
Estimation of Muon Detection Efficiency in CMS Detector using Tag and Probe method

A report submitted in fulfilment of the
requirements for the Summer Program
of
Master of Science

by

Sayan Dhani
(M.Sc. Physics, IIT Bombay
Summer Intern, SINP)

under
Prof. Dr. Surbir Sarkar (SINP)

to the
Centre for Advanced Research and Education (CARE)



Saha Institute of Nuclear Physics

July 2023

Abstract

This project delves into the world of high-energy particle physics, focusing on the CMS detector at the Large Hadron Collider (LHC). The central theme is the Tag and Probe method employed to estimate muon detection efficiency. This efficiency is vital for various physics analyses, from measuring cross-sections to the quest for rare phenomena and investigation of particle decay. Using real data, as opposed to simulations, allows for more precise assessments, calibration, and correction, ultimately enabling the extraction of signal events from background contributions with higher accuracy. The study primarily employs two methods: ratios of histograms after consecutive selection cuts and fitting of mass histograms in different PT ranges. This endeavour ultimately enables more precise extraction of signal events from background contributions, promising finer measurements of physical properties and the potential discovery of new particles.

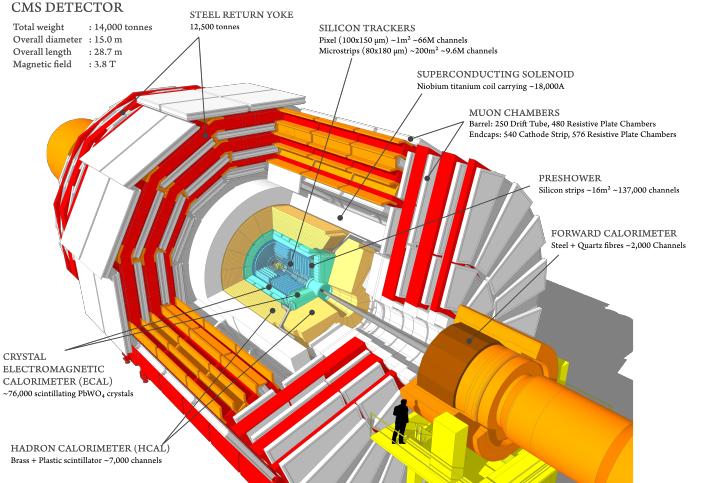
Contents

1	Introduction	1
2	Overview of the CMS(Compact Muon Solenoid) Experiment	1
2.1	Detectors of CMS	1
3	Muon Detection in CMS Experiment	2
3.1	Signal and Background	2
3.2	Motivation for Studying Muon Detection Efficiency	2
3.3	Theoretical Framework	3
3.3.1	Muons as fundamental particles	3
3.3.2	Principles of muon detection in CMS detector	3
3.3.3	Correlation between Muon Detection Efficiency and Momentum	3
3.3.4	Key Variable and Terminology used in the analysis :	3
4	Tag and Probe Method and Muon Detection Efficiency:	4
4.1	Tag and Probe Method:	4
4.1.1	What is “Tag” and “Probe”?	4
4.2	Calculating the efficiency:	4
4.3	Tag and Probe: Object and Event Selection	4
4.4	Estimation of Muon Detection Efficiency:	5
4.4.1	Method 1: Taking Ratios of the Histogram after Consecutive Selection cuts:	5
4.4.2	Method 2: Fitting the mass histogram in different P_T ranges	5
5	Tools and Techniques	6
5.1	Introduction to ROOT-PyROOT Framework	6
5.1.1	Coding Languages for Analysis:	6
5.1.2	Utilisation of PyROOT in Python for Data Analysis	6
5.2	RooFit-Signal and Background Modeling:	6
6	Result and Discussion	6
7	Conclusion	11
7.1	Summary of Findings and Contributions	11
7.2	Observations	11

1 Introduction

High-energy particle physics plays a significant role in advancing our understanding of the fundamental nature of the universe. At the heart of this quest lies the Large Hadron Collider (LHC), the world's most powerful particle accelerator, located at CERN in Switzerland. In this project, we explore the intricacies of the CMS detector of LHC and its significant contributions to cutting-edge research. Our primary focus lies on the experimental procedure named the Tag and Probe method, which is an essential data-driven method used here for estimating muon reconstruction efficiency[4].

The Tag and Probe method is applied to an unbiased data set, where it uses a very well-defined reconstructed object as a "Tag" (a reference) to find a very minimal biased sample of the "probe" object. These probe objects are used to measure the efficiency of the detection. Here, we discuss the Muon, a fundamental particle in the standard model[2].



2.1 Detectors of CMS

Some of the main components of the CMS detectors are :

Tracker : The tracker is a fundamental component of the CMS detector, located closest to the collision point. It consists of silicon-based detector layers that can accurately measure the trajectories of charged particles, providing essential information such as particle momentum, charge, and primary vertex position. This data is crucial for precise particle reconstruction and physics analyses. But only the charged particles leave tracks in the tracker.

Electromagnetic Calorimeter(ECAL) : Surrounding the tracker, the Electromagnetic Calorimeter (ECAL) is designed to measure the energy of electrons and photons. It comprises lead tungsten crystals that produce light when struck by electrons or photons. Precisely measuring their energy is essential for identifying and reconstructing these particles, contributing to a wide range of physics studies.

Hadron Calorimeter(HCAL) : The Hadron Calorimeter (HCAL) surrounds the ECAL and is responsible for measuring the energy of hadrons, particles composed of quarks. It consists of alternating layers of dense absorber and scintillator tiles. HCAL plays a critical role in detecting and measuring the energy of hadrons, providing valuable information for various physics analyses.

