



TECHNISCHE  
UNIVERSITÄT  
WIEN  
Vienna University of Technology

# Optimizing Muon Selection in an Environment with many Jets and high Luminosity

Clemens Jochum

March 13, 2013

## **Abstract**

After a short overview of the LHC and the CMS experiment important principles of particle creation and reconstruction methods are explained. Essential isolation parameters and their calculations are presented and the consequential concept of efficiency and purity for lepton selection is introduced. A representative event is used to clarify the concept and parameters of Particle Flow reconstruction. Finally resulting isolation histograms and figures of merit of efficiency versus purity for various cut configurations are presented.

Under the guidance of:  
Dr. Dipl-Ing. Robert Schöffbeck

# Contents

<b>1</b>	<b>List of Important Variables</b>	<b>2</b>
<b>2</b>	<b>Introduction</b>	<b>3</b>
2.1	The Large Hadron Collider . . . . .	4
2.1.1	General . . . . .	4
2.1.2	Experiments . . . . .	5
2.2	The CMS Experiment . . . . .	6
2.2.1	General . . . . .	6
2.2.2	Structure . . . . .	6
2.2.3	Collection and Selection of Data . . . . .	8
<b>3</b>	<b>Calculations</b>	<b>9</b>
3.1	General . . . . .	9
3.1.1	Lepton Production . . . . .	9
3.1.2	The Cone of a Lepton . . . . .	11
3.2	Calculation of Weights of Simulated Events . . . . .	12
3.3	Calculation of Isolations . . . . .	12
3.3.1	Particle Flow Reconstruction . . . . .	12
3.3.2	Standard Reconstruction . . . . .	13
3.3.3	Pileup correction . . . . .	13
3.4	Calculation of Efficiency and Purity . . . . .	14
3.5	Calculation of Errors . . . . .	15
<b>4</b>	<b>Representative Event</b>	<b>16</b>
4.1	Event selection . . . . .	16
4.2	Detector Overview . . . . .	16
4.3	Event Overview . . . . .	17
4.4	Charged Hadron Isolation . . . . .	18
4.5	Neutral Hadron Isolation . . . . .	19
4.6	Photon Isolation . . . . .	20
<b>5</b>	<b>Results</b>	<b>21</b>
5.1	General . . . . .	21
5.2	Isolations . . . . .	21
5.3	Plots . . . . .	23
<b>6</b>	<b>Conclusion</b>	<b>24</b>
<b>A</b>	<b>Isolations</b>	<b>26</b>
<b>B</b>	<b>Plots</b>	<b>34</b>

# 1 List of Important Variables

- $\phi$ : Azimuthal angle in the detector
- $\theta$ : Polar angle in the detector
- $\eta$ : Pseudorapidity (see Eq. 1)
- $\Delta R$ : Cone size parameter (see Eq. 3)
- $w^i$ : Weight for events of type i (see Eq. 4)
- $\sigma^i$ : Cross section for process i
- $L$ : Luminosity
- $N_m^i$ : Number of simulated events of the type i
- $P_{t,x}$  [GeV]: Transversal momentum of a particle of type  $x$
- $I_x$  [GeV]: Lepton's isolation of particles of type  $x$  (see Eq. 5)
- $I_{rel}^x$ : Lepton's relative isolation constructed using method  $x$  (see Eq. 6 and 7)
- $\sigma_x$ : Uncertainty of variable  $x$
- $\cancel{E}_T$  [GeV]: Missing transversal energy (for example due to not detected neutrinos)
- $H_T$  [GeV]: Scalar sum of transversalrelative isolation momenta of jets above a certain threshold

## 2 Introduction

If one compares an electron positron collider, such as the LEP, to a proton proton collider, like the LHC, many fundamental differences can be observed. Electrons and positrons are elementary particles and do not possess inner structure. Therefore the total energy induced by the collision is contained in the created photon or Z boson and the resulting energy distribution of created particles is a very narrow line. Thus electron positron colliders are used to do experiments with a well defined momentum transfer, but are not well suited in discovery mode, when the energy dependend cross section is unknown. Hence proton proton colliders are used for this purpose.

Each proton involved in the collision is composed of three valence quarks, gluons and virtual  $q\bar{q}$  pairs (sea quarks). Each of these constituents, which are collectively referred to as partons, carries a fraction of the momentum and energy of the proton. These quantities are characterized by the parton distribution function which can be computed to a large extent in the framework of perturbative QCD and which is improved experimentally.

Interactions which only include QCD couplings are by far the most due to their huge cross section compared to other processes. Initially interactions between partons release a lot of energy and result in a lot of created partons. Since the strong interaction can be described via an asymptotical free theory, these first interactions can be calculated perturbatively, due to their high energy. But the more particles are created this way and the less energy is available for each of them, the bigger the strong coupling constant becomes, and non-perturbative physics become important. These circumstances lead to incalculable hadronic particle showers known as hadronization which is simulated by specifically tuned heuristic models. In analysis, these sprays of particles are clustered to jets, which should be kinematically similar to the Leptons produced by QCD events are either "fake"-leptons, misinterpreted as leptons, or leptons indirectly produced by QCD events (e.g. via  $b$ -jets), although they do have a small cross-section compared to other strong interacting processes. Since they are produced in jets, QCD events contain very few isolated leptons. Also, QCD events generally do not contain missing energy as there are no neutrinos or other undetectable particles created, but "fake" missing energy is sometimes reconstructed, due to detector imperfections.

To characterize the lepton isolation, different methods can be used. While the Standard reconstruction method just analyzes the energy deposits in the respective detectors, the better and more accurate method is the Particle Flow reconstruction method, where energy deposits are first linked to a reconstructed particle, which is then analyzed.

The following thesis presents a method for filtering events based on their isolation values in order to extract well defined event-samples, which then can be used for further analysis. By using simulated data the functionality of this method could be analyzed more thoroughly.

Many thanks to the advisor of this thesis, Dr. Robert Schöffbeck, for his support and guidance through the many highs and lows of this project.

## 2.1 The Large Hadron Collider

### 2.1.1 General

The LHC (**L**arge **H**adron **C**ollider) situated at the CERN research facility near Geneva is a superconducting proton-proton beam collider. It uses two parallel, circular, concentric beamlines, which intersect at four points. At these interaction points the proton beams, which circulate in the beamlines in opposing directions, are brought to collision.

The collider was under construction from 1998 to 2008 and it is built in the tunnel formerly used by the LEP (**L**arge **E**lectron-**P**ositron Collider) experiment. It is located about 100 m underneath of Geneva with a perimeter of approximately 28.6 km. It is the biggest part of the CERN complex, using several experiments for preacceleration and various other experiments for detection and analysis of the resulting collisions. The LHC has been in operation since 2008 and will be presumably run until 2030.

Usually one month a year heavy ion collisions are executed. These mostly include lead ion collision, which are evaluated primarily at ALICE.

In order to achieve the necessary energy of 450 GeV, protons are preaccelerated before being injected into the LHC. After being obtained from a hydrogen source they are accelerated up to 50 MeV in the linear accelerator LINAC 2 (**L**INear **A**Ccellerator **2**). By then travelling through the PSB (**P**roton **S**ynchrotron **B**ooster), the PS (**P**roton **S**ynchrotron) and the SPS (**S**uper **P**roton **S**ynchrotron) they gain momentum and reach the energy of 450 GeV required for injection into the Large Hadron Collider. In the current design total collision energies of up to 14 TeV are possible. The LHC and the most important associated experiments are shown in Fig. 1.

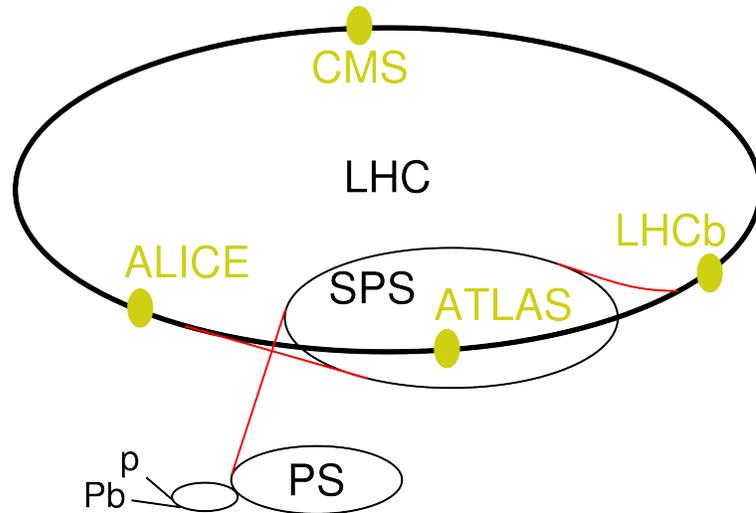


Figure 1: A schematic overview of the LHC experiments (created by Arpad Horvath)

To keep the particles from colliding with the walls of the beamlines 1232 superconducting dipole magnets and 392 quadrupole magnets, operating at 8.3 T, are used. The magnets' operating temperature of 1.9 K is maintained by using liquid helium. Furthermore a very powerful trigger system distinguishes between undesired low energetic processes and processes with a high energy transfer.

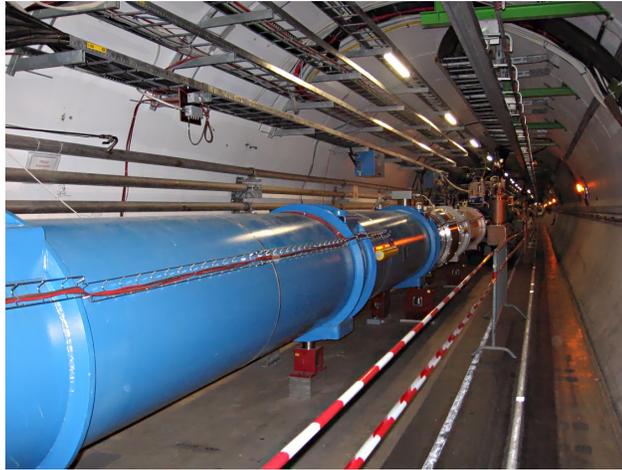


Figure 2: View of the pipe containing the beamlines (created by Julian Herzog)

The LHC Computing Grid is designed to process the 15 PB produced by the LHC and LHC-related simulations per year. It consists of private and public high-speed networks transferring data from CERN to academic institutions worldwide.

### 2.1.2 Experiments

The proton beams can be focused at four different locations, at which seven experiments, which also serve as detectors, are situated:

- ALICE (**A** **L**arge **I**on **C**ollider **E**xperiment) is located at one of the intersection points, where lead ions collide at high energies resulting in very high temperatures and generating quark-gluon plasmas, which then are studied.
- The ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) experiment is also located at a beamline intersection and was designed to contribute research data for different areas of physics. The most important of these areas are the search for the Higgs boson, physics beyond the standard model, CP violation and properties of the top quark and the W boson.
- The CMS (**C**ompact **M**uon **S**olenoid) detector has similar goals as the ATLAS experiment, but also investigates other topics such as characteristics of heavy ion collisions. It is also situated at a beam collision point.
- The LHCb (**L**arge **H**adron **C**ollider **b**eauty) experiment, which is too positioned at an interaction point, specialises in studying CP violation in the processes involving hadrons containing a bottom (beauty) quark.
- The LHCf (**L**arge **H**adron **C**ollider **f**orward) experiment is placed near the ATLAS experiment and is supposed to gather information for calibrating cosmic ray detectors and particle detectors in general.
- The MoEDAL (**M**onopole and **E**xotics **D**etector **A**t the **L**HC) is an addition to the LHCb experiment with the primary purpose of finding a magnetic monopole.

- The TOTEM (TOTal Elastic and diffractive cross section Measurements) experiment is used for the measurement of cross sections, elastic scattering and diffractive processes. It is situated near the CMS experiment.

The reason for the similarities in the research areas of the CMS experiment and the ATLAS experiment is the achievement of higher certainties of the results by letting the two detectors control and complement each other. [6, 10]

## 2.2 The CMS Experiment

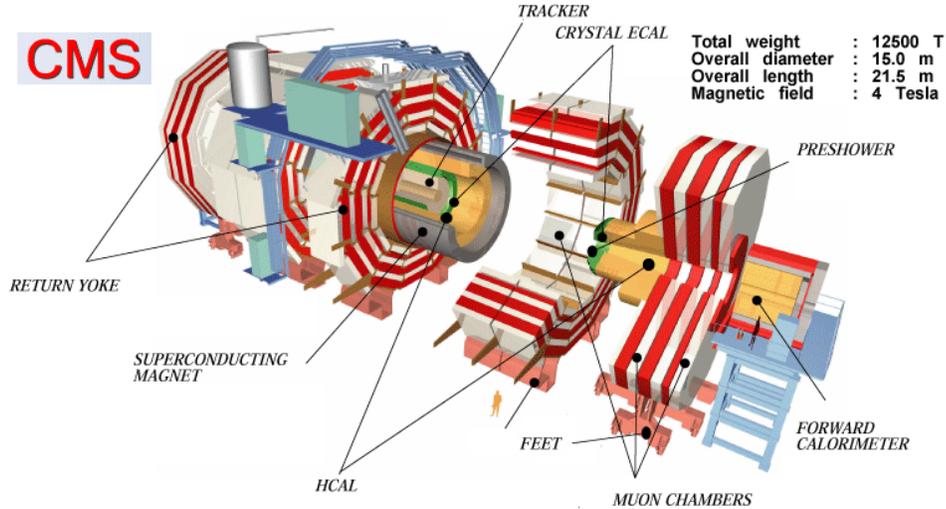


Figure 3: Overview and most important informations of the CMS detector (created by the CMS Collaboration)

### 2.2.1 General

As data examined in this thesis was taken from the CMS detector or an algorithm simulating the CMS detector, a short description of this experiment is necessary.

The Compact Muon Solenoid is one of seven detector experiments at the LHC and is designed to detect all stable and quasi stable particles created by the collisions at center-of-mass energies of up to 14 TeV except for neutrinos. However the presence of neutrinos can be indicated by missing transverse energy  $\cancel{E}_T$ . The CMS gathers data about the particles' tracks, energies and momenta.

