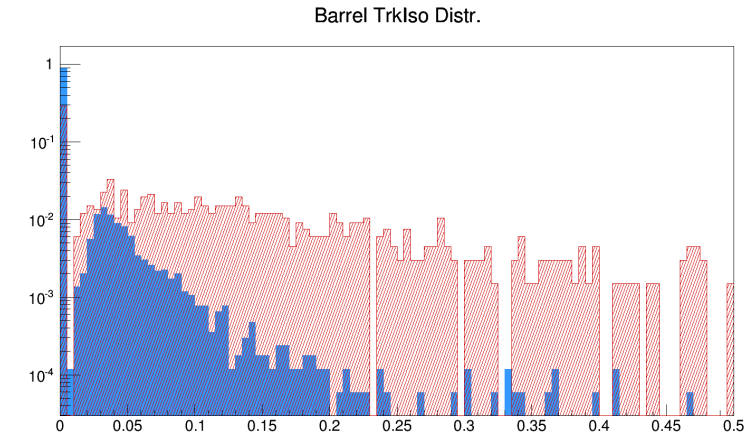


(c) Barrel $\Delta\phi$ efficiency

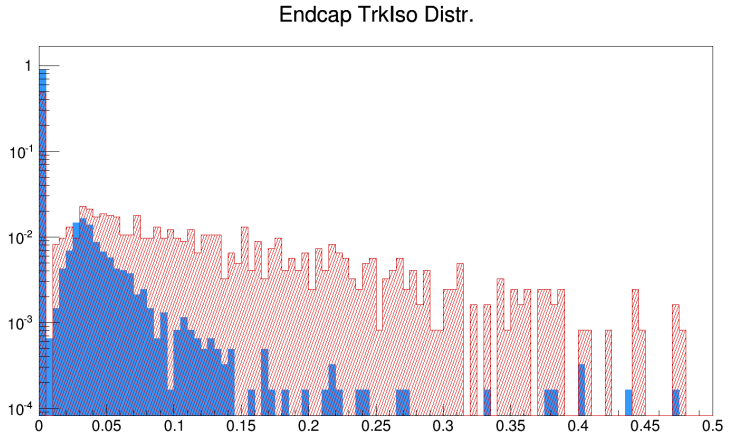


(d) Endcap $\Delta\phi$ efficiency

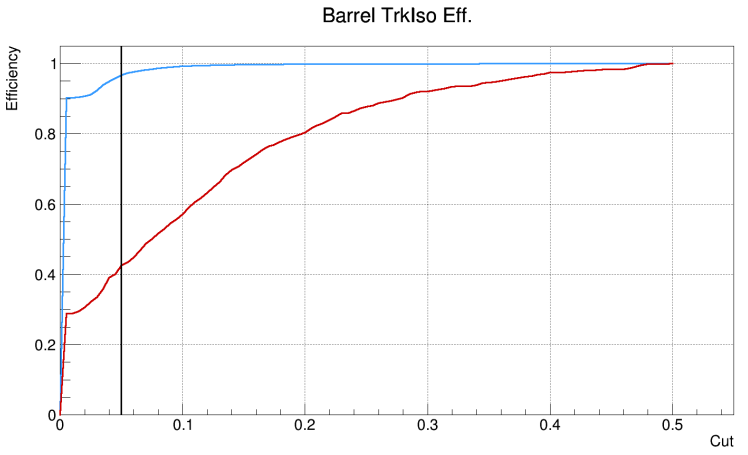
Figure 4.11: Signal (blue) and background (red) distribution and efficiencies of the $\Delta\phi$ variable. For the barrel region, the optimized cut was chosen to be 0.021, which provided a 34.9% background rejection at 96.6% signal efficiency. For the endcap region, the optimized cut was chosen to be 0.035, which provided a 24.5% background rejection at 96.5% signal efficiency. The vertical lines in the efficiency graphs denote the optimized cut point and the corresponding signal and background efficiencies.



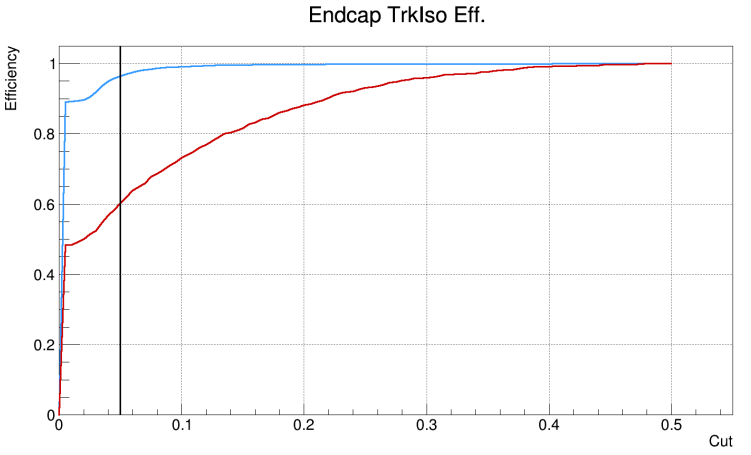
(a) Barrel TrackIso distribution



(b) Endcap TrackIso distribution



(c) Barrel TrackIso efficiency



(d) Endcap TrackIso efficiency

Figure 4.12: Signal (blue) and background (red) distribution and efficiencies of the relative tracker isolation. For the barrel region, the optimized cut was chosen to be 0.05, which provided a 57.5% background rejection at 96.7% signal efficiency. For the endcap region, the optimized cut was chosen to be 0.05, which provided a 39.7% background rejection at 96.5% signal efficiency. The vertical lines in the efficiency graphs denote the optimized cut point and the corresponding signal and background efficiencies.

background efficiencies. Although this method provides a more accurate estimation of the rate, it has the disadvantage of being much slower, as very large number of events are needed to obtain an accurate event yield necessary for the estimation. It is for this reason this method is infeasible where many iterations are involved in the optimization method, like the single muon trigger optimization described in Chapter 3. In the single electron trigger optimization however, where the differing framework allows the optimization to be done in a single iteration over the full phase space, this is the method of choice due to its higher accuracy.