Superconducting Solenoid Magnet : The CMS detector incorporates a superconducting solenoid magnet, generating a powerful magnetic field. This magnetic field enables the precise measurement of particle momenta, crucial for particle identification and accurate physics measurements. The solenoid magnet also aids in determining the charge of particles passing through the detector.

Muon System : The outermost layer of the CMS detector is the Muon System, dedicated to detecting and identifying muons—highly penetrating particles that can traverse through the inner layers of the detector. The Muon System plays a crucial role in physics analyses, as muons are vital for studying a wide range of phenomena and searching for new particles. It consists of several layers of gas detectors, drift tubes, and other components, ensuring precise measurement and identification of muons in the CMS experiment.

2 Overview of the CMS(Compact Muon Solenoid) Experiment

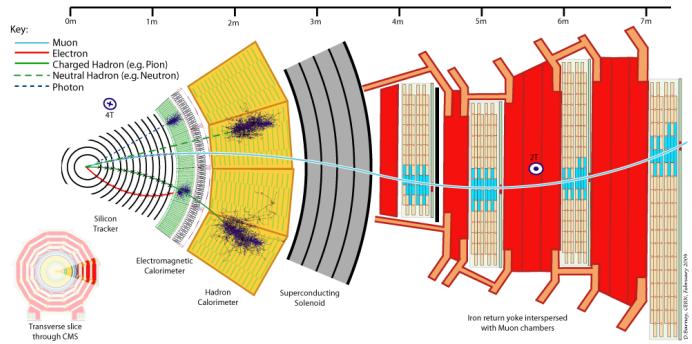
The Large Hadron Collider (LHC) is a powerful particle accelerator at CERN (European Organization for Nuclear Research) in Switzerland. The LHC consists of a 17-mile ring of superconducting magnets that accelerate protons and heavy ions to nearly the speed of light, creating high-energy collisions. Within the LHC, there are four major experiments, one of which is the Compact Muon Solenoid (CMS) experiment[6].

The CMS experiment is an important component of the LHC and is designed as a general-purpose detector. It played a pivotal role in discovering the Higgs boson in 2012. CMS is involved in searching for physics beyond the Standard Model, including exploring heavy ion collisions and rare particles like super-symmetric particles. With its exceptional resolution and various sub-detectors, CMS can differentiate even very close mass particles, enabling comprehensive measurements and research in diverse areas of particle physics.

The LHC, operating at its highest luminosity levels, provides a unique platform to study a wide range of physics processes. It has not only achieved the discovery of the Higgs boson but also enabled precise measurements of its properties. Additionally, the LHC experiments, including CMS, continuously explore the properties of quark-gluon plasma and top quarks, advancing our understanding of fundamental particles and their interactions.

Trigger System of CMS and Their Application :

The trigger system in the CMS experiment is a vital component responsible for efficiently selecting interesting collision events from the overwhelming number of proton-proton collisions that occur every second. It acts as a real-time filter, identifying events that may contain rare or significant physics processes worth further analysis. The trigger system is designed to work in two stages: 1) the Level-1 Trigger, which uses hardware-based algorithms for fast event selection, and 2) the High-Level Trigger (HLT), which employs sophisticated software-based algorithms to refine event selection further. Together, they ensure that the most relevant collision events are recorded and processed for detailed physics analysis, optimising data storage and computational resources.



o Level-1 Trigger

- a. 1st stage trigger system, based on hardware like ASICs and FPGAs, used for fast trigger mechanisms.
- b. It operates very fast(within 10 ns) and makes real-time decisions to select only the interesting events and reject others.
- c. The trigger primitives include information about energy, momentum, and position of particles produced in the collision, which is used for the primary decision-making to select any event based on the information of these properties.
- d. After the L1 trigger's decision, it is passed to the HLT for further processing and analysis through the electronic channels

o High-Level Trigger

- a. 2nd stage trigger system consists of Cluster – HPC, FPGAs, Disk arrays, high-speed networks, and sophisticated event selector.
- b. It operates in a highly parallel manner, utilises multiple processing nodes to analyse the data from detectors and is also very fast(only a few ms).
- c. This trigger system reconstructs particles' tracks and identifies physics objects using advanced algorithms.
- d. Then, it filters and selects only the very interesting events, reducing and compressing data by storing only relevant information.

3 Muon Detection in CMS Experiment

The experimental data can be classified into two categories:

3.1 Signal and Background

Signal : The specific physical processes or particles that researchers are interested in studying or discovering.

- It is characterised by specific experimental signatures or properties that distinguish them from the background events
- Our goal is to identify and isolate the signal events from the background events to study and measure their properties as accurately as possible and compare them with theoretical predictions or models.

Background : It refers to events or processes (like the QCD process) that mimic or resemble the signal events but are produced from non-interesting physical processes.

- Arise from well-understood and known physical processes in particle collisions at the LHC.
- Our challenge is to minimise the background events as much as possible and enhance the signal-to-background ratio. It can be achieved through careful data analysis techniques, selection criteria, and various statistical background estimation methods.

3.2 Motivation for Studying Muon Detection Efficiency

In high-energy physics experiments, the study of muon detection efficiency is important because, through this efficiency, calculations derived from data are essential for physics analyses, such as cross-section measurements, searches for rare processes, studies of particle decay and many more.

Now, we can use real or simulated data, but using real data instead of Monte Carlo (MC) simulations is crucial for several reasons.