It is built around one of the beamline intersections, where the two proton beams, which are focused to a radius of only  $17 \mu\text{m}$ , collide at an angle of  $285 \mu\text{rad}$ .

### 2.2.2 Structure

In order to understand the structure of the CMS it is best to explain the detector layer by layer. There are five different layers for the purpose of particle detection that cover the barrel and the endcap regions to obtain an angle resolution of  $4\pi$ . Each of them has different kinds of detectors. In Fig. 4 the profile of the CMS detector illustrating the various layers can be seen.

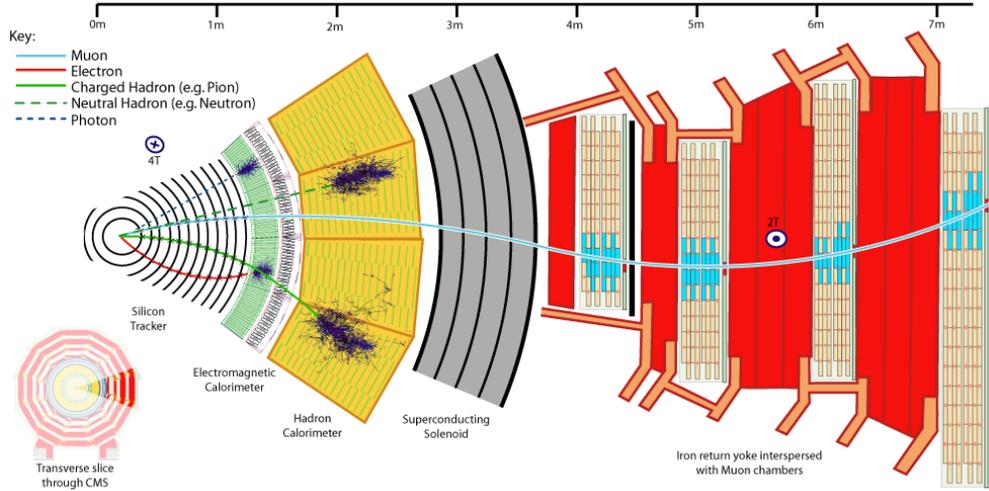


Figure 4: A slice through the CMS detector (created by the CMS Collaboration)

- **The Silicon Tracker**

The tracker is located immediately around the collision point and utilises silicon trackers, which have faster response times than gaseous detectors, a very essential feature in the LHC environment. With a detector region of more than  $200\text{ m}^2$  it is the world's largest silicon detector.

Comprising a central barrel region of 13 detection layers and an endcap region of 14 detection layers this detector is used for identifying particle tracks and secondary vertex finding, which links detected particles to their point of origin.

Particle tracks can be measured in a range of  $0 < \phi < 2\pi$  in the azimuth angle and  $|\eta| < 2.5$ , where  $\eta$  is the pseudorapidity defined by Eq. 1 with the polar angle  $\theta$  of the particles' track. (See section 3.1.2.)

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \quad (1)$$

The spatial resolution is approximately  $20\ \mu\text{m}$  in  $z$  direction, and about  $10\ \mu\text{m}$  in  $\phi$  direction.

Information gathered by this tracker is important for distinguishing between different particle types and measuring charge and momentum.

- **The Electromagnetic Calorimeter (ECAL)**

The ECAL is a calorimeter designed to measure energies of particles that interact via the electromagnetic force, such as electrons and photons. Crystals with a very short radiation length are used to slow down these particles, which lose their energy via ionisation, and simultaneously act as scintillators.

These crystals are positioned in a matrix of carbon fibres to isolate them optically and are supported by photodiodes, which analyse the reemitted light of the scintillators.

Muons or other heavier particles, which interact electromagnetically, can not be detected due to their big mass.

- **The Hadronic Calorimeter (HCAL)**

Similarly the HCAL measures energies of particles interacting via the strong force, such as protons, neutrons and mesons.

In this calorimeter dense layers of brass serve as absorbers while plastic scintillators and photodiodes determine the absorbed energy.

The barrel HCAL covers  $|\eta| < 1.3$  while the endcap HCAL covers  $1.3 < |\eta| < 3$  and an additional outer HCAL, which is located outside of the muon detectors, covers  $2.9 < |\eta| < 5$ . These characteristics enable a sufficient sealing of the interaction region thus making it possible to determine events with missing energy  $\cancel{E}_T$ .

- **The Superconducting Solenoid Magnet**

The tracker and the calorimetry are compact enough to fit inside this solenoid magnet, which is 13 m long and 7 m wide and is able to produce a 4 T magnetic field.

By measuring the deflection of charged particles in this magnetic field identifying the  $q/m$  ratio is possible.

- **The Muon Detectors**

Since muons can not be detected using calorimetry, a separate detector is required. This detector consists of drift tubes and resistive plate chambers in the barrel region ( $|\eta| < 1.2$ ) and cathode strip chambers and resistive plate chambers in the endcap region ( $0.9 < |\eta| < 1.4$ ). Using this detector, muons and their momenta can be measured with a muon reconstruction efficiency of 95-99%.

### 2.2.3 Collection and Selection of Data

Due to the high luminosity and the therefore extremely high collision rate of about  $10^9 \text{ s}^{-1}$  not every event, especially not events with a low momentum transfer are stored for further analysis. therefore a trigger has to separate undesired “soft processes” with a low momentum transfer from “hard processes” with a high momentum transfer. This is done in two steps by applying a powerful software trigger after first using a hardware trigger.

The “Level 1” hardware based trigger analyses collision data, which is temporarily stored in buffers, and tries to recognise certain characteristics such as high energy jets or missing energy. This process decreases the number of resulting events per second to an order of magnitude of  $10^5$ .

The software based HLT (**H**igh **L**evel **T**rigger) reduces this number even further so that every second ultimately around 100 events are stored for more analysis. The software for this trigger, which is mainly programmed in C++, runs on server farms and reconstructs essential variables such as lepton momentum or jet energies and dismisses events based on cuts applied to these variables. [1, 3]

# 3 Calculations

## 3.1 General

### 3.1.1 Lepton Production

Collisions at the LHC can produce leptons (in this case electrons) by several processes, of which the most important are listed below. As leptons originating from a QCD event do not contain interesting properties, they were categorized as background. Leptons created by the other processes listed below were used as signal events.

- $t\bar{t}$ -jets: Leptons coming from  $t\bar{t}$ -jets are created in processes caused by the strong interaction in collisions of two protons. The production mechanism is as follows:

$$pp \rightarrow t\bar{t} \rightarrow W^+bW^-\bar{b} \rightarrow l\nu_l b\bar{b}q\bar{q} \quad (2)$$

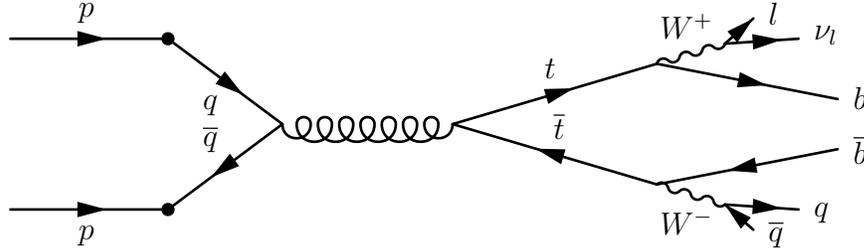


Figure 5: Feynman diagram of a typical  $t\bar{t}$ -jet event

Muons and electrons produced this way are only a small amount of the overall created particles, as only 15% go into the  $e^-$  channel and 15% go into the  $\mu^-$  channel, but 45% go into the  $q\bar{q}$  channel. In Fig. 5 the characteristics of such a processes can be seen:

- There exist four jets of which two are heavy quark jets ( $b\bar{b}$ -jets) and two are light quark jets
  - One lepton (muon or electron) is created.
  - Due to the undetectable neutrino missing energy  $\cancel{E}_T$  is present.
- $W$ +jets: As the name indicates these processes are composed of a  $W$ -boson together with jets. They have a similar signature as  $t\bar{t}$ -jets.

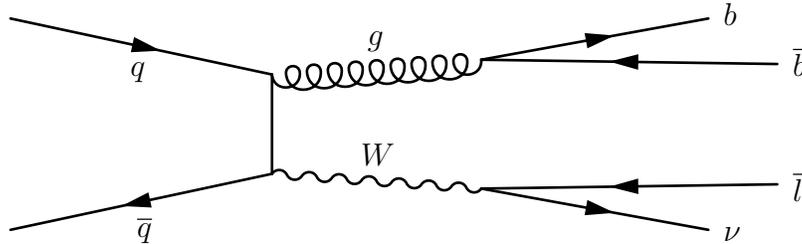


Figure 6: Feynman diagram of a possible  $W$ +jets event

- Single top processes: In a single top process a single top quark decays into a lepton, the corresponding neutrino and a heavy quark. The neutrino causes missing energy  $\cancel{E}_T$ .

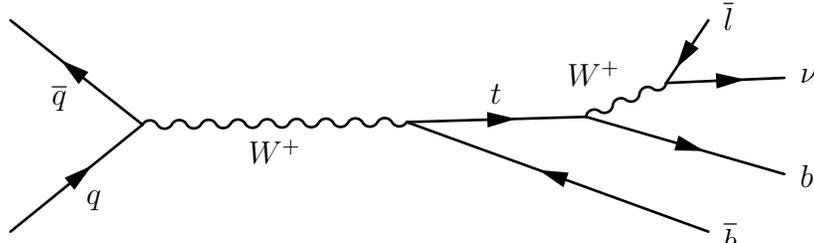


Figure 7: Feynman diagram of a possible single top event

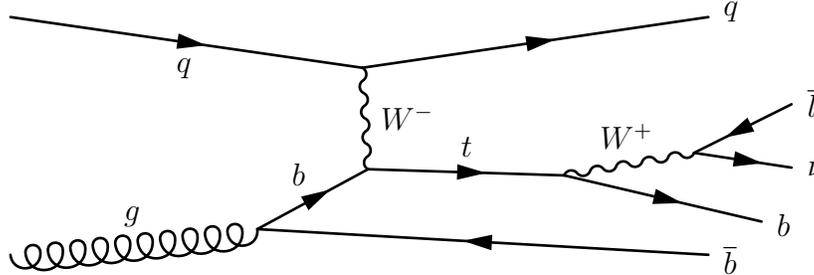


Figure 8: Feynman diagram of another possible single top event

- Drell-Yan processes: A quark and an antiquark annihilate forming a virtual photon or Z-boson which then decays into a pair of oppositely charged leptons. Due to the absence of Neutrinos, Drell-Yan processes do not have missing energy  $\cancel{E}_T$  and are therefore strongly reduced by applying an  $\cancel{E}_T$  cut. Nevertheless, not detected leptons, due to detector imperfections, can result in a detection of "fake"-missing energy.

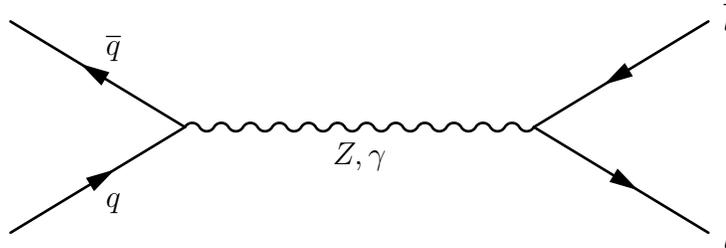


Figure 9: Feynman diagram of a Drell-Yan event

- QCD background: QCD events are generally produced via strong coupling. Leptons from QCD events are either "fake" leptons, which are signals in the detector, misinterpreted as leptons, or more rarely directly produced leptons from QCD events (e.g. from  $b$ -jets). They form a background process due to their comparatively high cross-section. Since each initial parton carries an momentum according to the PDF and since there is no intermediated high-mass state in these processes, final state particles of QCD events tend to be boosted in the forward (or backward) direction.

For the calculation of efficiency and purity it is necessary to distinguish between signal events and background events. Since the calculations were done with simulated events, it was possible to distinguish directly between signal events and background events. For more information see [5].

### 3.1.2 The Cone of a Lepton

In the following chapters many lepton related variables are defined by properties of particles, which reside within the cone of the lepton. This cone is located around the trajectory of the lepton and has a cone size parameter  $\Delta R$ , which is defined by Eq. 3.

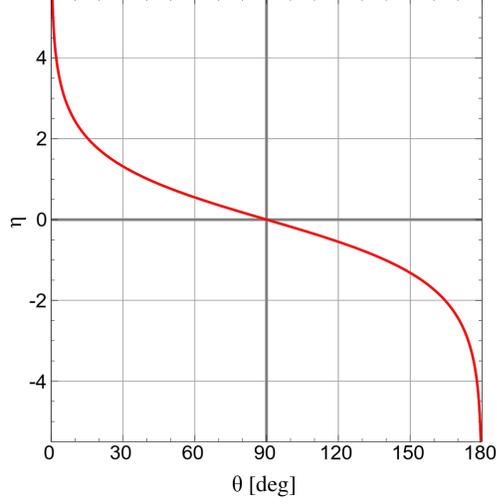


Figure 10:  $\eta$  plotted against  $\theta$

While the polar angle  $\theta$  is measured from the positive z-axis, the azimuthal angle  $\phi$  is measured in the x-y plane, as it can be seen in Fig. 11.

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \quad (3)$$

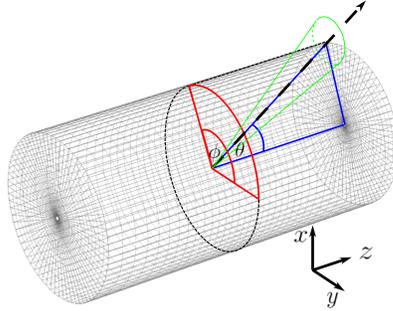


Figure 11: Definition of  $\phi$  (red) and  $\theta$  (blue) and an exemplified cone (green) for a trajectory (dashed) in the schematic tracker barrel

In Fig. 11 a schematic example of a cone can be seen. For the different reconstruction methods the cone is defined by different maximum values for  $\Delta R$ . The definition of  $\eta$  can be seen in Eq. 1 and the relation between  $\eta$  and  $\theta$  can be seen in Fig. 10. More information can be found in the first sections of [7] and [9].