In order to estimate the rate, one has to run the trigger with the desired set of cuts such as WP75 over common processes at the LHC, which are taken to be dijet QCD, $W \to e\nu$, dielectron production in Drell-Yan process [27] and top quark pair production. After obtaining the event yield in all these processes, the rate, r measured in Hz, and its uncertainty Δr, are given as:

Table 4.2: Cut points of all identification variables of single electron WP75 in barrel and endcap.

Variable	Barrel	Endcap
$\sigma_{i\eta i\eta}$	0.011	0.031
H - 0.01E	4.0	13.0
EcalIso	0.21	0.14
HcalIso	0.11	0.21
1/E - 1/P	0.032	0.032
$\Delta\eta$	0.0035	0.0065
$\Delta\phi$	0.021	0.035
Fit χ^2	3.0	2.8
TrackIso	0.05	0.05
Sig. Eff.	78.4%	75.8%
Bkg. Eff.	2.3%	5.3%
Rate (Hz)	64.8	109.0

$$r = c\left(1 - e^{\frac{\sigma\ell\epsilon}{c}}\right), \Delta r = \sigma\ell\sqrt{\frac{p + p\epsilon}{t}} \tag{4.2}$$

Meanwhile, the variables inside the rate formula are defined as:

- Number of particles in a (maximally) filled bunch n (m)
- Bunch crossing time b
- Collision rate $c = \frac{n}{mb}$
- Instantaneous luminosity $\ell = 1.4 \times 10^{34}$ cm^{-2} s^{-1}
- Cross section of the process σ in cm^2, 1 pb $= 1 \times 10^{-36}$ cm^2
- Passing (total) event p (t)
- Efficiency $\epsilon = \frac{p}{t}$

Finally, the cross sections used in the rate calculations are the standard ones provided by the STEAM group of CMS for rate calculations. These are summarized in Table 4.3.

Table 4.3: Cross sections of common processes considered in rate calculation.

Process	Cross Section (pb)
QCD $\hat{p}_T$ 30 - 50 GeV	161500000
QCD $\hat{p}_T$ 50 - 80 GeV	22110000
QCD $\hat{p}_T$ 80 - 120 GeV	3000114
QCD $\hat{p}_T$ 120 - 170 GeV	493200
$W \rightarrow e\nu$	16000
$DY \rightarrow ee$	6960
$t\bar{t}$	806

4.5 Additional Single Electron Studies

As the goal of the trigger system is to record events pertaining to a particular physics needs, it is crucial to fully understand the behavior of an optimized working point and explore the available alternative options. It would also be desirable to increase the signal acceptance of the path whenever possible without increasing the rate. This section describes the additional studies done on the single electron WP75 with these goals in mind.

4.5.1 Additional Identification Handles: Valid and Missing Hits

Although the variables described in Section 4.3 and used in the WP75 of single electron trigger are the main handles with the highest separation power, they are by no means the only ones. From the information provided by the calorimeters and Tracker, many others could be defined, such as $\sigma_{\phi\phi}$, which is the shape of the shower in ϕ direction.

As the track reconstruction takes as input the hits in the Tracker, the number of valid hits is a viable handle to identify the electron candidates with, as a high quality track is usually reconstructed with a high hit count. Conversely, the number of missing hits (or more specifically, the number of active Tracker layer in which no hits are registered) is also a viable handle, as in the presence of a missing hit, the trajectory along this layer is simply the product of track fitting procedure without any hits to constrain it.

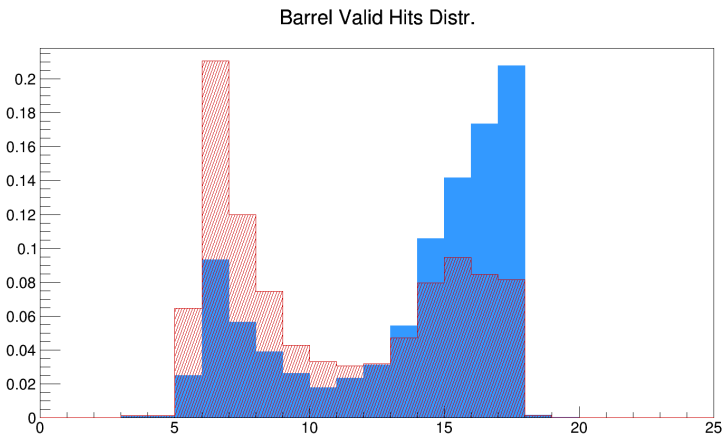
As could be seen in Figure 4.13 however, these do not provide a good discrimination between the signal and background. The number of valid

hits do not distinguish the type of particles leaving them in the Tracker and therefore do not discriminate between electrons and the copious amount of charged hadrons produced in background processes. While missing hits seem to provide a usable discrimination in the endcap, perhaps due to the fact that objects from QCD events typically have high values of $|\eta|$, thus increasing the chances of them missing a Tracker layer, this variable is sensitive to detector alignment issues and is therefore unstable. Because of their low performance in the case of valid hits and lack of stability in the case of missing hits, both variables are not used in the WP75.