First, data accounts for various detector-specific effects, such as inefficiencies, misalignment, or calibration uncertainties, which MC simulations might not fully capture. This allows for a more accurate evaluation of the detector's performance and identifying areas for improvement.

Second, comparing data-driven efficiencies with those obtained from MC simulations enables the application of corrections and calibrations to the simulated data, aligning it with the observed data and improving the accuracy of simulation predictions.

Therefore, quantifying the efficiency of selecting specific particles or processes from real data allows for more accurate signal extraction, distinguishing signal events from background contributions more precisely and leading to more precise measurements of physical properties or the potential discovery of new particles.

3.3 Theoretical Framework

3.3.1 Muons as fundamental particles

Muons are elementary particles (belonging to the Lepton family), like electrons but approximately 200 times heavier. They carry an electric charge of -1 and have a spin of -1/2. With a mean lifetime of 2.2 microseconds, muons are relatively short-lived particles. Despite their brief existence, muons play a crucial role in particle physics as their properties and interactions provide valuable insights into the fundamental forces and subatomic interactions that govern the universe. Their unique characteristics make them highly valuable for studying high-energy phenomena and exploring the nature of matter and the universe.

3.3.2 Principles of muon detection in CMS detector

1. The muons are highly penetrating particles that can traverse through the inner layers of the detector without much loss of energy due to their low interaction with matter.
2. All the detectors like Tracker, ECAL, HCAL, and Muon systems detect and reconstruct muon trajectories.
3. The Muon system, consisting of drift tubes, cathode strip chambers, and resistive plate chambers, is designed to identify and measure muons' properties.
4. After combining information from different detector components, it can accurately determine the momentum, charge, and other properties of detected probable muon candidates.
5. The detection efficiency determines how accurately the detector can identify the probable muon candidates.

3.3.3 Correlation between Muon Detection Efficiency and Momentum

- A correlation can be observed as follows: higher momentum muons tend to have higher detection efficiency.
- This correlation arises due to the characteristics of detector systems and some energy loss through photons for the low-energy muons.

3.3.4 Key Variable and Terminology used in the analysis :

There are many variables corresponding to an event that are associated with the objects produced in the event, like electrons, photons, muons, Jets, b-jets, MET, etc. These observables are obtained from the different reconstructions of the objects based on the information about the properties available from the detector. Some of these are -

1. **Transverse Momentum(p_T):** It is the momentum of the particle in the transverse plane(r, θ) of the detector, which is the perpendicular plane to the beam. Let us say \vec{P} is the momentum 3 vector, then

$$P_T = \sqrt{p_x^2 + p_y^2}$$

2. **Pseudo-Rapidity(η):** expressed as

$$\eta = -\ln(\tan \frac{\theta}{2})$$

3. **Angle θ :** The angle between the \vec{r} and the Z-axis in the usual spherical coordinate system.
4. **Angle ϕ :** Angle of the object's trajectory in the plane transverse of the directions of the proton beams.
5. **Mass (m) :** This is the rest mass of the muons. We also use the invariant mass, which is the Lorentz invariant mass of the Muon Pair.
6. **Charge :** There is also a quantity mentioning the charge of the object.
7. **Missing Transverse Energy (MET):** The initial momentum of the colliding partons (quarks + gluons) along the beam axis is not known – the energy of each hadron splits and is constantly exchanged between its constituents – so the amount of total missing energy cannot be determined. However, the initial energy in particles travelling transverse to the beam axis is zero, so any net momentum in the transverse direction indicates missing transverse energy, also called **missing E_T** or **MET**. Missing transverse momentum is the negative vector sum of the transverse momenta of all detected particles in an event. The magnitude of the missing transverse momentum vector is called missing transverse energy.

8. ΔZ : The distance between the point where the track of a particle meets the beam axis after extrapolating and the primary vertex is called the ΔZ . It gives information about whether the particle is coming from (close to) the primary vertex.

9. ΔR : It is the measure of the angular separation between two particles in a detector.

$$dR = \sqrt{(\delta\eta^2 + \delta\phi^2)}$$

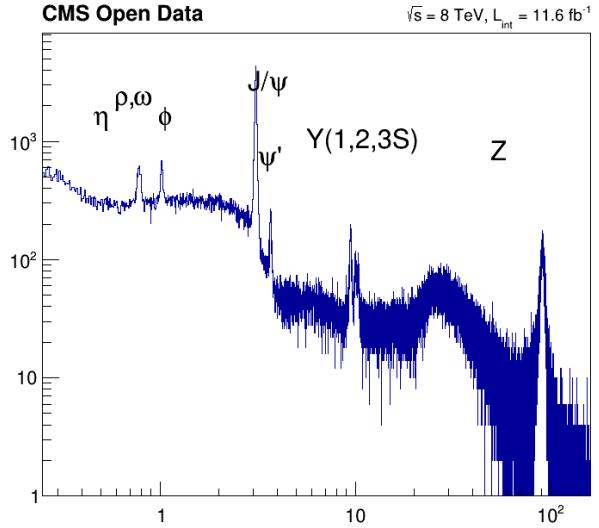
Where $\delta\eta$ and $\delta\phi$ represent the difference in pseudo-rapidity and the azimuthal angle values between the two particles, respectively. When R is small (close to zero), the two particles are very close in the detector, while the large value of R indicates that they are well-separated in the detector.