## 3.2 Calculation of Weights of Simulated Events

Events passing the trigger are saved for further analysis, which is performed using the software framework CMSSW (Compact Muon Solenoid SoftWare).

In order to analyse the measured processes and their contributions more thoroughly, simulations of events are produced using the Monte Carlo method. Unfortunately due to the limited computing power, it is not possible to produce enough events so that the results can be compared to the real data. therefore simulated events, multiplied with weights calculated from the corresponding cross sections  $\sigma^i$  taken from literature, the Luminosity  $L$  and the simulated number of events  $N_m^i$  are used.

The weight  $w^i$  of a simulated event is given by:

$$w^i = \frac{\sigma^i L}{N_m^i} \quad (4)$$

The final number of events is then given by summing up events multiplied with their corresponding weight and should match the number of actual events taken from real data.

## 3.3 Calculation of Isolations

### 3.3.1 Particle Flow Reconstruction

A particle generally gives rise to several Particle Flow elements in the respective detectors: calorimeter deposits (ECAL/HCAL) and/or charged particle tracks. These deposits have to be connected by a linking algorithm to reconstruct every single particle and to avoid double counting in different detectors. This reconstruction method is done in several steps: [8]

- First “Particle Flow” muons are defined after corresponding global muons in the muon chambers are identified. The associated trajectories of the Particle Flow muons in the tracker are then removed for the following reconstruction process.
- Subsequently Particle Flow electrons are reconstructed. The requirements for a Particle Flow electron are trajectories in the tracker, connected with an energy deposit in the ECAL. After the identification, the electrons and additional photons produced by bremsstrahlung, determined via an additional algorithm using deposits in the ECAL, are removed.
- For the remaining tracks tighter quality criteria are applied: The relative uncertainty on the measured  $P_t$  has to be smaller than the expected energy resolution for charged hadrons. This requirement dismisses about 0.2% of tracks. Now deposits in the HCAL and corresponding tracks and ECAL deposits are searched for. These are collectively identified as Particle Flow charged hadrons and are afterwards removed.
- Remaining energy deposits in the HCAL or ECAL are then respectively identified as Particle Flow neutral hadrons or Particle Flow photons, depending if the energy deposit is found in the HCAL for neutral hadrons, or in the ECAL for photons.

In Particle Flow reconstruction the isolations of a lepton are then determined by the transverse momenta  $P_{t,x}$  of the corresponding surrounding particles, within the cone of a

given size. Thus, the exact formulae for the charged hadron isolation  $I_c$ , the neutral hadron isolation  $I_n$  and the photon isolation  $I_p$  of a lepton are as given in 5, where only the particles within the cone are considered in the summation.

$$I_c = \sum_i P_{t,c}^i \quad I_n = \sum_i P_{t,n}^i \quad I_p = \sum_i P_{t,p}^i \quad (5)$$

The relative isolation  $I_{rel}$  is then just the overall isolation in comparison to the  $P_{t,l}$  of the lepton.

$$I_{rel} = \frac{I_c + I_n + I_p}{P_{t,l}} \quad (6)$$

### 3.3.2 Standard Reconstruction

In the usual event reconstruction method, which is here called Standard reconstruction, the isolations of a lepton are calculated differently. In this method the relative isolation is not calculated by summing up  $P_{t,x}$  of reconstructed particles within the cone, but by computing the ratio of the sum of the total energy deposited in the hadronic and the electromagnetic calorimeter within the cone ( $I_{hcal}$  and  $I_{ecal}$ ) and the total transverse momenta within the cone detected by the tracker ( $I_{trac}$ ) to the transverse momentum  $P_{t,l}$  of the lepton. In this calculation the contributions of the lepton to the calorimeter must be ignored.

Therefore the equation to calculate  $I_{rel}$  for this reconstruction method is:

$$I_{rel} = \frac{I_{hcal} + I_{ecal} + I_{trac}}{P_{t,l}} \quad (7)$$

But the disadvantage of this method in comparison to the Particle Flow reconstruction is, that it is possible for the contribution of one particle to be counted multiple times. A charged hadron, for example, can deposit energy in the hadronic and the electromagnetic calorimeter and shows up in the tracker.

Therefore Particle Flow reconstruction generally is the better choice.

### 3.3.3 Pileup correction

The definitions of isolation would be correct in ideal circumstances for a lepton created by a single vertex, but particles may also be created in other vertices, called “pile-up”, which distort the isolation measurement. The energy from additional collisions, occurring close enough in time to be included in the calorimeter energy for the lepton producing event, are referred to as pileup. Particles with tracks associated to a PU-vertex are removed, but residual neutral deposits need to be corrected for. This correction thus increases efficiency, since it reduces the energy (isolation) within the cone. These so called pileup corrections have become increasingly important since the number of additional vertices increases with higher energies. The pileup correction is not done for standard reconstruction, because the errors induced by double counting energies in standard reconstruction, are producing bigger errors, than pileup does. Therefore the following correction is only done for particle flow reconstruction. For jets, pileup is corrected using the so called Jet-Area-Correction [4]:

For each event an average  $P_t$ -density  $\rho$  per jet-area is estimated which characterizes the soft jet activity and is a combination of the underlying event, the electronics noise, and the pileup.

This jet area, measured on the rapidity ( $\eta$ ) and azimuth ( $\phi$ ) cylinder, is insofar a non-trivial concept since jets consists of pointlike particles which themselves do not have a specific area. This jet area ( $A$ ) is different for each jet and depends on the substructure of the underlying event. Given a reasonable definition of the jet area (the exact calculations are not carried out here, since they are irrelevant for further considerations), the modification of the jet's transverse momentum can be shown to be:

$$\Delta P_t = A\rho \pm \sigma\sqrt{A} - L \quad (8)$$

where  $\rho$  the level of diffuse noise refers to the momentum added per unit area.  $\sigma$  is the standard deviation of the noise when measured across many regions of unit area. The first term is therefore the geometrical contamination with an uncertainty (second term), while the third term accounts for the occasional loss of part of the jet's content.

The correction is now based on the assumption that the uncertainties are small  $\sigma \ll \sqrt{A}\rho$  and that the losses ( $L$ ) can be neglected compared to the geometrical contamination. Therefore each measured jet  $j$  has to be corrected via the subtraction:

$$P_t^j = P_t^j - A\rho \quad (9)$$

This procedure can now be used to correct the isolation of electrons.

$$I_{rel} = \frac{I_c + I_n + I_p - \rho A}{P_t} \quad (10)$$

To correct the isolation of muons, another method, the so called  $\Delta\beta$  correction is used, where one estimates the residual deposits, using the charged deposits weighted by a suitable constant. This constant has been shown to be 0.5, therefore following method of corrected particle flow isolation for muons is used in the CMS detector [2]:

$$I_{rel} = \frac{I_c + I_n + I_p - 0.5 \sum PUPt}{P_t} \quad (11)$$

### 3.4 Calculation of Efficiency and Purity

These quantities are calculated in the same way for Particle Flow reconstruction and Standard reconstruction. The purity  $P$  (not to be confused with  $P_t$ ) is given by the number of signal events  $n_S$  divided by the sum of the number of background events  $n_{BG}$  and the number of signal events  $n_S$ ,

$$P = \frac{n_S}{n_{BG} + n_S} \quad (12)$$

In order to get less of the undesired events (QCD-events) and therefore increase  $P$ , cuts respective to  $I_{rel}$  are applied. The number of events after the cut will be denoted by  $n'_x$ . By dividing  $n'_S$  by  $n_S$  we gain the efficiency  $E$ , which characterizes how many signal events are lost,

$$E = \frac{n'_S}{n_S} \quad (13)$$

Of course  $P$  is also affected by the cut and can be calculated by replacing  $n_x$  by  $n'_x$  in Eq. (12). This process of cutting is then repeated with different cut-parameters and the corresponding  $E$  and  $P$  are calculated.

### 3.5 Calculation of Errors

Since we are expecting non-correlated events it is necessary to use a Poisson distribution given by

$$P(X = k) = \frac{\langle n \rangle^k e^{-\langle n \rangle}}{k!}. \quad (14)$$

In Poisson statistics the uncertainty  $\sigma_{n_x}$  of the particular  $n_x$  is given by

$$\sigma_{n_x} = \sqrt{n_x}, \quad (15)$$

But as the number of events always consists of the numbers of events in different bins, which have different weights  $\omega_i$  the equation for  $\sigma_{n_x}$  for a whole sample becomes

$$\sigma_{n_x} = \sqrt{\sum_{bins} n_{x,bin} \omega_i^2}. \quad (16)$$

The propagation of the uncertainties in  $P$  and  $E$  are calculated in (15) and (16).

$$\sigma_P = \sqrt{\left(\frac{\partial P}{\partial n'_S} \sigma_{n'_S}\right)^2 + \left(\frac{\partial P}{\partial n'_{BG}} \sigma_{n'_{BG}}\right)^2} = \sqrt{\left(\frac{n'_{BG} \sigma_{n'_S}}{(n'_{BG} + n'_S)^2}\right)^2 + \left(\frac{n'_S \sigma_{n'_{BG}}}{(n'_{BG} + n'_S)^2}\right)^2} \quad (17)$$

$$\sigma_E = \sqrt{\left(\frac{\partial E}{\partial n'_S} \sigma_{n'_S}\right)^2 + \left(\frac{\partial E}{\partial n_S} \sigma_{n_S}\right)^2} = \sqrt{\left(\frac{\sigma_{n'_S}}{n_S}\right)^2 + \left(\frac{n'_S \sigma_{n_S}}{n_S^2}\right)^2} \quad (18)$$

The values calculated by equations in 3.3, 3.4 and 3.5 for different cuts were used for the plots and histograms in 5.2 and 5.3.

To distinguish between the Particle Flow method and Standard method, calculated quantities will from now on be denoted by the superscripts  $^{pf}$  and  $^{st}$ .

## 4 Representative Event

### 4.1 Event selection

In this section the different Particle Flow isolations for a representative event will be calculated. A simple algorithm, which assured that all isolations were relatively high, chose the respective event. After the important data of this event were gathered, it was identified by its Run-, Lumi- and Event-Number ( $r/l/e$ ) and analysed using the “Fireworks”-tool for visualisation.

### 4.2 Detector Overview

In figure 12 the whole detector, as it is displayed in the “Fireworks”-environment, can be seen with all details enabled. In order to analyse the events only the tracker barrel was viewed from close up without any details.

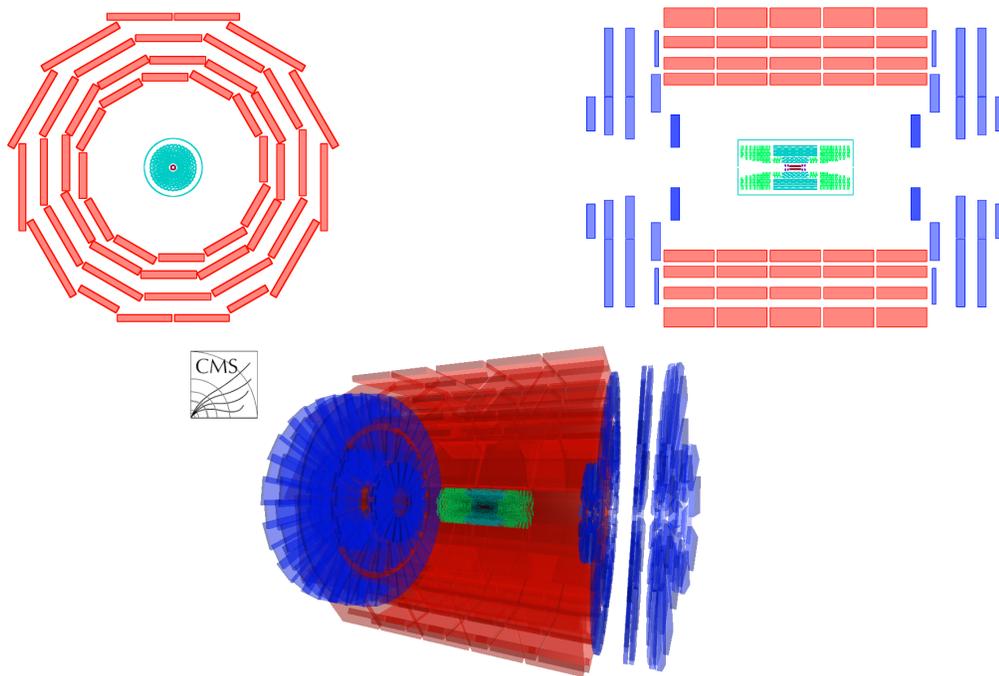


Figure 12: Overview of the CMS detector

### 4.3 Event Overview

The selected event was taken from the root file situated at `/scratch/trauner/MCsamples/Summer11_QCD_Pt_150_MuPt5Enriched_TuneZ2_7TeV_pythia6_PU_S4_START42_V11_v1_AODSIM.root` on the server of the “Institut für Hochenergiephysik” (HEPHY).

An overview of this event showing all important particles can be seen in figure 13, where muons are coloured red, charged hadrons are coloured blue, neutral hadrons are coloured orange and photons are coloured teal. The essential data are recorded in Tab. 1.