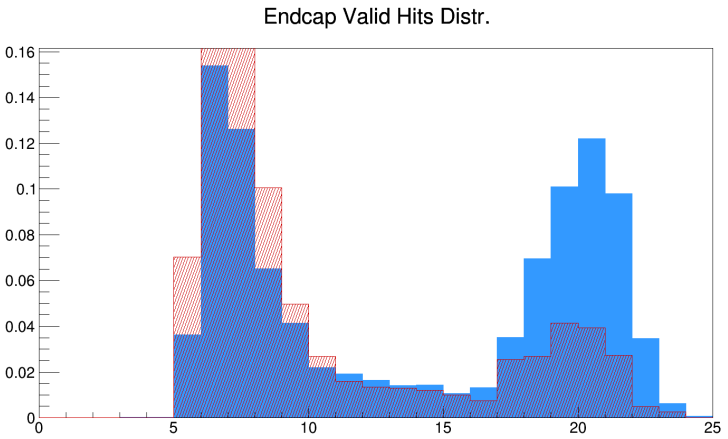
4.5.2 Barrel-Only Restriction of Single Electron Trigger

Other than the type of objects selected, two major selection criteria of a given trigger path are the minimum p_T of the candidate and its η region. Although it would be desirable to open the phase space as much as possible to maximize the signal acceptance, these are usually determined by the rate of the trigger, which is determined by the overall resources available to the experiment.

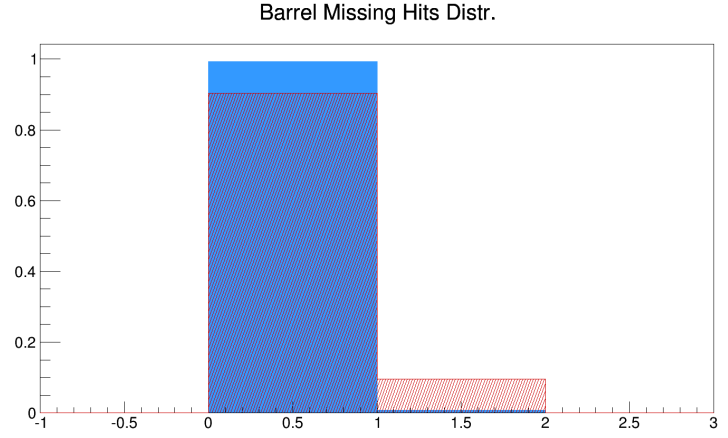
In the case of single electron trigger, the p_T threshold is 32 GeV, as indicated by the path name in Table 4.1. As roughly only 70% of electrons resulting from top quark decays have p_T above 30 GeV (Figure 4.14), the trigger threshold is seen as too high for top physics, particularly considering that the p_T cut in a physics analysis is set typically a few GeV higher than the trigger threshold. A proposal to lower this threshold by sacrificing the endcap region was investigated. As the main limiting factor of this threshold is the very high rate at L1, the investigation is conducted by comparing the rates of the default acceptance scenario (p_T threshold 32 GeV and $|\eta|$ threshold 2.1) with the lowered p_T threshold scenario and tightened $|\eta|$ cut to cover only the barrel region.



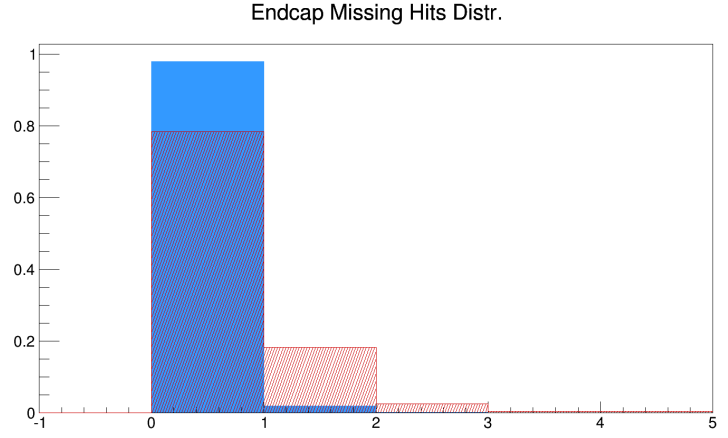
(a) Barrel valid hits distribution



(b) Endcap valid hits distribution



(c) Barrel missing hits distribution



(d) Endcap missing hits distribution

Figure 4.13: Distribution of valid and missing hits in barrel and endcap for signal (blue) and background (red) processes.

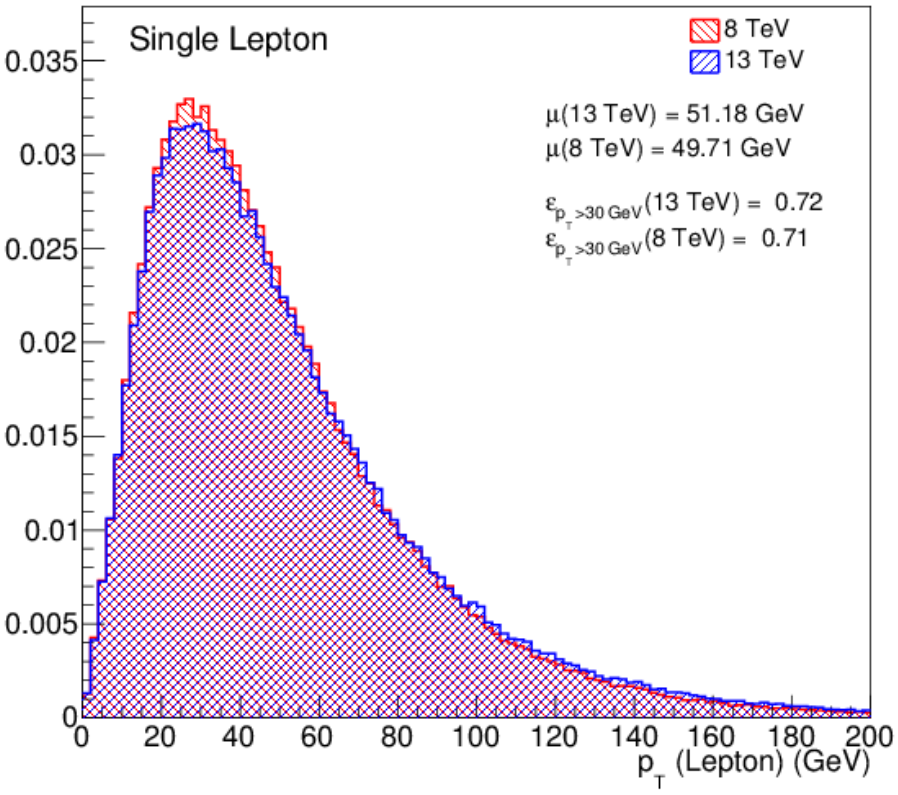


Figure 4.14: Lepton p_T distribution in semileptonic $t\bar{t}$ events at two different energy; 8 TeV and 13 TeV [28].