10. **Isolation:** (for muon) The isolation of a muon is often defined by selecting a cone around the muon's direction and calculating the sum of the transverse momenta (P_T) of all other particles (excluding the muon itself) within that cone.

$$\text{Isolation} = \frac{\sum_i P_T^i (\text{other particles within cone around muon})}{P_T^\mu}$$

Commonly used values of isolation are in the range of 0.2 to 0.4, representing a cone with a radius of 0.2 to 0.4 unit in the R space around the muon's direction.

4.1.1 What is “Tag” and “Probe”?



The peaks in the above figure correspond to the resonance particles. The resonance particle, which is used to calculate the efficiencies, decays into a pair of particles: the tag and the probe[2]. **Tag muon** = well-identified, triggered muon (tight selection criteria). **Probe muon** = unbiased set of muon candidates (very loose selection criteria), either passing or failing the criteria for which the efficiency is to be measured.

4.2 Calculating the efficiency:

The efficiency is given by the fraction of probe muons that pass given criteria like reconstruction criteria and isolation criteria. The formula for efficiency is given by :

$$\epsilon = \frac{\text{No. of probe muons Passing the criteria}}{\text{No. of all probe muons}}$$

The denominator corresponds to the number of resonance candidates (**tag+probe pairs**) reconstructed from the data set. The numerator corresponds to the subset for which the probe passes the criteria.

The (tag+probe) invariant mass distribution is used to select only the signal that can be identified through the resonance peak in the invariant mass distribution, that is, only the true Z candidates decaying into di-muons.

4.3 Tag and Probe: Object and Event Selection

1. **Event Selection:** Select the interesting events having
 - (a) There must be at least two muon candidates.
 - (b) HLT Condition: There must be at least one isolated muon with p_T of at least 24 GeV and $|\eta| < 2.1$.

- (c) The MET must have P_T less than 40 GeV.
2. **Tag Muon Selection:** Select those muons from each event that have
- (a) $P_T > 24$ GeV, isolation < 0.3 , $|\eta| < 2.1$
 - (b) Have the tight/best reconstruction.
 - (c) Not to keep any bias, the muons should be shuffled, and the 1st one from that is chosen,
3. **Probe Selection:** from the same event,

- (a) Loop over all the muons except the tag muon.
With conditions to satisfy :
- (b) Opposite charge to the tag muon and $|\eta| < 2.1$,
- (c) Gives mass of the parent particle (which can be Z) (adding tag and probe) between 20 – 160 GeV

Apply Vetoos : Even if the tag and probe are selected and fall under the mass-range criterion, there could be electrons, jets, and b-jets present in the event. If they present, then it is not an event that we are interested in, so we must apply some vetoes to remove them. These vetoes are as follows:

1. **Electrons Selections:** for selecting good electrons, the electrons must have
 - (a) $P_T > 10$ GeV, $|\eta| < 2.5$, Multivariate-analysis id = 0,
 - (b) $\Delta R > 0.4$ (space angle between electron and the probe or tag muon)
2. **Jets Selection:** for selecting good Jets, the Jets must have
 - (a) $P_T > 30$ GeV, $|\eta| < 4.7$
 - (b) $\Delta R > 0.4$ (space angle between Jet and the probe or tag muon)
3. **b-Jets Selections:** for selecting good b-Jets, the b-Jets must have
 - (a) $P_T > 30$ GeV, $|\eta| < 2.5$
 - (b) $\Delta R > 0.4$ (space angle between Jet and the probe or tag muon)
 - (c) b-tag Score of the Jets > 0.33

Now, if the event has

1. Any selected electron, skip the event
2. More than 3 Jets skip the event
3. Any b-Jet, skip the event.

After all these selections, we can say that we have the events that could generate Z-particles, decayed into two correlated muons. Now, we can estimate how the CMS detector detects the probable muon candidates.

4.4 Estimation of Muon Detection Efficiency:

We can estimate the muon detection efficiency in many methods and ways; here, 2 of them are used. They are - 1) Method 1, by taking the ratio of the histograms after consecutive selection cuts[<empty citation>], 2) Method 2, by fitting the mass histogram in different P_T ranges[1].

4.4.1 Method 1: Taking Ratios of the Histogram after Consecutive Selection cuts:

Let, after event selection: No. of probe muons = N_0 (integrated over the histogram of P_T distribution)

Now apply tight ID cut: No. of probe muons = N_1 (,,)

Again, apply Isolation cut: No. of probe muons = N_2 (,,)

The efficiency of tight selection cut, $\epsilon_1 = \frac{N_1}{N_0}$ (taking the ratios of the P_T histogram after tight ID or predefined tight selection criteria applied on the probes to the P_T histogram without tight ID cut)

The efficiency of Isolation cut, $\epsilon_2 = \frac{N_2}{N_0}$ (taking the ratios of the P_T histogram after Isolation criteria applied on the probes to the P_T histogram without any isolation cut or tight ID cuts)

The efficiency of tight selection and Isolation Cut, $\epsilon_{2,1} = \frac{N_2}{N_1}$ (taking the ratios of the P_T histogram after Isolation and tight ID criteria applied on the probes to the P_T histogram without only tight ID cuts)

But as efficiency depends on the P_T of the muons, high P_T muons can be detected more efficiently. Still, the stat is reduced in the high P_T region, so the histograms need to be re-binned to keep the statistics high enough for every region.

Hence, we estimate the efficiency by plotting the muon P_T histograms(after different cuts) and dividing the histograms. These histograms are called efficiency histograms.

4.4.2 Method 2: Fitting the mass histogram in different P_T ranges

Although after the tag and probe method and event vetoes, we managed to get a purer sample of signal, we cannot remove the background completely. So now, one can estimate the signal and background using known functions. It will be good to model the background and signal and try to fit the data (using RooFit).