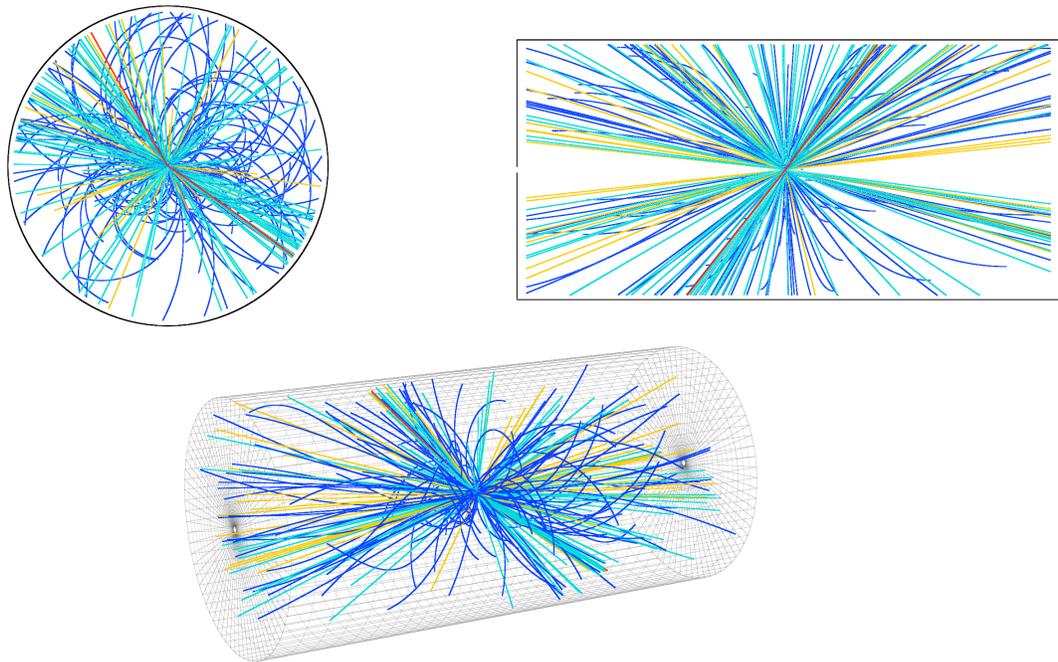


Figure 13: Overview of the selected event

$r/l/e$	$I_{rel}^{pf}$	$I_c^{pf}$	$I_n^{pf}$	$I_p^{pf}$	$P_{t,l}$
1/1782/887720	2.06	15.35	19.35	33.72	33.20

Table 1: Data of the selected event

In the following sections the particles within the cone were identified for charged hadrons, neutral hadrons and photons. They were then labelled using the nomenclature of the “Fireworks”-environment and the corresponding isolation was calculated using the particles’  $P_{t,x}$ , which were taken from the particles’.

By inserting the required values from the Tab. 2, 3, 4 and 1 in Eq. 6  $I_{rel}^{pf}$  could be calculated. The result matches the value from Tab. 1.

## 4.4 Charged Hadron Isolation

The identified and labelled charged hadrons can be seen in figure 14 and their  $P_{t,c}$  with the resulting isolation can be seen in Tab. 2.

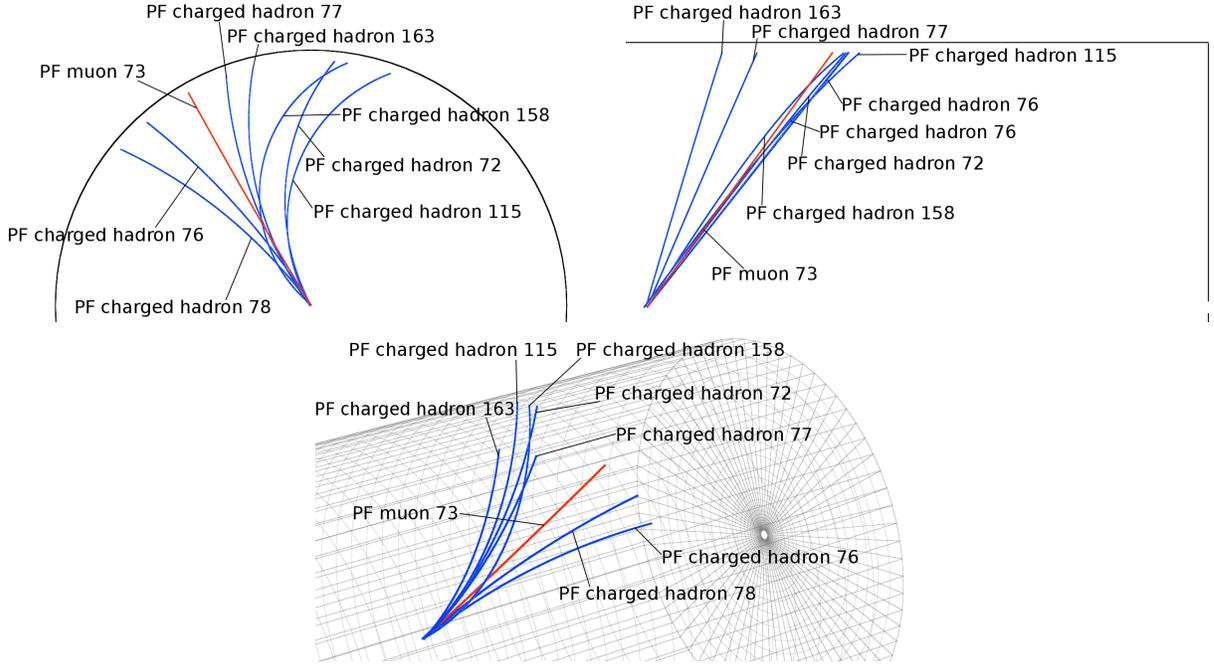


Figure 14: Tracks of charged hadrons (coloured blue) within the cone of the muon (coloured red) from different angles

Object Label	$P_{t,c}$
PF charged hadron 72	1.37
PF charged hadron 76	4.42
PF charged hadron 77	2.74
PF charged hadron 78	3.03
PF charged hadron 115	0.98
PF charged hadron 158	0.88
PF charged hadron 163	1.93
Sum of $P_{t,c}$ ( $I_c^{pf}$ )	15.35

Table 2:  $P_{t,c}$  of the relevant charged hadrons and the resulting isolation  $I_c^{pf}$

Determining the correct particles was difficult for charged hadrons, as their tracks are curved more or less strongly depending on their weight and speed due to the magnetic field.

## 4.5 Neutral Hadron Isolation

The identified and labelled neutral hadrons can be seen in figure 15 and their  $P_{t,n}$  with the resulting isolation can be seen in Tab. 3.

Object Label	$P_{t,n}$
PF neutral hadron 74	16.68
PF neutral hadron 353	2.67
Sum of $P_{t,n}$ ( $I_n^{pf}$ )	19.35

Table 3:  $P_{t,n}$  of the relevant neutral hadrons and the resulting isolation  $I_n^{pf}$

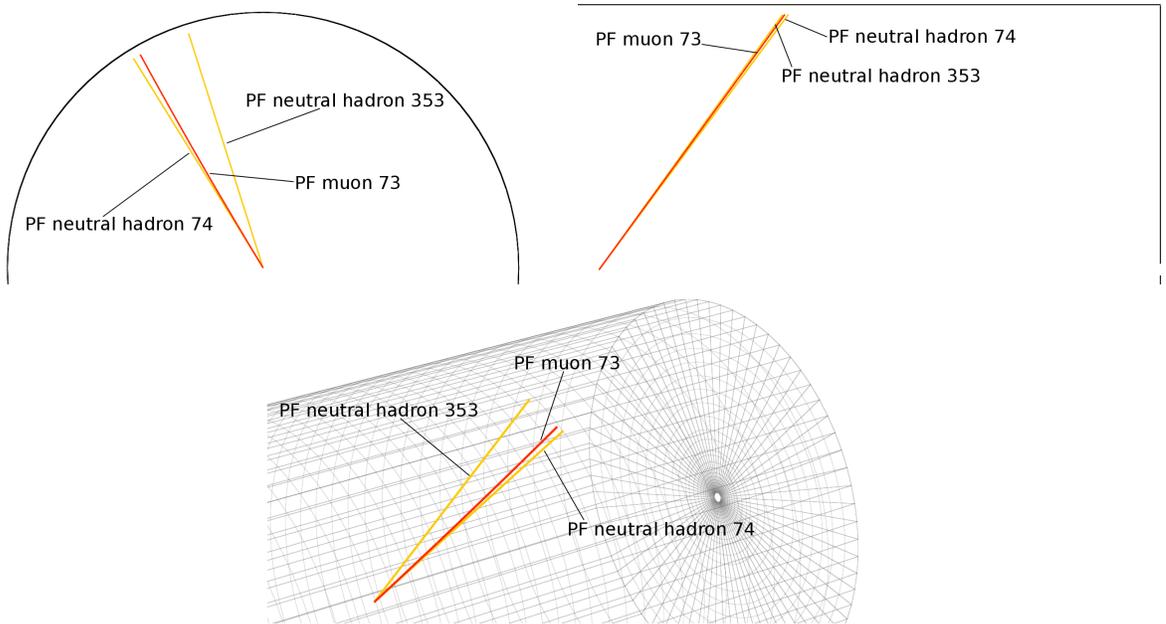


Figure 15: Tracks of neutral hadrons (coloured orange) within the cone of the muon (coloured red) from different angles

## 4.6 Photon Isolation

The identified and labelled photons can be seen in figure 16 and their  $P_{t,p}$  with the resulting isolation can be seen in Tab. 4.

Object Label	$P_{t,p}$
PF photon 75	11.92
PF photon 377	15.76
PF photon 378	1.03
PF photon 384	2.40
PF photon 394	0.89
PF photon 413	1.72
Sum of $P_{t,p}$ ( $I_p^{pf}$ )	33.72

Table 4:  $P_{t,p}$  of the relevant photons and the resulting isolation  $I_p^{pf}$

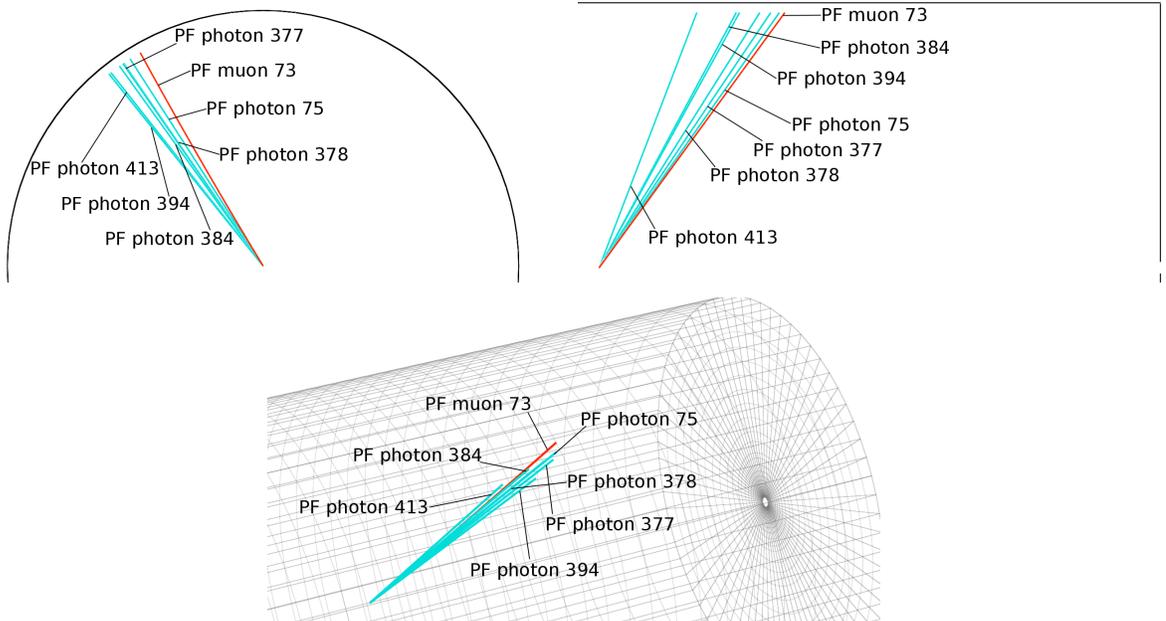


Figure 16: Tracks of photons (coloured teal) within the cone of the muon (coloured red) from different angles

# 5 Results

## 5.1 General

The simulated Monte Carlo data for Particle Flow reconstruction for the following plots and histograms were taken from the directory `/data/mhickel/pat_120917/mc8TeV/` on the server of HEPHY, while the simulated data for Standard reconstruction were taken from the directory `/data/trauner/pat_120518_PA/Mu/`. The reason for the two different sets of data is that due to unfortunate circumstances the first set does not contain Standard reconstruction values, while the second set does not contain the aforementioned correction for Particle Flow values. Real data for Particle Flow reconstruction were taken from the directory `/data/mhickel/pat_120927/data8TeV/` and for Standard Reconstruction the directory `/data/trauner/pat_120626/Mu/` was used. In all cases only the bins relevant for muons were selected.

The general cut was `'isopfRA4Tupelizer_singleMuonic && isopfRA4Tupelizer_jet2pt>40'` for Standard reconstruction and `'muonsisPF && muonsisGlobal && muonsPt>=20 && fabs(muonsEta)<=2.4 && muonsNormChi2<=10 && muonsNValMuonHits>0 && muonsNumMatchedStadions>1 && muonsPixelHits>0 && muonsNumtrackerLayerWithMeasurement>5 && fabs(muonsDxy)<0.02 && fabs(muonsDz)<0.5 && muonsPFDeltaPT<5 && jet2pt>40'` for Particle Flow reconstruction. These cuts correspond to tight muons. Furthermore various cuts  $\cancel{E}_T$  and  $H_T$  were applied. In section 5.3 a cut of 1.5 for the corresponding relative isolations of the different reconstruction methods was used, because the relative isolation of the Standard reconstruction had an inherent cut of 1.5 and therefore it would otherwise not be possible to compare the different reconstruction methods.

## 5.2 Isolations

For this section at first isolation values  $I_c^{pf}$ ,  $I_n^{pf}$  and  $I_p^{pf}$  were calculated for muons according to Eq. 5. Then the relative isolation was calculated using Standard reconstruction (Eq. 7) and Particle Flow reconstruction (Eq. 6) with and without pileup correction. For histograms of relative isolation with  $\cancel{E}_T > 150$  and  $H_T > 400$  real data were also added, because this cut is equivalent to the real data trigger and thus a comparison makes sense. The complete results can be seen in appendix A.

Below a set of results with  $\cancel{E}_T > 60$  and  $H_T > 400$  as an example. It can be seen that there is a cut of 1.5 on the relative isolation of the Standard reconstruction. Also it is obvious that QCD events generally have a higher relative isolation, which is why cuts on relative isolation are preferable to eliminate background events without reducing signal events.