Table 4.4 reports the rates in the two lower p_T and the default scenarios. Note that rate uncertainties of zero mean that the uncertainties in the rate estimation is smaller than the degree of accuracy considered for the study.

From the rates, taking the L1 of the full region at p_T threshold 32 GeV to be an acceptable rate, it was observed that we can go down to 27 GeV by restricting ourselves only to the barrel region. However, the overall signal rate (defined as rate contribution from non-QCD processes) decreases by roughly 10% in doing so, thus reducing the overall signal acceptance. From this result it was seen that restricting the trigger to the barrel region is not a commendable course of action.

Table 4.4: L1 rates in the full and barrel acceptance region for three different trigger p_T thresholds.

Process	$p_T > 27$ GeV		$p_T > 30$ GeV		$p_T > 32$ GeV	
	Full (Hz)	Barrel (Hz)	Full (Hz)	Barrel (Hz)	Full (Hz)	Barrel (Hz)
QCD $\hat{p}_T$ 30 - 50 GeV	10574.7 $\pm$ 89.9	6904.6 $\pm$ 72.5	7065.4 $\pm$ 73.3	4661.4 $\pm$ 59.5	5380.0 $\pm$ 64.0	3573.9 $\pm$ 52.1
QCD $\hat{p}_T$ 50 - 80 GeV	5161.2 $\pm$ 23.3	3072.7 $\pm$ 17.9	4344.6 $\pm$ 21.3	2619.6 $\pm$ 16.5	3858.6 $\pm$ 20.1	2338.7 $\pm$ 15.6
QCD $\hat{p}_T$ 80 - 120 GeV	1179.7 $\pm$ 4.1	707.9 $\pm$ 3.2	1112.7 $\pm$ 4.0	669.1 $\pm$ 3.1	1072.2 $\pm$ 3.9	645.4 $\pm$ 3.0
QCD $\hat{p}_T$ 120 - 170 GeV	343.2 $\pm$ 0.9	229.4 $\pm$ 0.7	334.8 $\pm$ 0.9	224.1 $\pm$ 0.7	330.0 $\pm$ 0.9	221.2 $\pm$ 0.7
$W \rightarrow e\nu$	79.2 $\pm$ 0.5	57.3 $\pm$ 0.4	71.6 $\pm$ 0.5	52.2 $\pm$ 0.4	65.9 $\pm$ 0.4	48.2 $\pm$ 0.4
$DY \rightarrow ee$	21.0 $\pm$ 0.0	15.9 $\pm$ 0.0	19.3 $\pm$ 0.0	14.7 $\pm$ 0.0	18.3 $\pm$ 0.0	13.8 $\pm$ 0.0
$t\bar{t}$	2.6 $\pm$ 0.0	2.1 $\pm$ 0.0	2.5 $\pm$ 0.0	2.1 $\pm$ 0.0	2.4 $\pm$ 0.0	2.0 $\pm$ 0.0
Total	17362.5 $\pm$ 93.0	10990.6 $\pm$ 74.7	12951.7 $\pm$ 76.4	8243.8 $\pm$ 61.8	10728.2 $\pm$ 67.2	6843.8 $\pm$ 54.5

Chapter 5

Data-Driven Measurement of Single Muon Trigger Efficiencies

"The true method of knowledge is experiment."

— W. Blake, 1757 – 1827

As the events that do not pass any triggers in real collisions are not saved into the data and therefore forever lost, data trigger efficiencies can not be measured using the simple procedure used in the previous chapters. Instead, a reliable method requiring no information beyond those provided by recorded events is necessary. This chapter describes the measurement of the muon leg efficiency of single muon cross trigger using the Tag & Probe method, done on 3.9 fb^{-1} of data taken by CMS at $\sqrt{s} = 7$ TeV.

5.1 Tag & Probe Method

The Tag & Probe method makes use of mass resonances to measure the efficiency of a selection criteria on objects of a desired type [29]. Within the context of this chapter, the $Z \to \mu\mu$ decay was used. The method was named by the fact that one of the muon is required to be a *tag* muon, meaning that it is required to pass a very tight selection criteria suitable for a control muon. The other muon is called the *probe* muon, with which the efficiency of an intended selection criteria is measured according to the equation:

$$\text{Eff.} = \frac{N_{\text{probe, pass}}}{N_{\text{probe, total}}} \tag{5.1}$$

In order to ensure the compatibility of the simulation with respect to the data, the measurement was done on both MC and real data samples. On the Drell-Yan MC samples containing perfect signal yield, the number of probes was simply counted and divided into the passing or failing categories following the invariant mass cut on the tag-probe pair. On the other hand, due to the presence of small but non-negligible presence of background in the real data, a fit on the signal and background is first performed before extracting the signal yield. The efficiencies were then measured in bins of kinematical properties of the probe. Finally, a scale factor (SF), defined as the ratio of the data and MC efficiencies, is calculated for each bin.