To do this, we will have to use the invariant mass histograms (for different pt ranges), which will look like the given figure.

The shape of the mass distribution is dominated by the backgrounds in the lower Pt ranges and dominated by signals in higher Pt ranges for this analysis. So, the background can be fitted using decaying power law, and the signal can be fitted using a double-sided Crystal ball function. After the fit, we get a number of signal events after each cut and calculate the efficiency, which should be more reliable.

5 Tools and Techniques

5.1 Introduction to ROOT-PyROOT Framework

ROOT is a powerful data analysis framework widely used in high-energy physics and other scientific disciplines. It offers a comprehensive set of tools for data storage, manipulation, analysis and data visualisation. One of its notable features is the PyROOT interface, which allows users to access ROOT's functionality from within Python. This Python integration makes it more accessible and user-friendly, enabling scientists and researchers to utilise ROOT's capabilities while leveraging Python's simplicity and extensive libraries for data analysis and visualisation. It's particularly valuable for those who want to combine ROOT's strengths with the versatility of Python in their data analysis workflows.

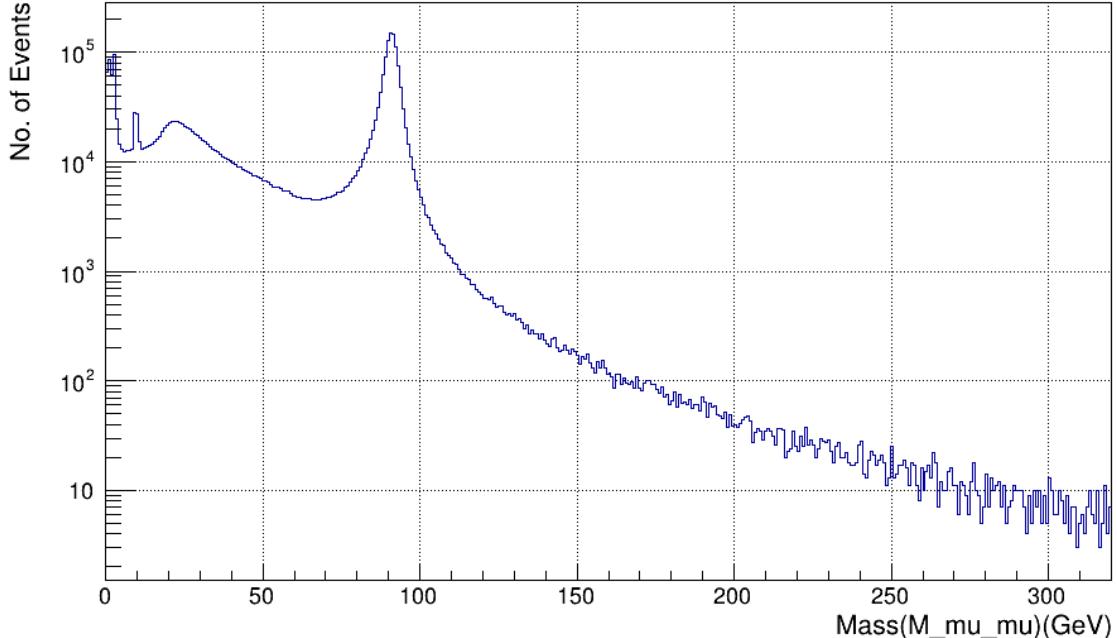
5.1.1 Coding Languages for Analysis:

In this analysis, mostly Python and some C++ languages are used.

6 Result and Discussion

Invariant Mass Distribution of Tags and Probes The peak implies that some tags and probes are produced through the decay of a resonant particle. It is likely to be the Z-particle, as the peak is around 91 GeV.