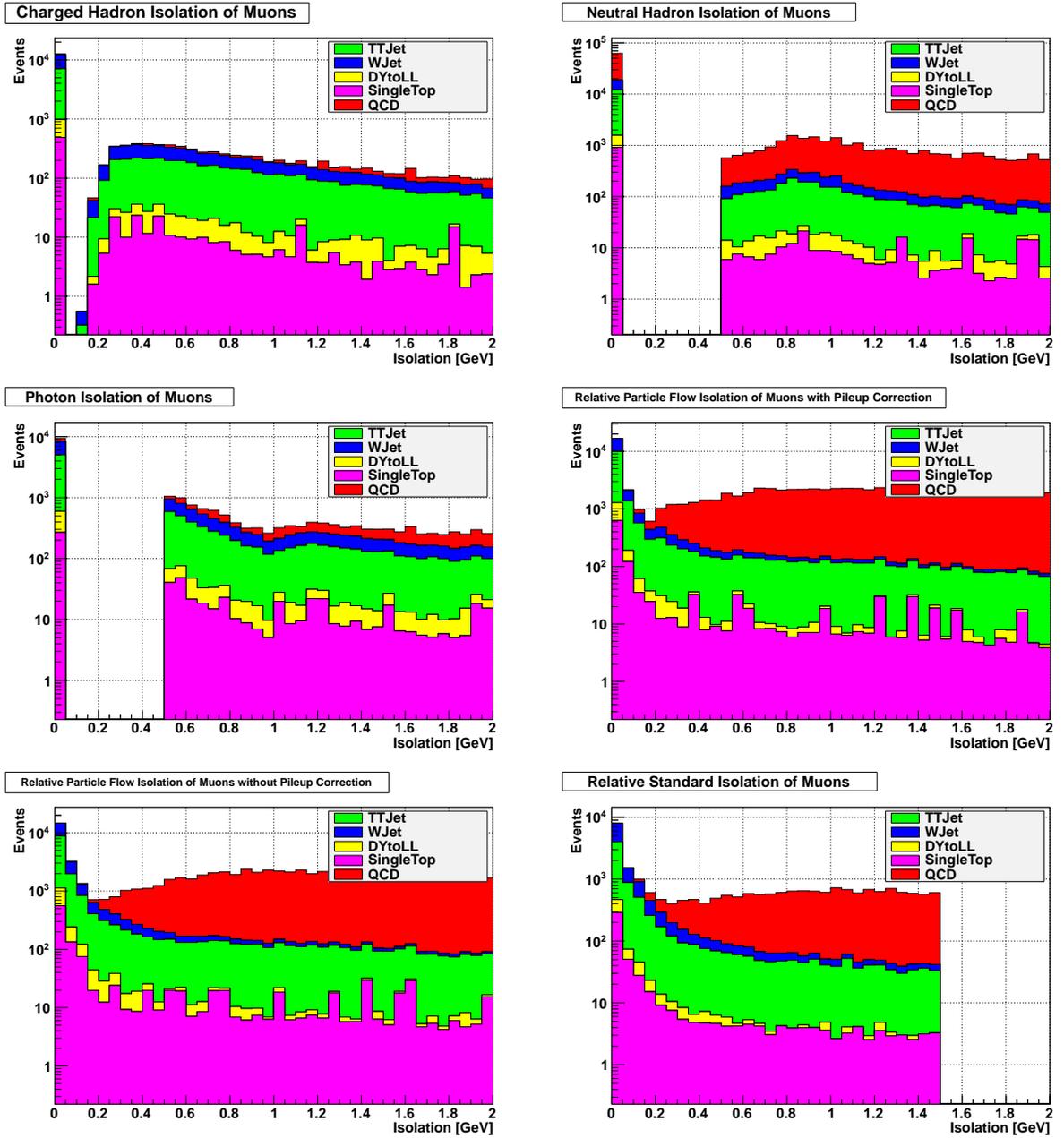


Figure 17: Isolation histograms with  $\cancel{E}_T > 60$  and  $H_T > 400$

### 5.3 Plots

In this section  $E$  and  $P$  were calculated for different cuts using Eq. 13 and 12. An algorithm identified equidistant points on the resulting curve and labelled them with their respective relative isolation cuts. An example plot with  $\cancel{E}_T > 60$  and  $H_T > 400$  can be seen below and the complete results can be seen in appendix B.

It is obvious that a tight relative isolation cut causes a high value of  $P$ , because leptons produced in QCD events generally have a higher relative isolation, but at the same time it results in a low value of  $E$ , because the number of events in general is reduced. A loose relative isolation cut causes the opposite situation.

As a high number of events and therefore a high value of  $E$  is necessary to calculate reliable results, but also the fraction of prompt leptons (which corresponds to  $P$ ) has to be high in order to execute meaningful calculations, a compromise between adequate statistics and the selection of good events has to be made. As it can be seen in the example relative isolation histograms of the previous section and the example plot below a relative isolation cut of approximately 0.2 is optimal as a tighter cut does not increase purity significantly while reducing efficiency rapidly.

The plots in appendix B show that cuts applied to  $\cancel{E}_T$  are very effective against QCD events, as they usually do not contain missing energy, but they are also quite effective against prompt leptons in general. This is why the  $E$  error bars become very large. Therefore a moderate  $\cancel{E}_T$  cut is preferable.

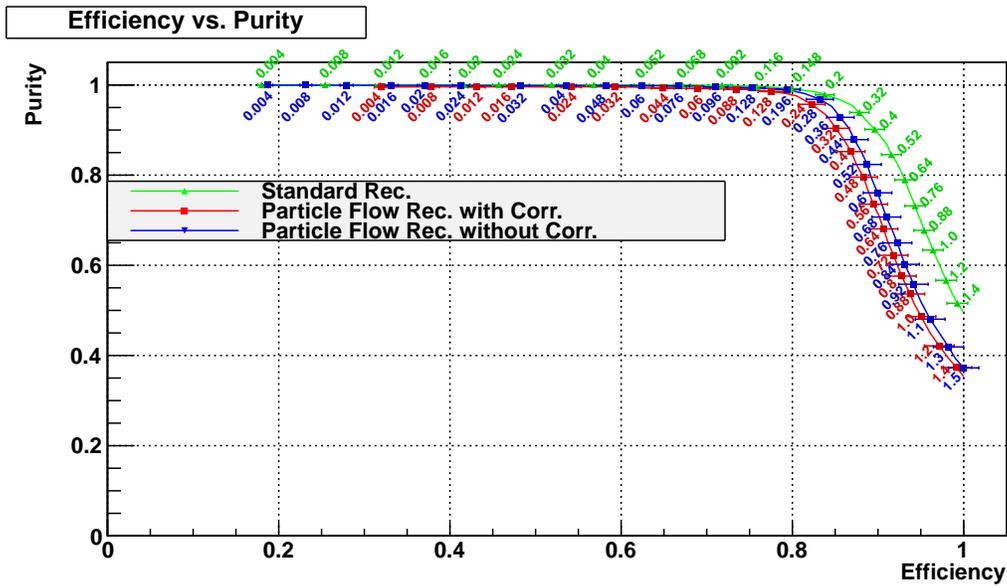


Figure 18: Efficiency plotted versus purity with  $\cancel{E}_T > 60$  and  $H_T > 400$

## 6 Conclusion

In order to coherently represent the immense number of particles created by the LHC and their signals captured by the detectors, efficient reconstruction techniques are necessary. The newer Particle Flow reconstruction aims to reconstruct individual particles and link them to their energy deposits in the CMS detector. Although there are still some minor flaws due to its recency, this method appears very promising, because it presents the data very comprehensibly and helps to simplify several computations. Furthermore, the different isolation values, which were defined in this thesis, provided the possibility of evaluating the quality of an event.

Since the LHC was designed to discover new particles with unidentified characteristics its precision is low and therefore a large part of the amount of produced particles is not just unusable, but rather constitutes a disruptive noise. In this thesis relative isolation cuts on simulated events were used to minimize this noise for muons and to make the desired data more accessible. As it is not possible to reduce the number of undesired events without also reducing the number of desired events, a compromise had to be made.

For this purpose the two parameters efficiency  $E$  and purity  $P$ , which are measures for the relative amount of desired and undesired events after the cut, were introduced. It was possible to determine the dependency of these two quantities on the applied relative isolation cut in different circumstances. Using these findings the corresponding optimized relative isolation cuts and therefore the optimized values of  $E$  and  $P$  could be identified. These results enable simple selection of optimized data sets for further calculation.

Hopefully the outcome of this thesis will be useful for the many endeavours, which still lie ahead in the LHC's bright future.

## References

- [1] Christian Adlef. *QCD Background Estimation in SUSY Signal Regions*.
- [2] CMS Collaboration. *Baseline muon selections*. 2012. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId>.
- [3] The CMS Collaboration. *CMS Physics Technical Design Report Volume I : Detector Performance and Software*. 2006.
- [4] M. Cacciari and G. P. Salam. “Pileup subtraction using jet areas”. In: *arXiv* 0707.1378v2 (2007).
- [5] Wolfer Peelaers. *Measurement of the Cross Section of  $t\bar{t}$  Production in the Semi-Leptonic Muon Channel at the LHC Using the CMS Detector*. 2008.
- [6] K. M. Potter. *The Large Hadron Collider (LHC) project of CERN*. 1996.
- [7] The CMS Collaboration. “Combined search for the quarks of a sequential fourth generation”. In: *arXiv* 1209.1062 (2012).
- [8] The CMS Collaboration. “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and  $E_T^{miss}$ ”. In: *CMS PAS PFT-09.001* (2009).
- [9] The CMS Collaboration. “Search for new physics with same-sign isolated dilepton events with jets and missing transverse energy at the LHC”. In: *arXiv* 1104.3168 (2011).
- [10] Inc. Wikimedia Foundation. “*LHC*” *Wikipedia, The Free Encyclopedia*. Sept. 2012. URL: <http://www.en.wikipedia.org/wiki/Lhc>.