5.2 Muon Leg Efficiency Measurement

A cross trigger is a trigger in which two or more objects are required to fire the trigger. The cross trigger studied in this chapter is the single muon plus a jet path, `HLT_IsoMu17_CentralPFJet30`, which is designed to record events involving W bosons. Leg is a term referring to the set of filters focusing on a particular object within a cross trigger. The advantage of crossing triggers is in rate reduction; requiring more than one object in an event greatly reduces the rate at which the trigger is firing. This in turn allows one to increase the acceptance of the trigger to a degree that is not feasible with inclusive triggers, which can be very beneficial for analyses suffering from low signal statistics such as the search for single top s-channel process.

5.2.1 Measurement Setup

Table 5.1 summarizes the setup for the study. The Tag & Probe tool used was the official package provided in CMSSW [30].

Table 5.1: Setup and samples for the muon leg efficiency measurement.

Setup	Information
CMSSW	`CMSSW_4_4_5`
MC	7 TeV `Fall11` Drell-Yan
Data	`SingleMu` Run 2011A and 2011B

5.2.2 Muon Identification and Isolation Requirement

In order to improve the purity of the sample, only tags and probes passing a set of offline selection criteria are considered in the efficiency measurement. The criteria used was the standard muon selection criteria used in 7 TeV CMS analyses [31]. The full list of the cuts is given in Table 5.2.

Collectively, this set is identified as the `GlobalMuonPromptTight` working point, the performance of which was reported in detail elsewhere [32,33]. The muon type cut refers to the algorithm with which the muon was reconstructed, such as PF [13], global or tracker muon [21]. The matched stations on the other hand refers to the number of stations involved in the measurement of the muon candidate, while track hits and pixel layers refer to the parameters used in the tracker track reconstruction. Relative isolation on the other hand is a quantity combining the tracker, electronic and hadronic calorimeter information to calculate the isolation of the candidate with respect to charged particles, photons and neutral particles respectively.

Table 5.2: List of offline cuts imposed on tag and probe muons.

Cut	Tag	Probe
$p_T >$	24 GeV	20 GeV
$\lvert\eta\rvert <$	2.1	
Muon type	PF, Global, Tracker	
Matched stations $>$	1	
Valid track hits $>$	10	
Registered pixel layers $>$	1	
Relative isolation $<$	0.15	

5.2.3 Tag and Probe Trigger Paths

To ensure that the muons selected according to the above selection criteria are also the ones registered by the trigger whose efficiency are being measured, the muon is required to be within a cone of $\Delta R = 0.1$ from the trigger object. As the triggering conditions changed throughout the data taking, it was necessary that the trigger path for the tag and muon leg for the probe be adapted accordingly. For this reason, the efficiency measurement is divided into two scenarios for MC and three for data, depending on the tag trigger used. In both cases the tag trigger path was taken to be the unprescaled inclusive single muon trigger with the lowest p_T threshold available in the data. Although the efficiencies measured by the probe were different for MC and data, the muon leg of the single muon plus jet cross trigger share the same set of cuts as the inclusive path available in MC and was therefore behaving in the same way, thus validating a straightforward comparison between the two through the SFs. The information of tag and probe trigger paths for all the scenarios is available in Table 5.3.

To eliminate the bias resulting from different conditions in which MC and data events were taken, the MC events were reweighted to match the pileup conditions of the data. The events in MC A scenario was reweighted to match the 11A scenario while MC B was separately reweighted to match the 11AB and 11B scenarios. The SFs are then calculated according to this MC-data pairing.

5.2.4 Z Resonance Selection

After the tag and probe are selected, their four-momentum are summed together to reconstruct the Z boson. To improve the purity of the sample, only

Table 5.3: Tag and probe trigger paths for all MC and data scenarios. The scenarios 11A, 11AB and 11B refer to the which data run (2011A or 2011B) the considered run ranges belonged to.

Flag	Run Range	Int. Lumi (fb^{-1})	Tag Path	Probe Path/Filter
MC A	n/a	n/a	IsoMu24	IsoMu17
MC B	n/a	n/a	IsoMu24_eta2p1	IsoMu17
11A	167913 - 173198	0.878	IsoMu24	hltMuIsoCenJet L3IsoFiltered17
11AB	173236 - 178380	2.135	IsoMu24_eta2p1	hltMuEta2p1IsoCenJet L3IsoFiltered17
11B	178420 - 180252	0.897	IsoMu24_eta2p1	hltMuEta2p1IsoCenJet L3withL2QIsoFiltered17

dimuon pairs within an invariant mass window of 70 - 110 GeV are considered for the efficiency measurement. Additionally, the distance between tag and probe in z-direction must be less than 4 cm in order to reject muon pairs not coming from the same vertex. In the case of data, to reject the irreducible background, the mass spectrum is fit with a Voitgian function for signal distribution and an exponential function for the background before the signal yield is extracted. An example of the fitted distribution for passing probe is shown in Figure 5.1.

5.2.5 Results

As mentioned earlier in the chapter, the probe from pairs passing all selection criteria were split into passing or failing probes, depending on whether they pass the muon leg listed in Table 5.3. The efficiency of the filter is then given by equation 5.1. The efficiency and SF for each p_T and η bins is shown in Figure 5.2.

As could be seen in the plots, the muon leg has a high efficiency in all bins of p_T and η. Additionally, the SF distribution is almost flat with values close to unity, an indication that the MC describes the data very well.