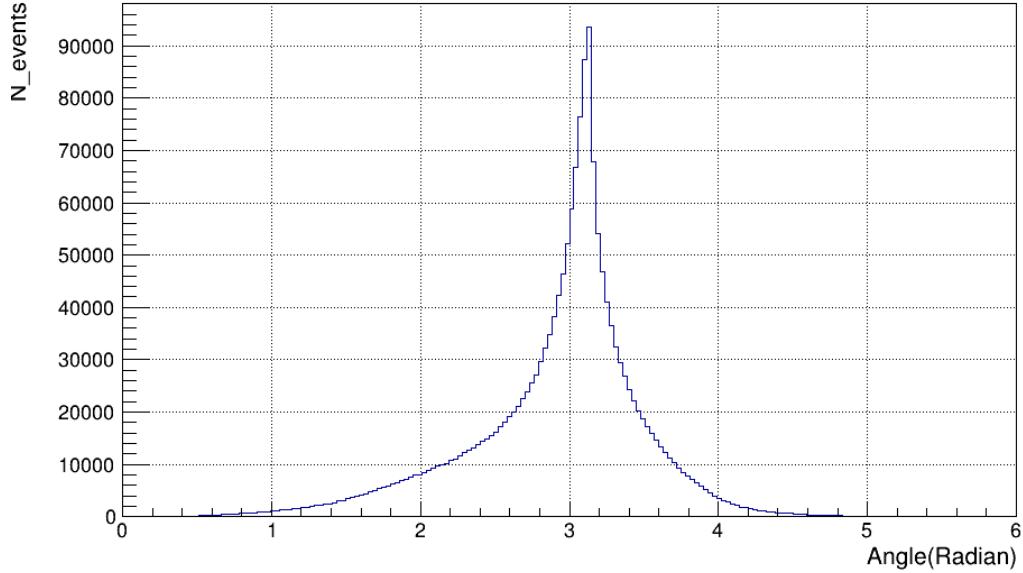
Invariant Mass of Tag & Probe Pair



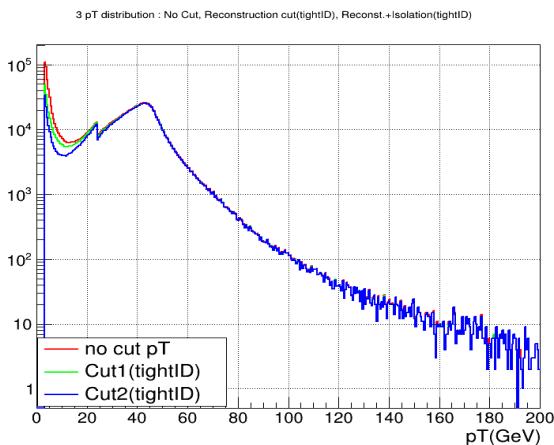
ΔR The distribution implies that most of the tags and probes are (180) apart, which implies most of them are

produced without any boost, indicating the parent particle was at rest while decaying.

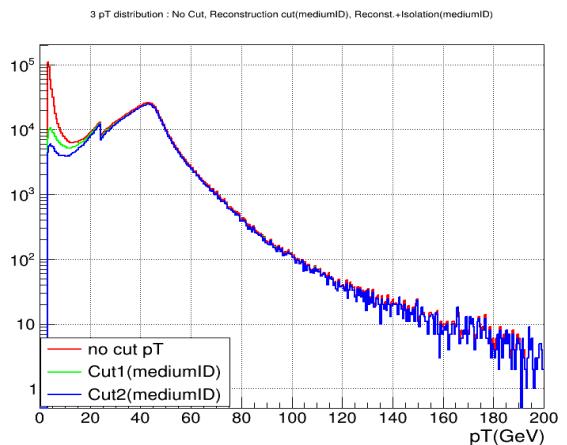
$\text{deltaR(tag, probe)}$



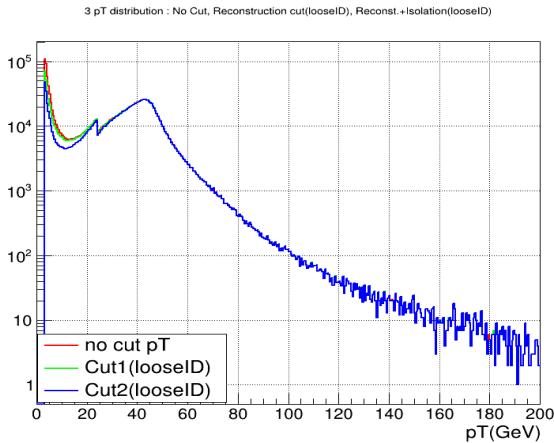
Transverse momentum distribution(P_T)