# A Isolations

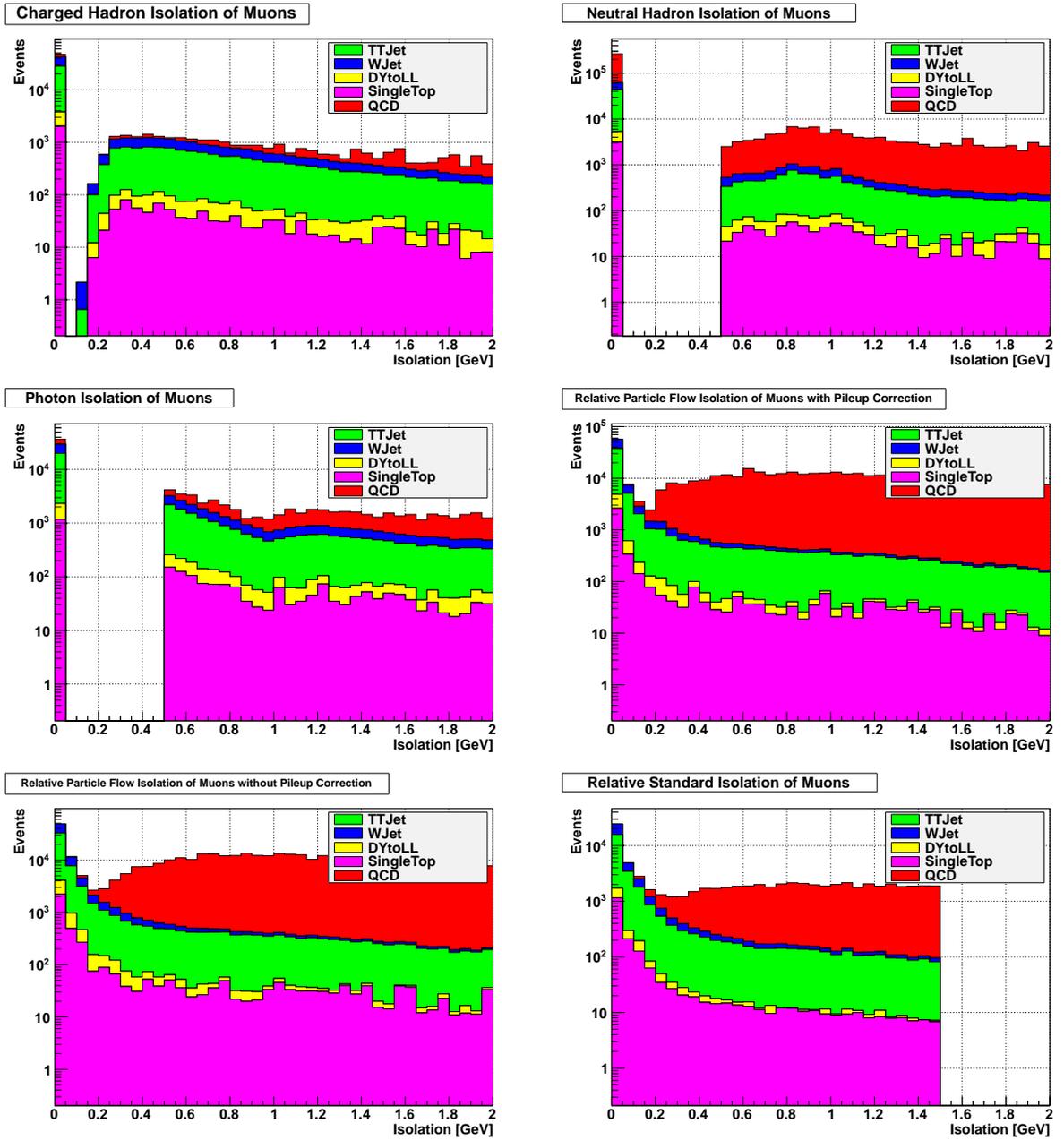


Figure 19: Isolation histograms with  $\cancel{E}_T > 60$  and  $H_T > 200$

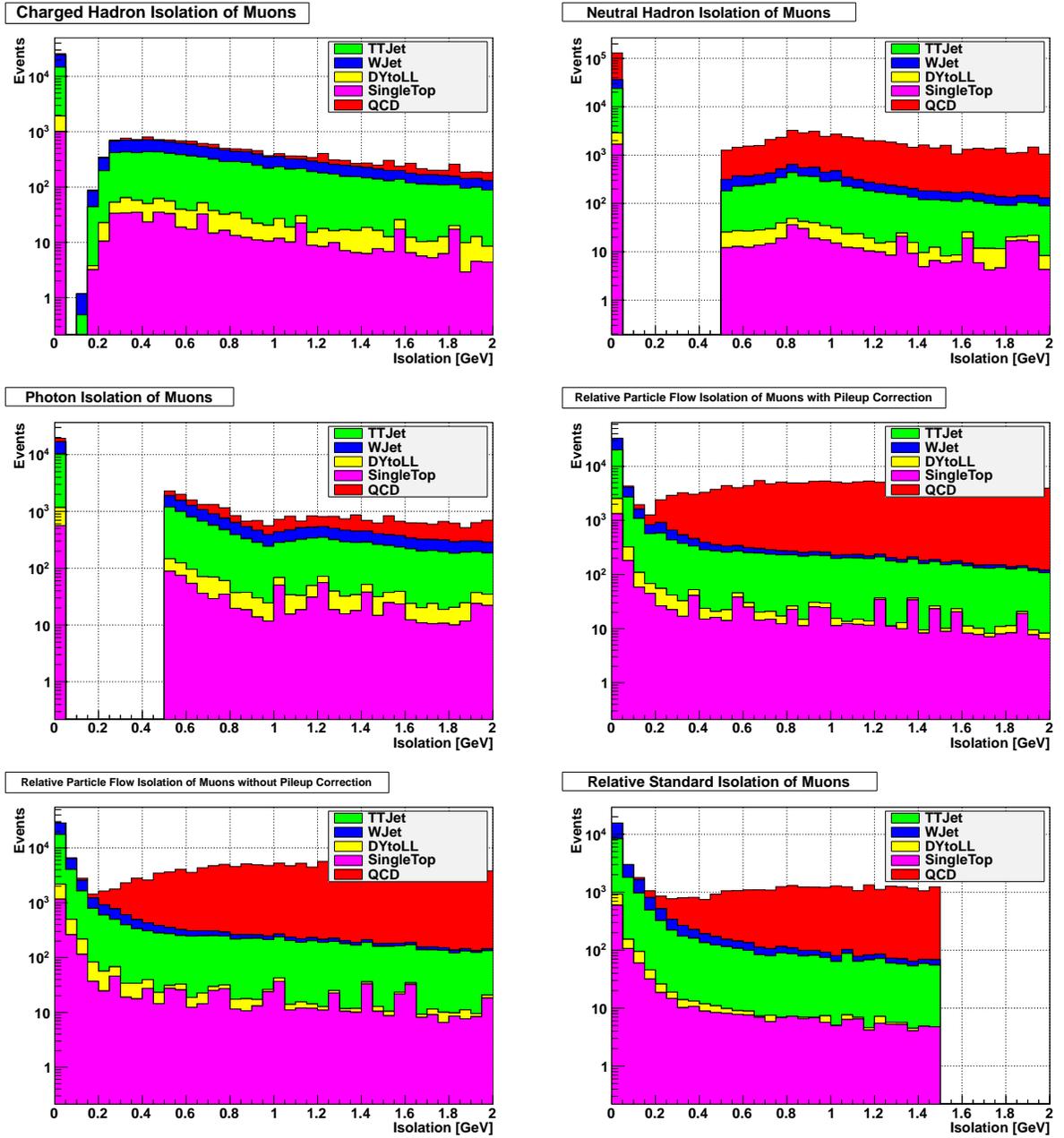


Figure 20: Isolation histograms with  $\cancel{E}_T > 60$  and  $H_T > 300$

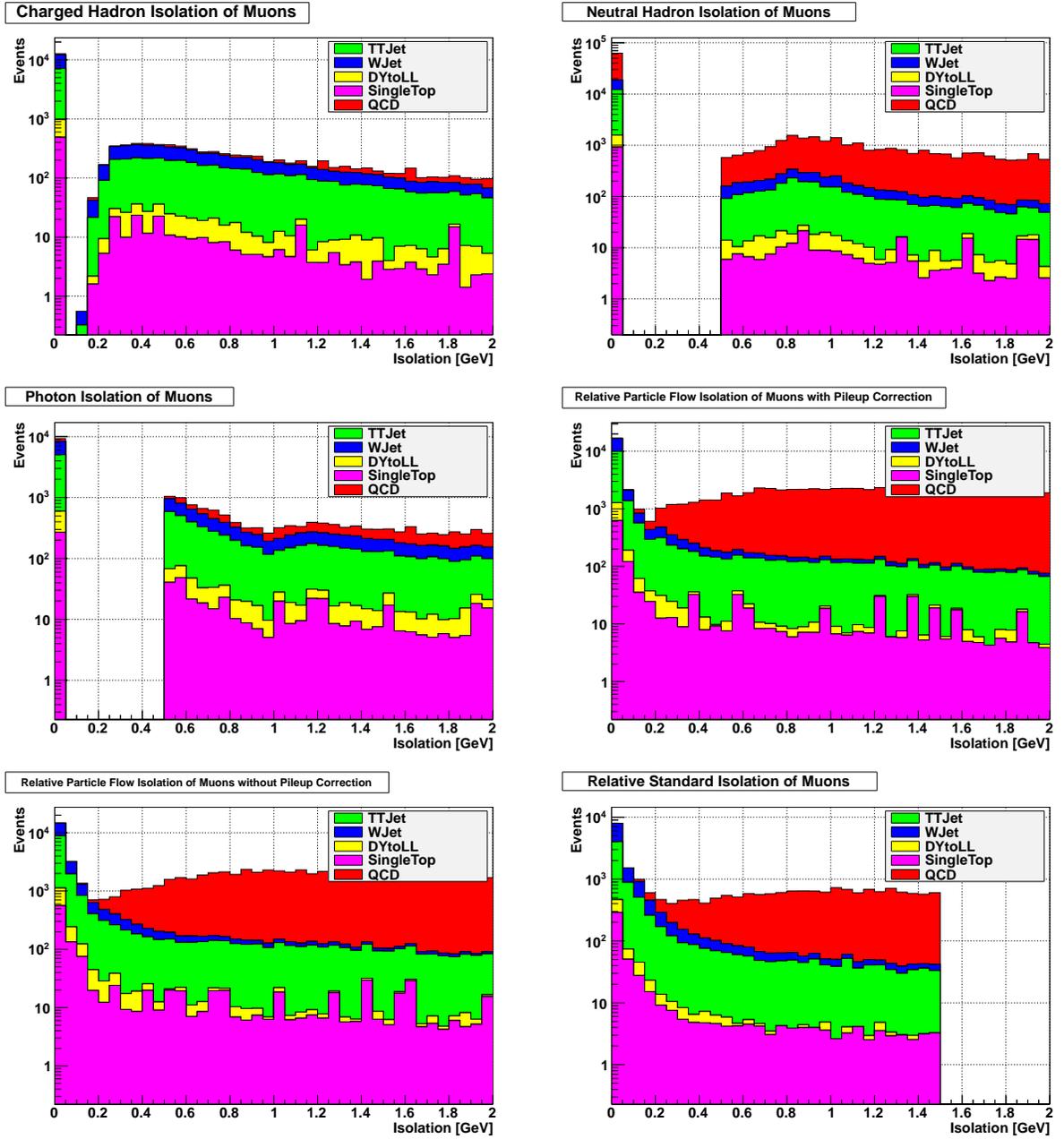


Figure 21: Isolation histograms with  $\cancel{E}_T > 60$  and  $H_T > 400$

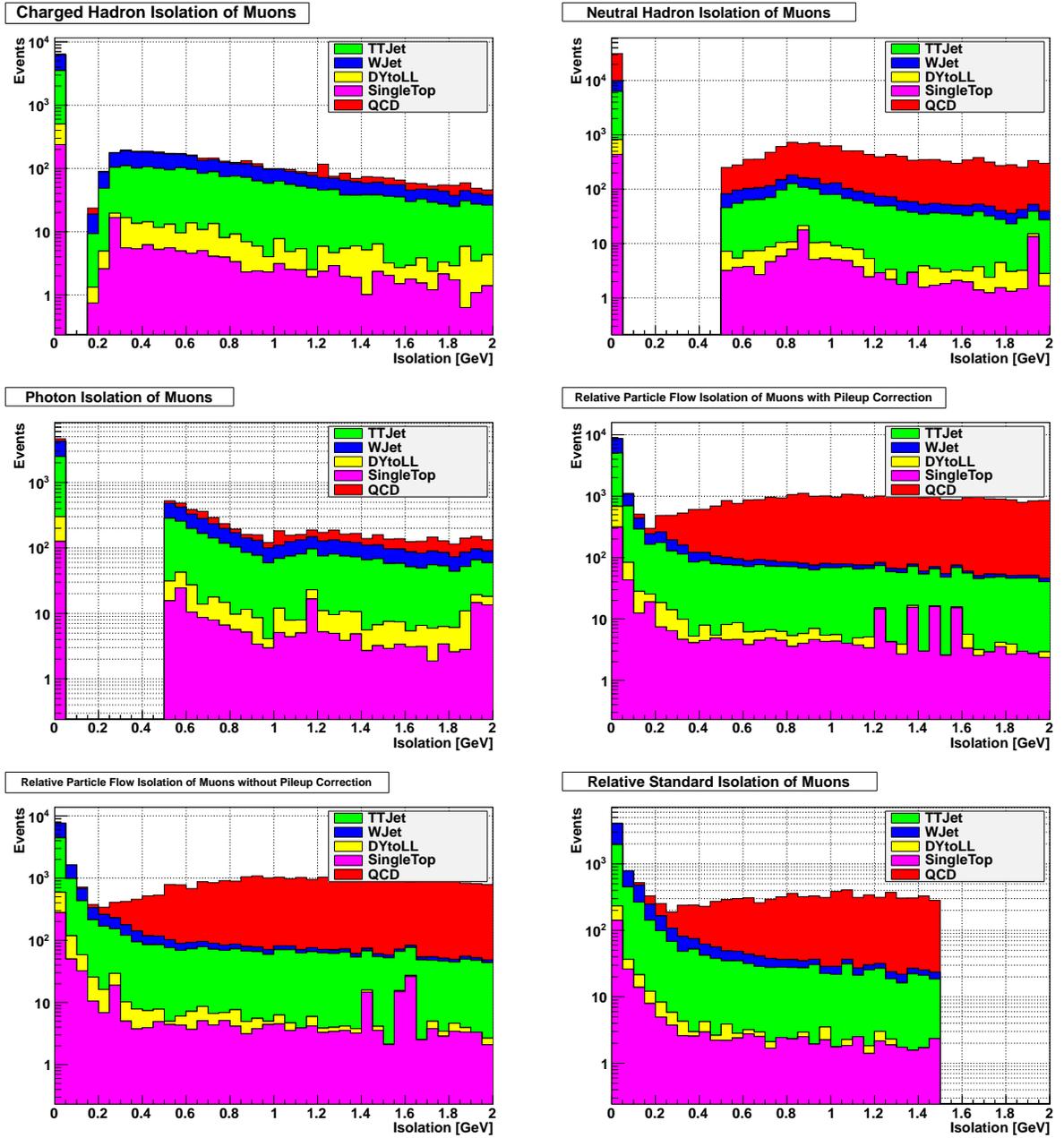


Figure 22: Isolation histograms with  $\cancel{E}_T > 60$  and  $H_T > 500$

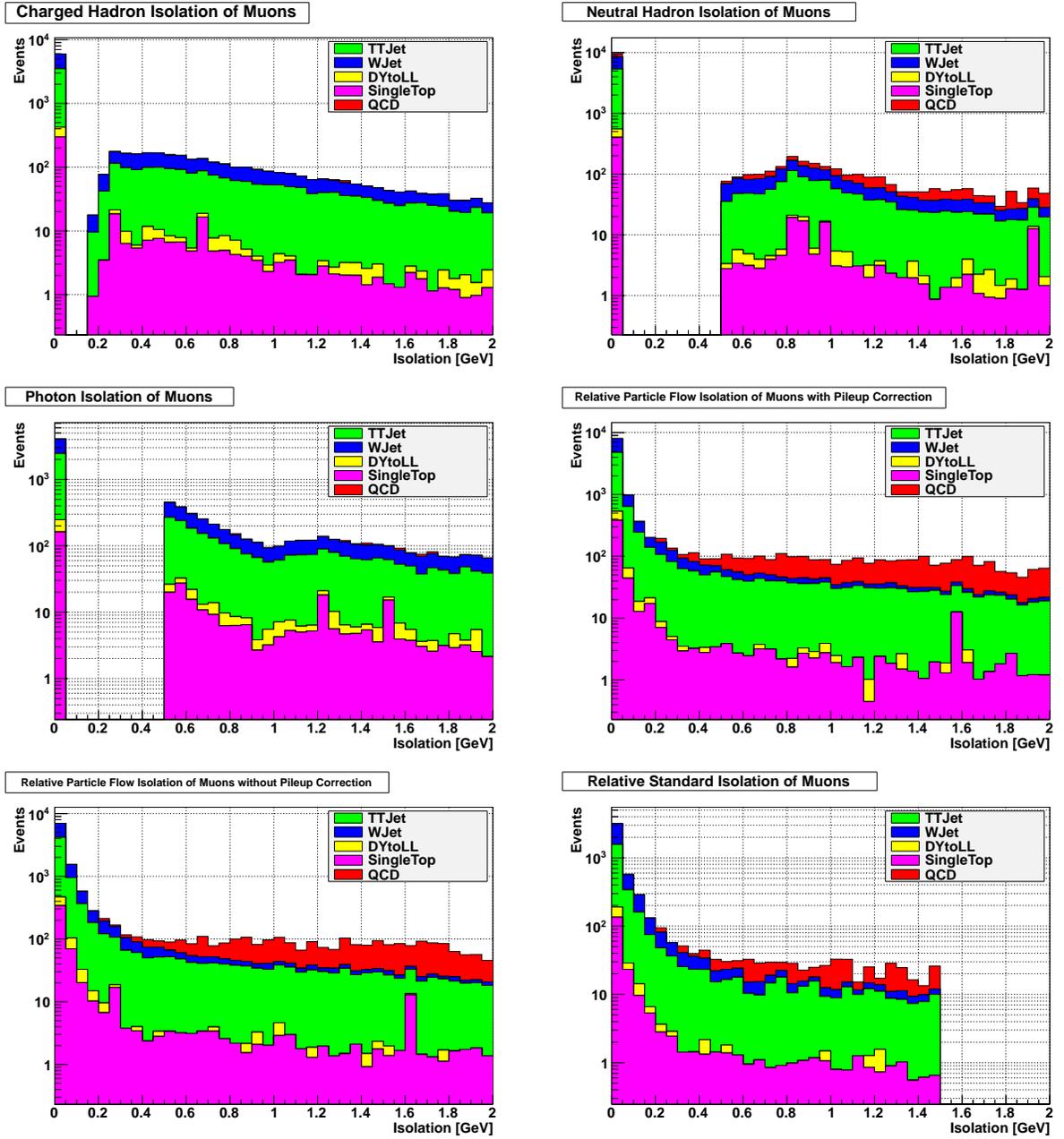


Figure 23: Isolation histograms with  $\cancel{E}_T > 150$  and  $H_T > 200$

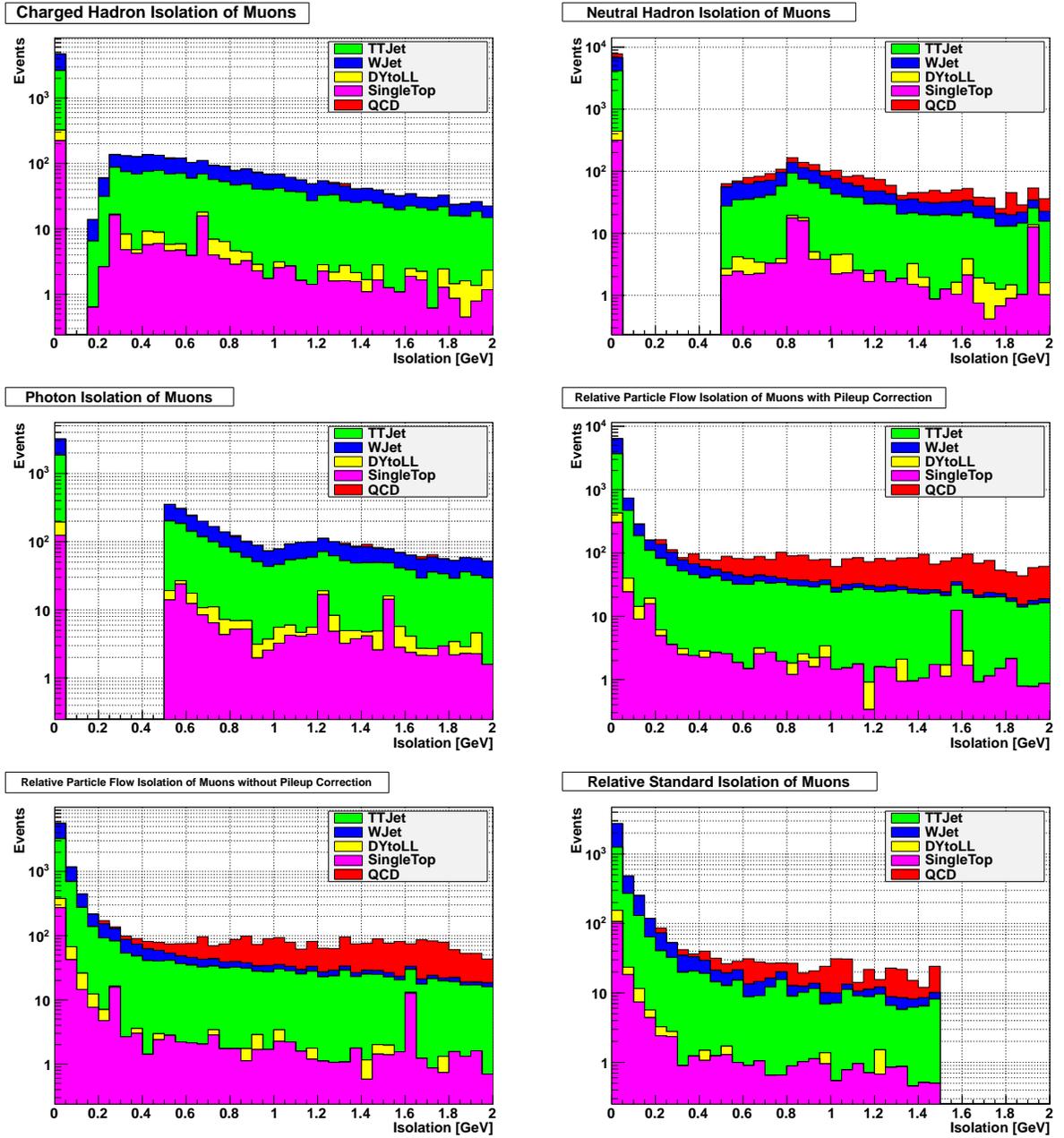