5.3 Cross Check With *s*-channel Single Top

Although it was verified that the MC is a good description of the data, there remains the question whether the trigger efficiencies derived using the Tag & Probe method is applicable to events with a significantly different topology, such as the single top *s*-channel process. In order to settle this issue, the muon leg efficiencies in the two processes were compared. Since the Tag & Probe method could not be used in the *s*-channel process due to the lack of a dimuon resonance, for this process the efficiency is measured in the same way

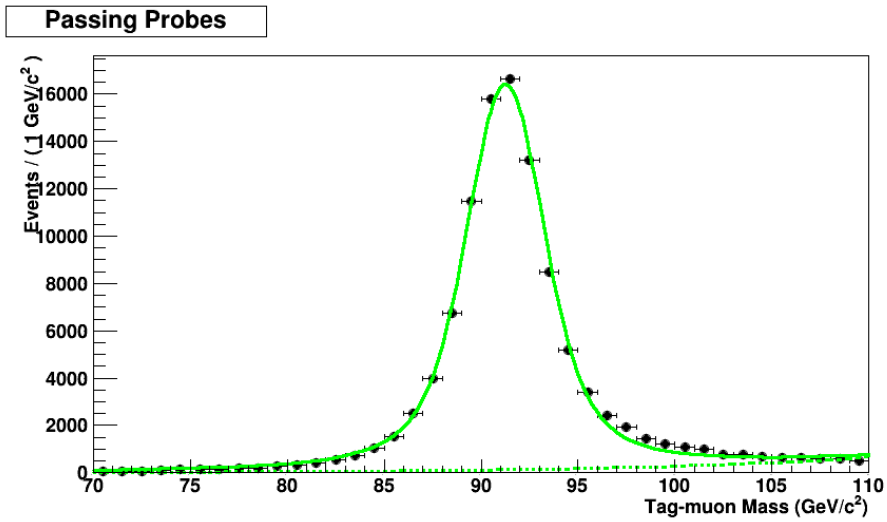
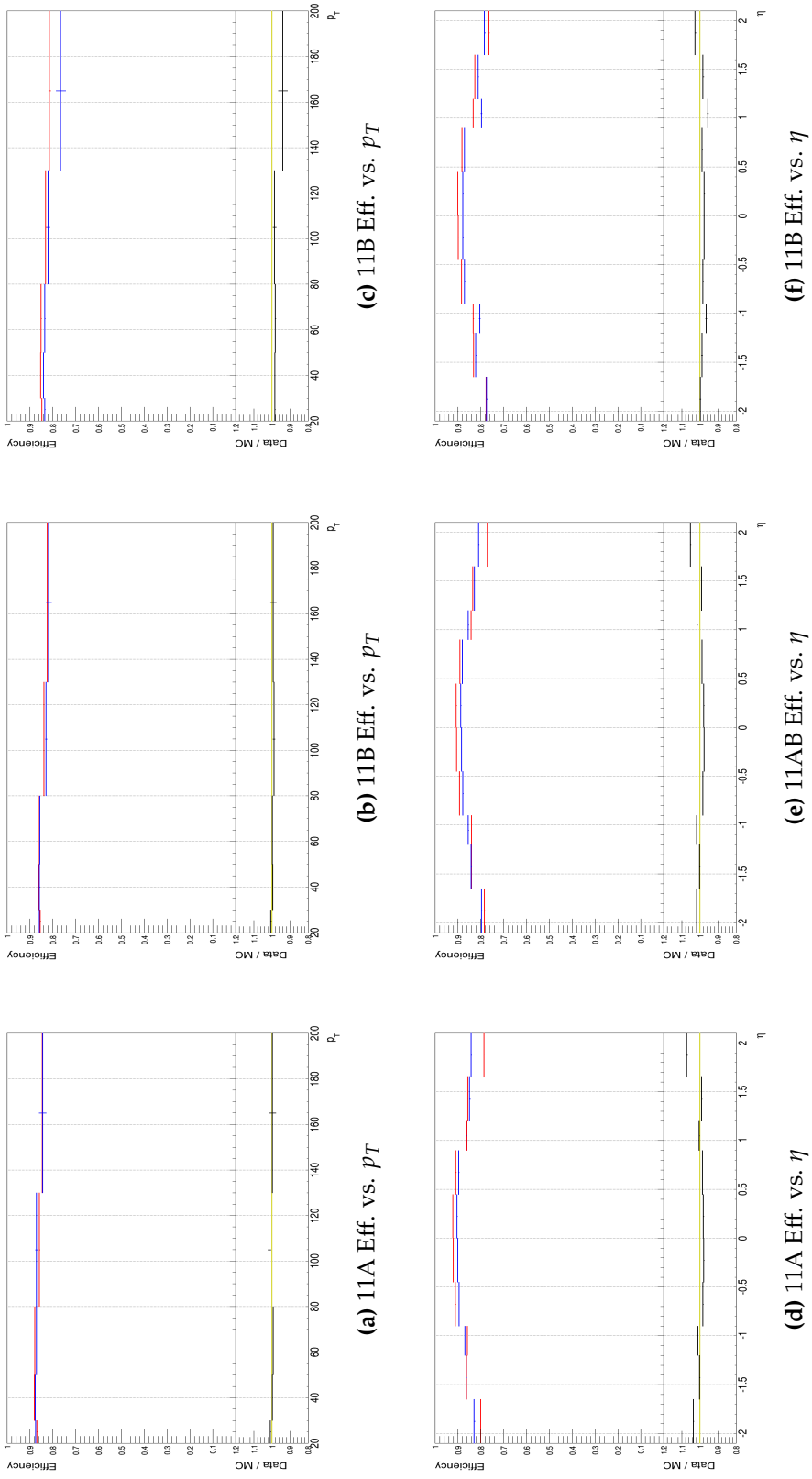


Figure 5.1: Tag and passing probe mass distribution for the 11AB scenario in the p_T bin 50 - 80 GeV. The mass spectrum is fitted with a Voitgian and exponential functions for signal and background respectively, with a peak at 91.260 $\pm$ 0.008 GeV, clearly compatible with that of Z-boson.