(a) with Tight ID

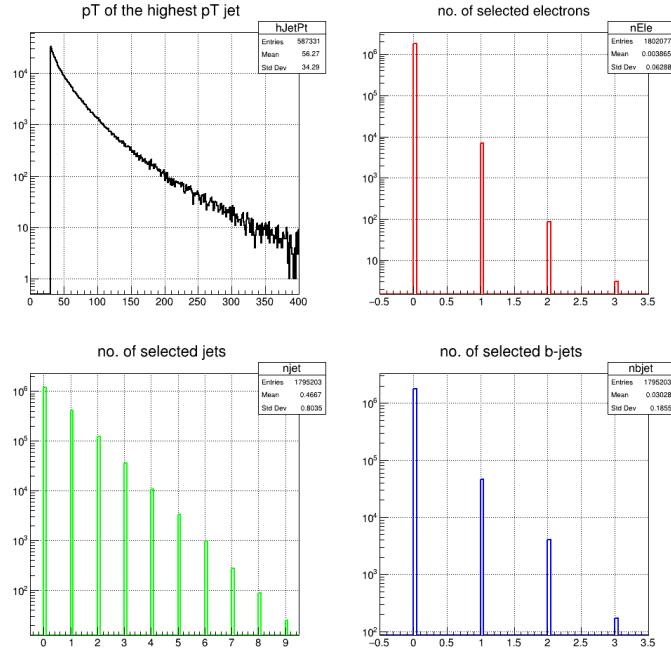


(b) with Medium ID

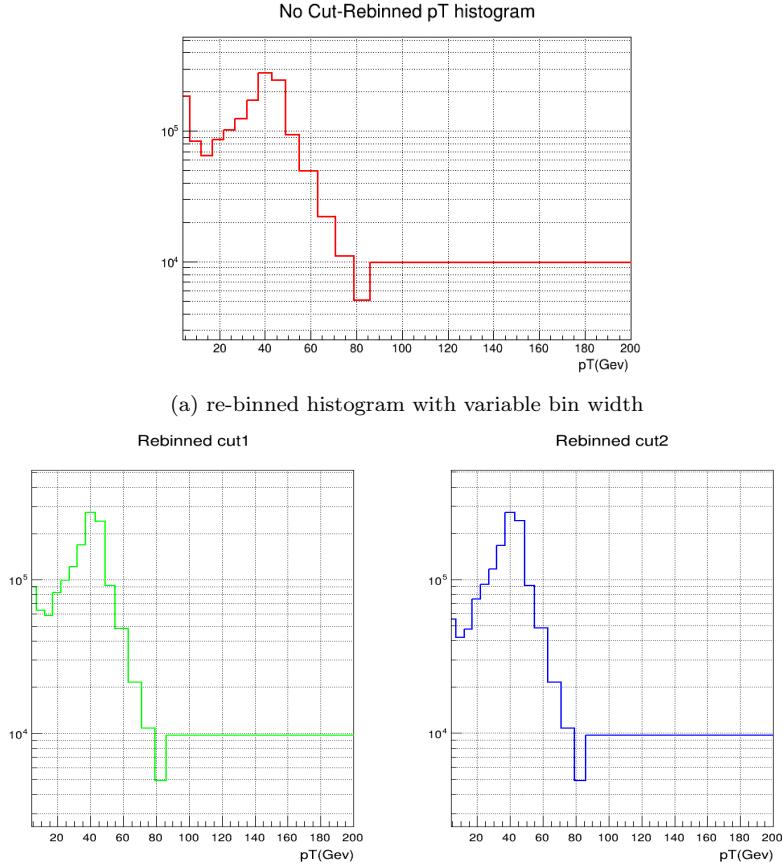


(c) with Loose ID

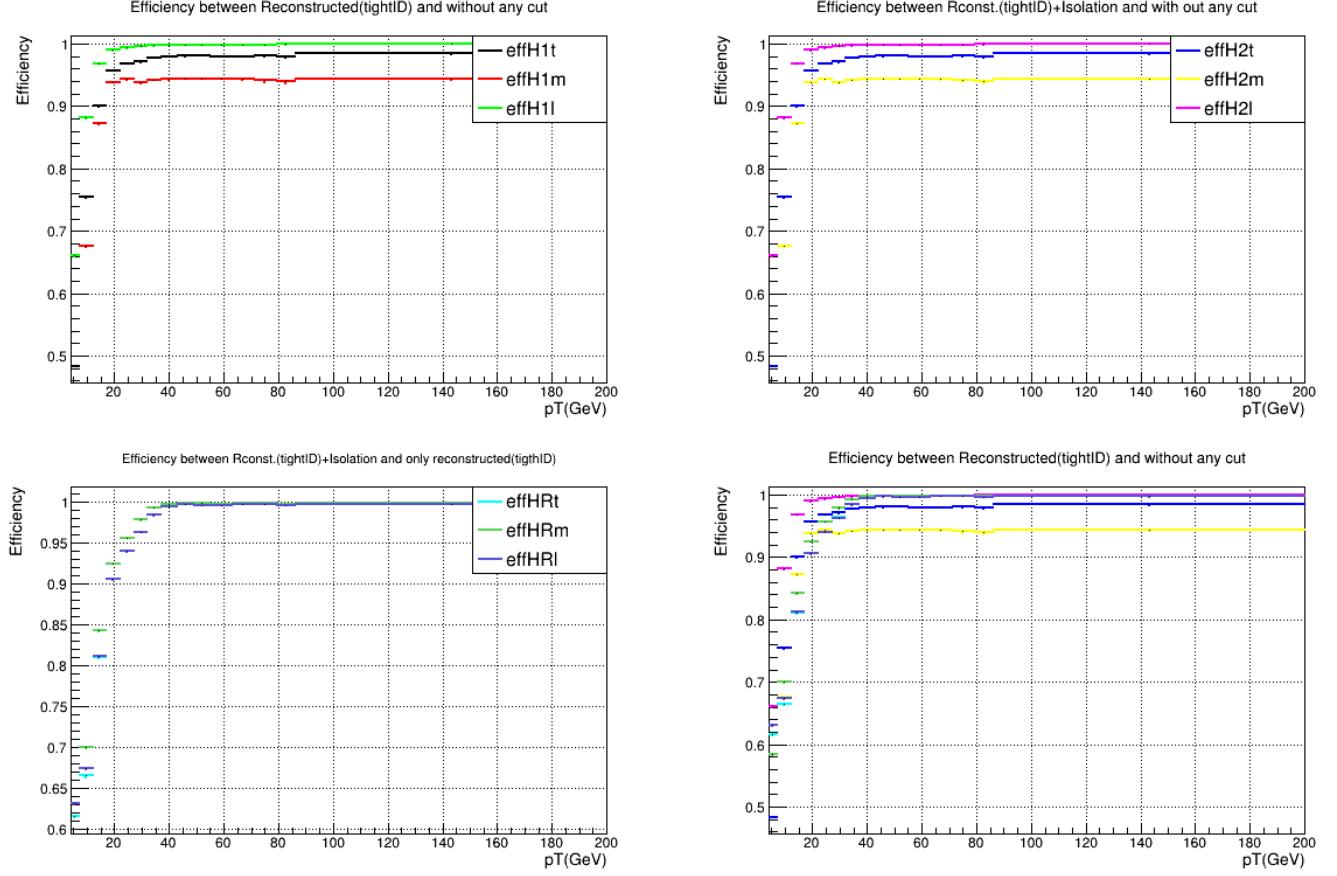
Veto distribution



After applying the vetoes, we have more purer sample. But from the P_T histogram, we can see fewer statistics in the higher P_T ranges. So, to reduce the error while deciding these momentum histograms, we re-bin it with variable bin width. After it looks like this:



After doing the re-binning, we can divide the P_T histogram without no cut to the P_T histogram with the first reconstruction cut and then also with the isolation cut. This gives us some ratio histograms. After a specific P_T value, saturation is observed; that ratio is the Reconstruction Efficiency. Some plots with different reconstruction IDs (tight, medium, loose) are given here.