Figure 24: Isolation histograms with  $\cancel{E}_T > 150$  and  $H_T > 300$

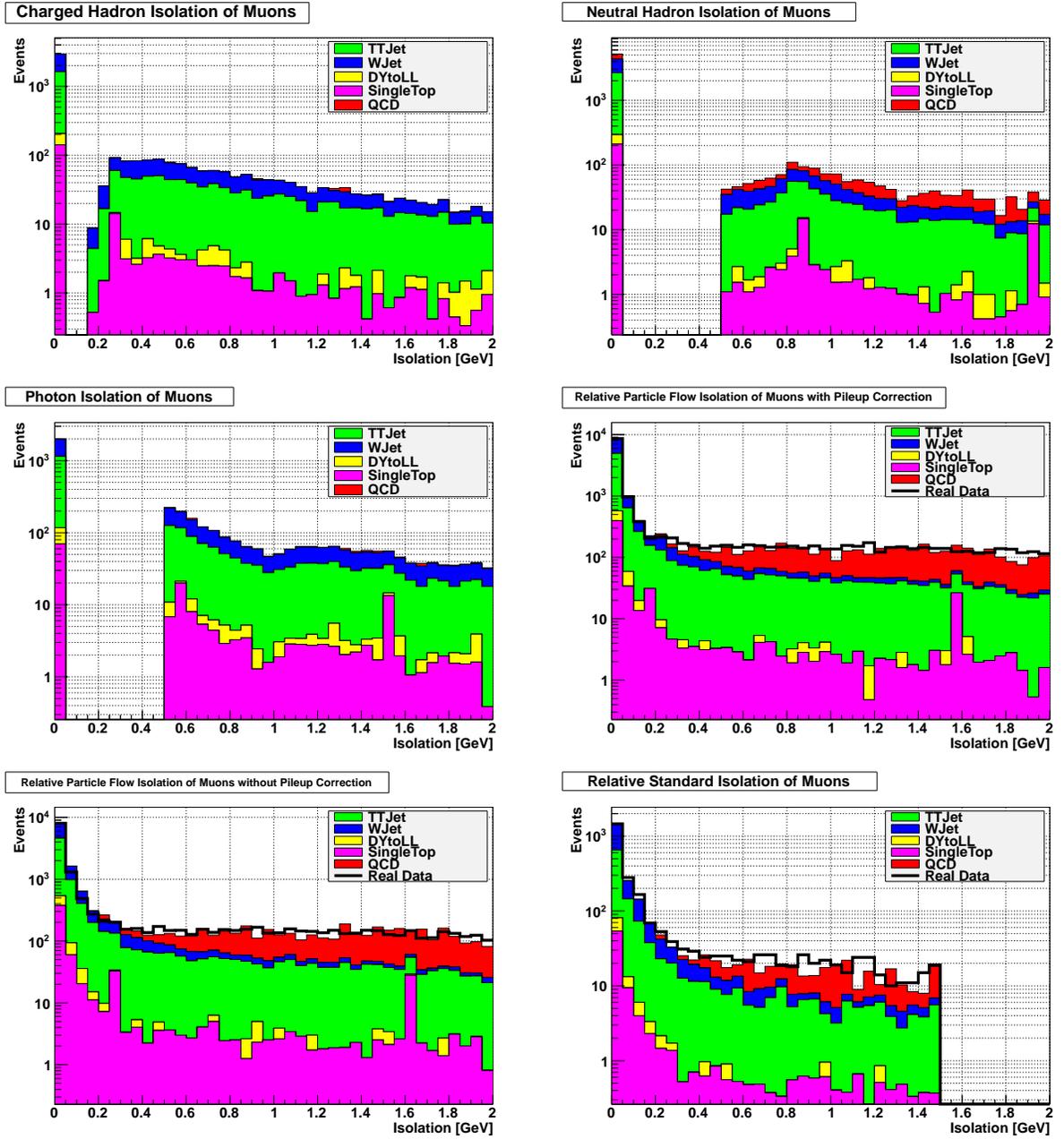


Figure 25: Isolation histograms with  $\cancel{E}_T > 150$  and  $H_T > 400$

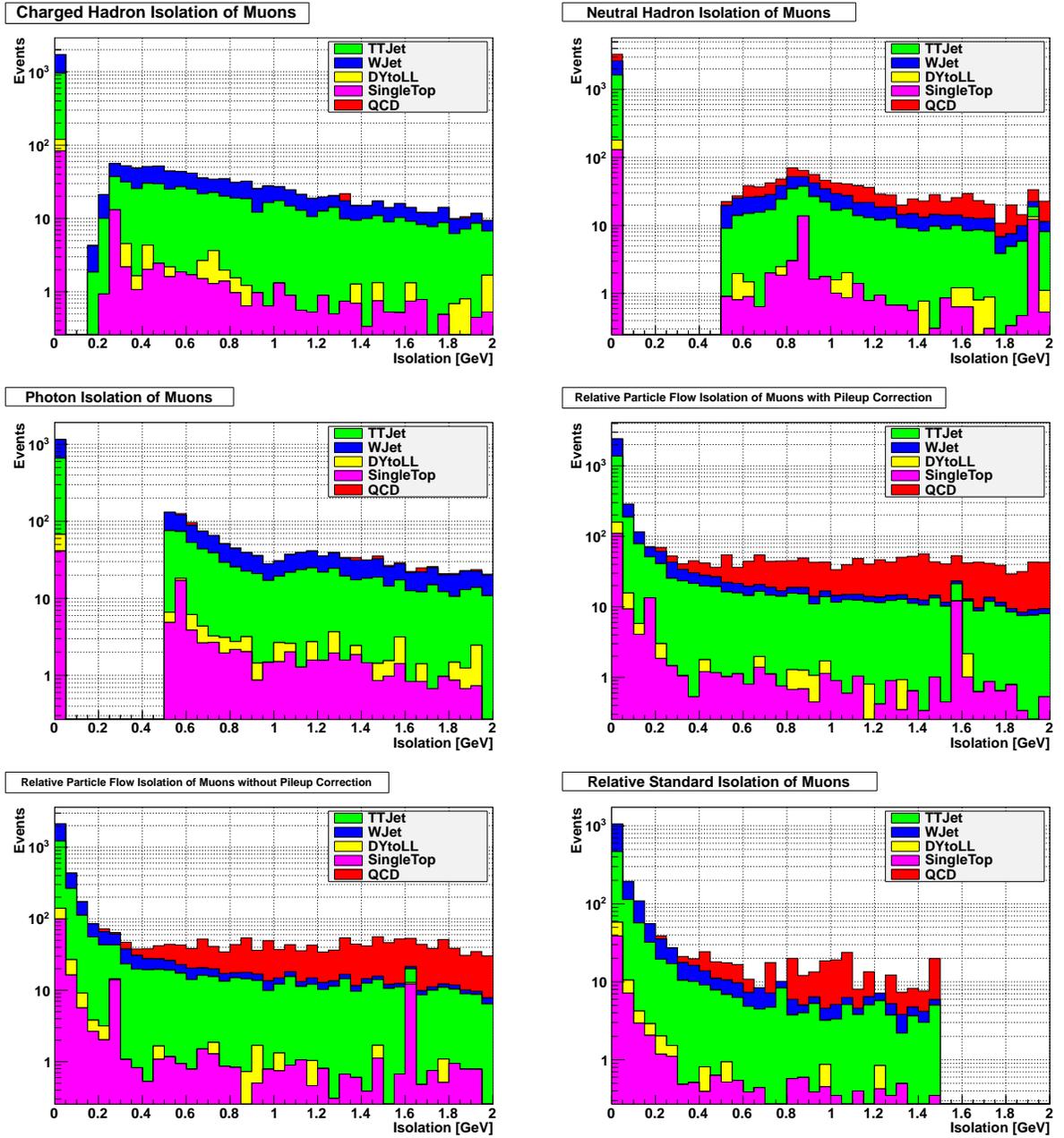


Figure 26: Isolation histograms with  $\cancel{E}_T > 150$  and  $H_T > 500$

## B Plots

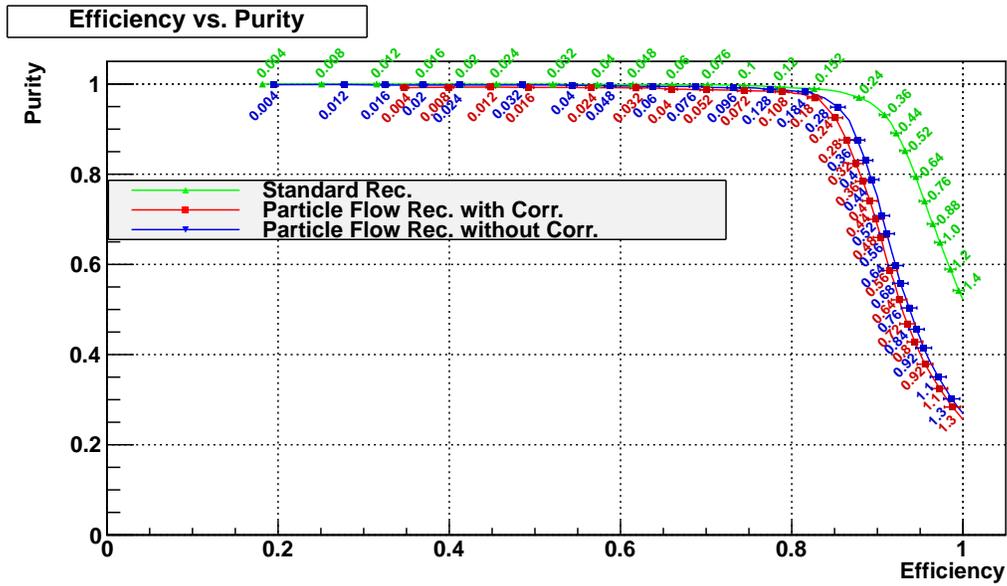


Figure 27: Efficiency plotted versus purity with  $E_T > 60$  and  $H_T > 200$

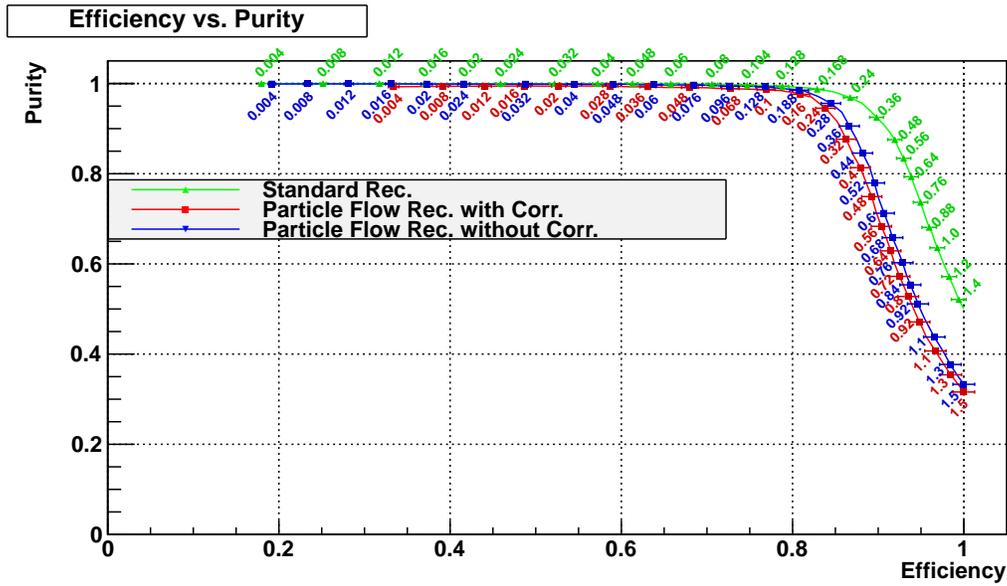


Figure 28: Efficiency plotted versus purity with  $E_T > 60$  and  $H_T > 300$

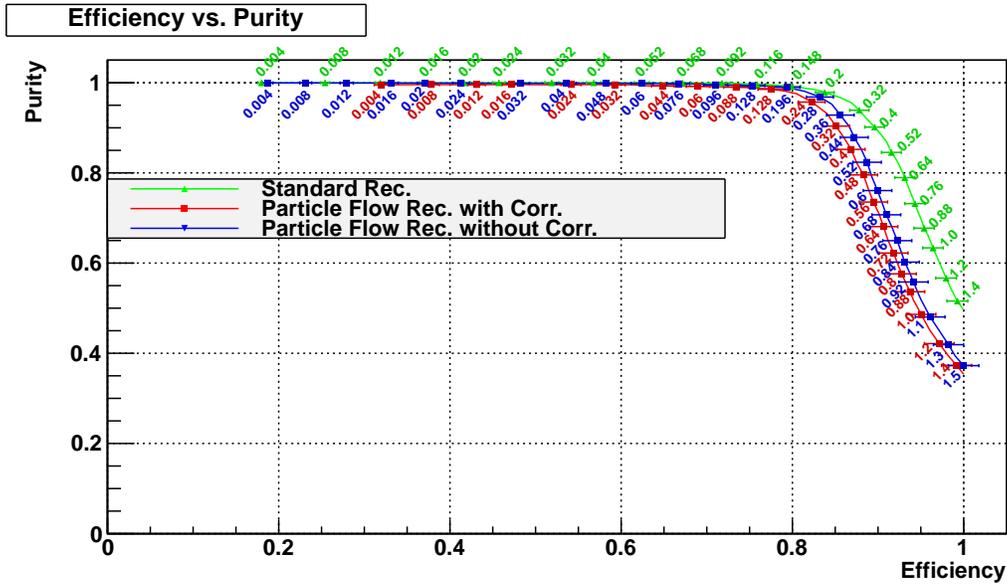


Figure 29: Efficiency plotted versus purity with  $E_T > 60$  and  $H_T > 400$