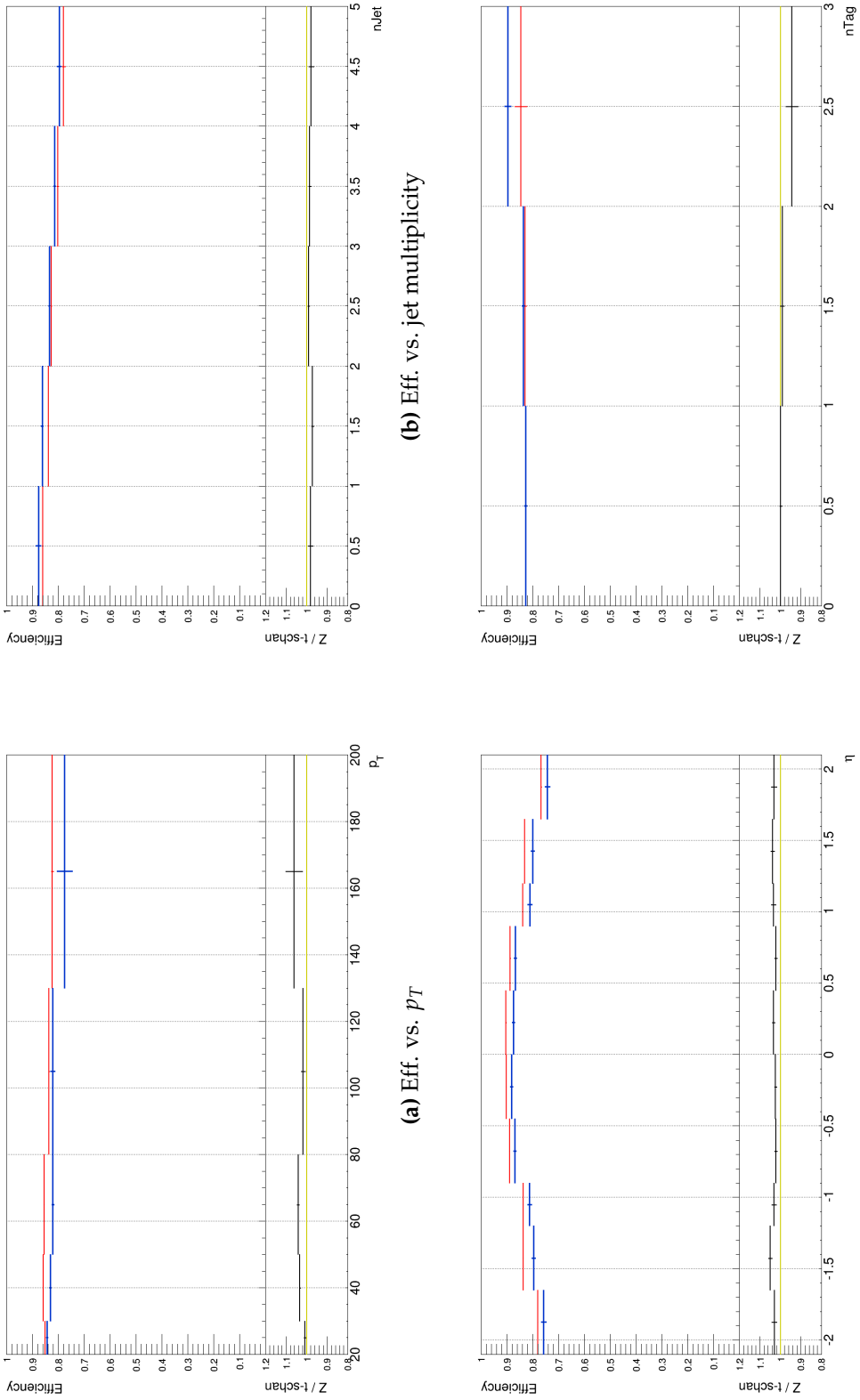
(a) 11A Eff. vs. p_T **(b)** 11B Eff. vs. p_T **(c)** 11B Eff. vs. p_T

(d) 11A Eff. vs. η **(e)** 11AB Eff. vs. η **(f)** 11B Eff. vs. η

Figure 5.2: Muon leg efficiencies in bins of p_T and η as measured using the probe. The red lines represent MC while blue represent data. Also shown is the distribution of ratios between the data and MC efficiencies, the SF.

as previous chapter; matching with a MC truth object. The results are shown in Figure 5.3.

In doing this cross check, care needs to be taken that one does not express the efficiencies in bins of a variable sensitive to the differences in event topology, as this would introduce an additional bias to the comparison. This could be seen in the p_T and η distributions, where the muon leg efficiencies of the two processes are influenced by the physics underlying the event topology. A fairer comparison could be made in variables insensitive to this effect, such as jet and b-tag multiplicity. In these variables it could be seen that the efficiencies are very similar to each other and their SF distribution is also almost flat. The inefficiency at b-tag = 2 region for DY is attributed to low statistics, as the DY events contain no genuine b-jets and therefore a significantly lower number of b-tagged jets.



(a) Eff. vs. p_T

(b) Eff. vs. jet multiplicity

(c) Eff. vs. η

(d) Eff. vs. tag multiplicity

Figure 5.3: Muon leg efficiencies for DY and single top s-channel MC. The red lines represent DY while blue represents single top s-channel. Also shown is the distribution of SF between the two.

Chapter 6

Conclusions and Outlook

"This was good training for research, because large parts of experimental work are sometimes boring or involve the use of skills in which one is not particularly gifted."

— M. L. Perl, 1927 – 2014

The lepton reconstruction algorithms used in inclusive single lepton triggers at CMS have been studied and optimized for the harsher conditions of Run 2, with an emphasis on minimizing the rate for a given signal efficiency.

For the single muon trigger, this was done using the OpenHLT tool using a simulated $t\bar{t}$ sample as the signal and the background approximated by the use of a zero-bias parked data. Muon identification parameters in both L2 and L3 reconstruction have been shown to already be at their optimal points from the large drop of signal efficiency when the cuts are tightened. The combined relative isolation cut on the other hand provided some window

for optimization, with the proposed cut of 0.12 leading to additional 10% background rejection with only 2% signal loss.

For the single electron trigger, the more complex reconstruction algorithm necessitated a different approach to the optimization. It was done by taken a signal and background yield over the entire phase space, obtained by imposing only very loose preselection on the electron candidates. MC samples are used for both signal and background, with the $W \to e\nu$ taken to be signal and dijet QCD taken as background. The set of optimized cuts yielded signal efficiencies of 78% and 76% in the barrel and endcap region respectively, while achieving background rejections of 98% and 95% in the respective regions. This translates to a total expected rate of 174 Hz. Several other proposals to further optimize the single electron trigger or increase its acceptance has also been explored.

Finally, the efficiency of the single muon trigger in data was measured using the Tag & Probe method using a large portion of the 7 TeV data. The choice of using an old data for this measurement is motivated by two reasons; the fact that the data-taking period for which the single lepton triggers developed in this thesis has yet to come and also that the trigger efficiencies for that particular portion of CMS data was yet to be measured but required for some CMS legacy physics analyses, in particular the single top s-channel cross section measurement. The choice to measure the efficiencies only for the single muon cross triggers and not single electron was similarly driven by availability and necessity within the CMS collaboration. Primarily, this measurement aims to serve as a practice for the Run 2 data-taking period to come, as well as establishing the technique to be used during that time.

Compared to Run 1, the reconstruction algorithms used at HLT are very close to the version used offline versions used at the analysis level. For example,

a new version of L3 muon reconstruction algorithm is used which is based on iterative tracking algorithm, the one used for offline track reconstruction, with a custom trajectory seeder code based on pixel hits to control the timing. Electron reconstruction on the other hand had been improved as well, including the use of the same clustering algorithm used offline, weekly laser corrections to account for transparency loss in Ecal crystals over time and the use of PF-based isolation instead of calorimeter-based as in Run 1. In general there have been also numerous code improvements and cleaning to make the reconstruction logic more robust, debugging simpler and also to benefit from the advances in the computational and programming fields in recent years. This has been a major milestone achieved in preparing for the harsher Run 2.

Bibliography

[1] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).

[2] S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. **D2**, 1285 (1970).

[3] L. Evans and P. Bryant, Journal of Instrumentation **3**, S08001 (2008).

[4] S. Chatrchyan *et al.*, Journal of Instrumentation **3**, S08004 (2008).

[5] D. Barney, Simple short introduction to CERN & CMS, A short talk to introduce CERN and CMS to a non-expert audience, 2014.

[6] CMS Collaboration, J. Man *et al.*, *CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter* (CERN, Geneva, 2012).

[7] CMS Collaboration, S. Chatrchyan *et al.*, JINST **5**, T03012 (2010), 0911.4991.

[8] CMS Collaboration, S. Chatrchyan *et al.*, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software* (CERN, Geneva, 2006).

[9] CMS Collaboration, S. Dasu *et al.*, *CMS The TriDAS project; Technical Design Report, Volume 1: The Trigger Systems* (CERN, 2000).

[10] CMS Collaboration, S. Cittolin, A. Racz, and P. Sphicas, *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project* (CERN, Geneva, 2002).

[11] CMS Collaboration, S. Chatrchyan *et al.*, (2012), CMS Physics Analysis Summary TRG-12-001.

[12] https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookDataFormats.

[13] CMS Collaboration, S. Chatrchyan *et al.*, (2009), CMS Physics Analysis Summary PFT-09-001.

[14] C. Vuosalo and D. Lange, MiniAOD: A new analysis data format for CMS, 21st International Conference on Computing in High Energy and Nuclear Physics, 2015.

[15] T. Sjostrand, S. Mrenna, and P. Z. Skands, Comput. Phys. Commun. **178**, 852 (2008), 0710.3820.

[16] GEANT4 Collaboration, S. Agostinelli *et al.*, Nucl. Instrum. Meth. **A506**, 250 (2003).

[17] CMS Collaboration, S. Chatrchyan *et al.*, (2010), CMS Physics Analysis Summary MUO-10-003.

[18] CMS Collaboration, S. Chatrchyan *et al.*, JINST **9**, P10009 (2014), 1405.6569.

[19] R. Fruhwirth, Nucl. Instrum. Meth. **A262**, 444 (1987).

[20] https://twiki.cern.ch/twiki/bin/viewauth/CMS/NewOpenHLT.

[21] CMS Collaboration, CERN Report No. CMS-AN-2008-097, 2009 (unpublished).

[22] CMS Collaboration, S. Chatrchyan *et al.*, (2012), CMS Report CMS-DP-2012-022.

[23] CMS Collaboration, S. Chatrchyan *et al.*, (2015), 1502.02701, CMS Physics Analysis Summary EGM-13-001.

[24] https://twiki.cern.ch/twiki/bin/viewauth/CMSPublic/L1TriggerDPGResults.

[25] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov, Journal of Physics G: Nuclear and Particle Physics **31**, N9 (2005).

[26] https://twiki.cern.ch/twiki/bin/viewauth/CMS/TmdRecipes.

[27] S. D. Drell and T.-M. Yan, Phys. Rev. Lett. **25**, 316 (1970).

[28] B. Stieger, Internal CMS Presentation, https://indico.cern.ch/event/304729/.

[29] CMS Collaboration, CERN Report No. CMS-AN-2009-111, 2009 (unpublished).

[30] https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonTagAndProbe.

[31] https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId.

[32] CMS Collaboration, S. Chatrchyan *et al.*, (2010), CMS Physics Analysis Summary MUO-10-002.

[33] CMS Collaboration, S. Chatrchyan *et al.*, J. Instrum. **7**, P10002. 81 p (2012).