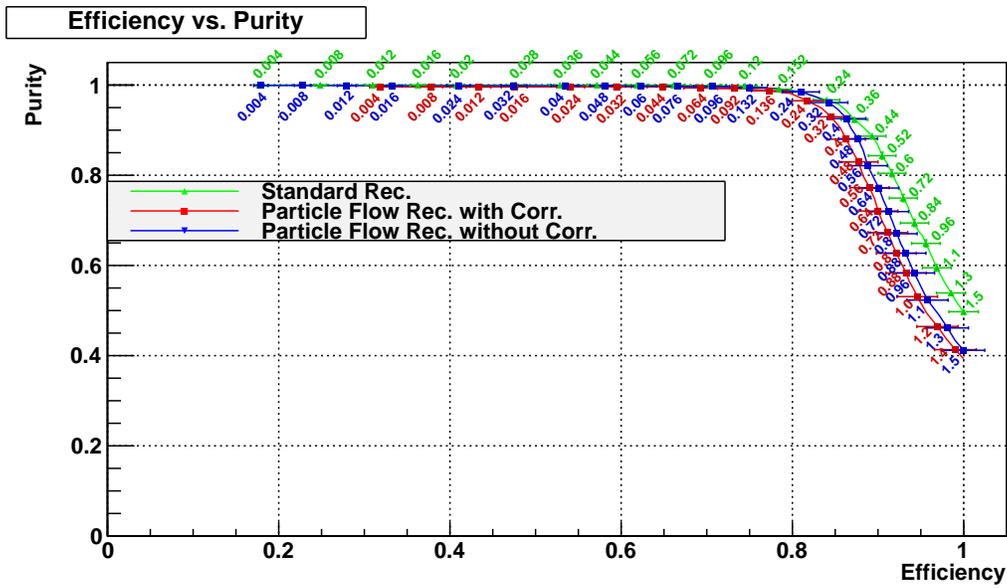


Figure 30: Efficiency plotted versus purity with  $E_T > 60$  and  $H_T > 500$

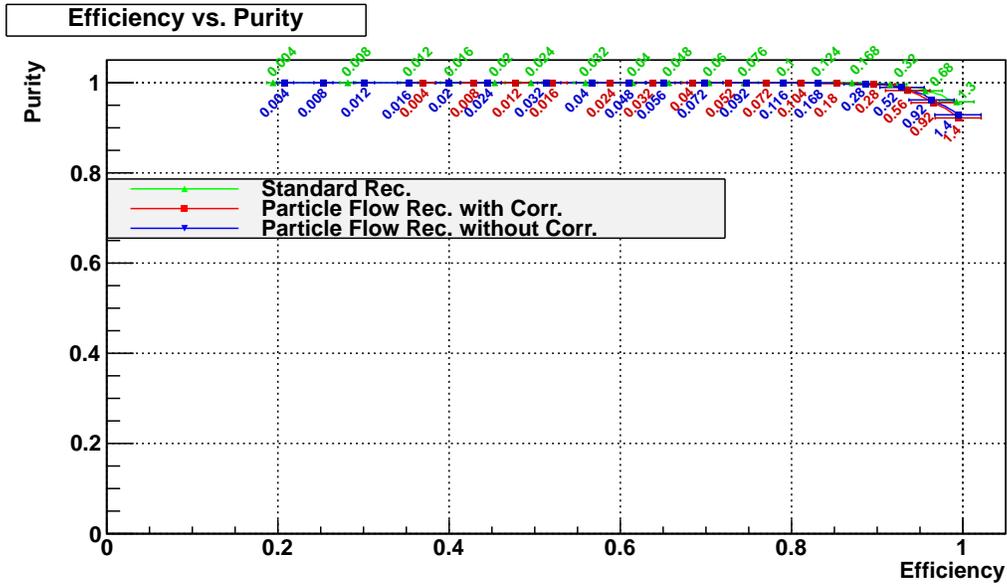


Figure 31: Efficiency plotted versus purity with  $\cancel{E}_T > 150$  and  $H_T > 200$

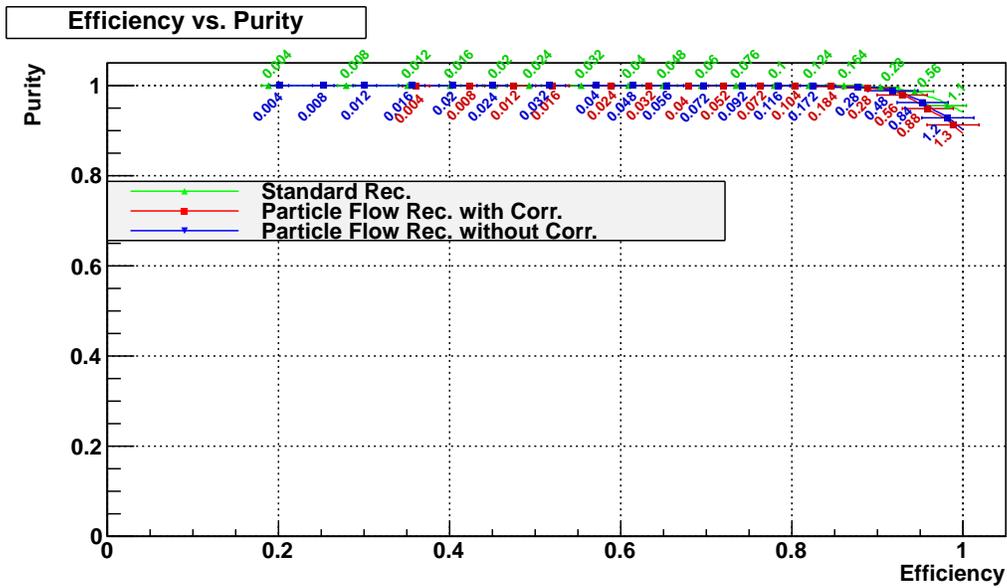


Figure 32: Efficiency plotted versus purity with  $\cancel{E}_T > 150$  and  $H_T > 300$

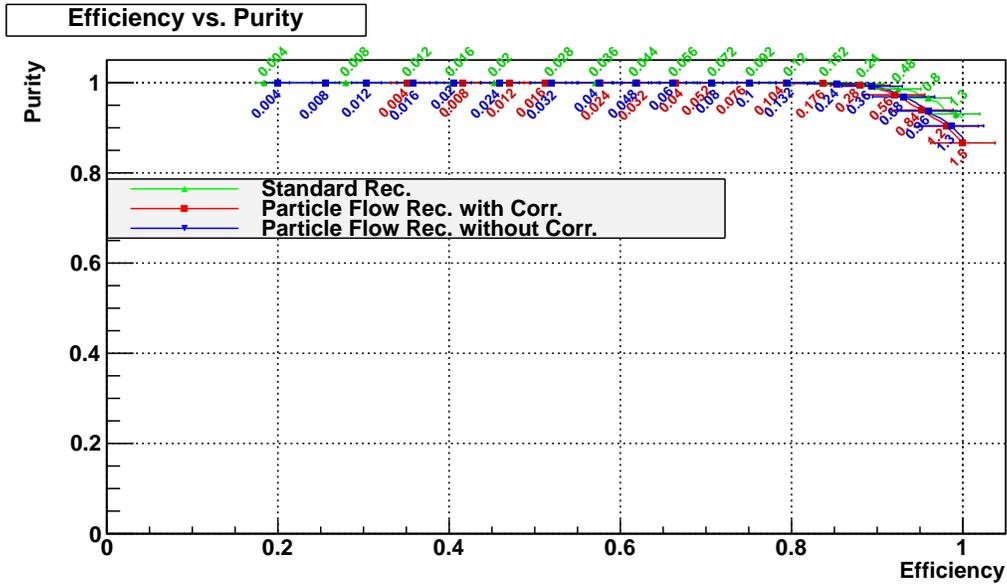


Figure 33: Efficiency plotted versus purity with  $\cancel{E}_T > 150$  and  $H_T > 400$

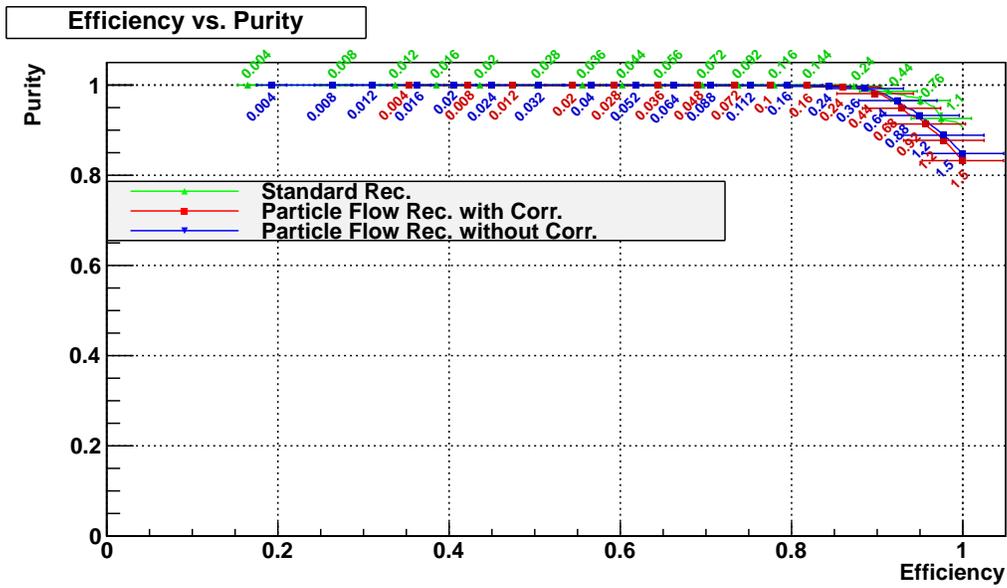


Figure 34: Efficiency plotted versus purity with  $\cancel{E}_T > 150$  and  $H_T > 500$