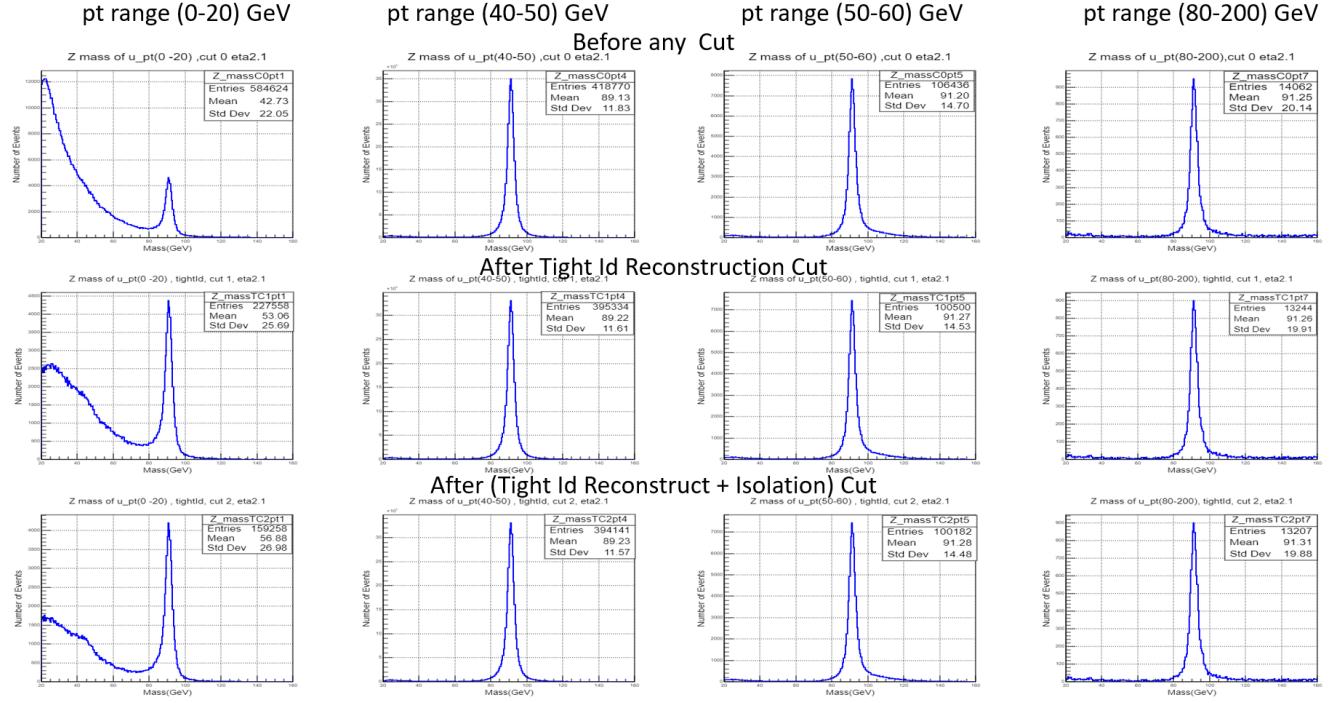
caption re-binned histogram with a variable bin width

So, from the histograms, we can see the efficiency is 94%.

Now apply the method 2 discussed above

Still, there may be some background, which we can't do with the other filtering conditions. We use the fitting method(method 2) for these kinds of backgrounds. We try to fit the signal and the background with the double-sided Crystal ball function; the peak model the signal, while the power law tail model the background. Using RooFit to do this fitting, we can extract the signal and background number and get the efficiency by taking the ratio of them. We take the invariant mass distribution in different P_T ranges. Here it is :

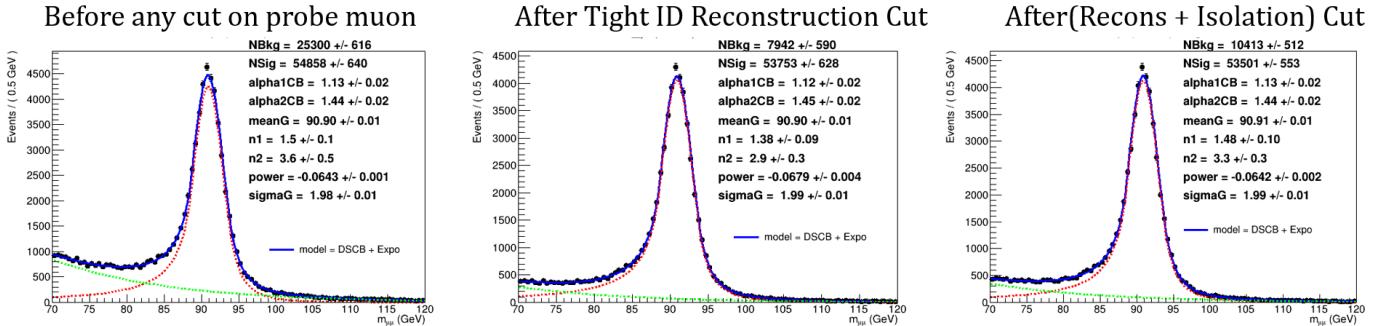
Invariant mass histogram of Z particle (Tag and Probe pair) for different pt ranges



Now, fitting these mass histograms in different P_T ranges, we get the number of signals after each cut and divide them to each other conservatively, and we get the efficiency again.

After fitting with RooFit,

Signal and Background: probe pt range(0-20)



Thus, we can fit the histogram and get the number of signals for each range. The result from the fitting is given in the table below:

Efficiency Table						
Sl. No.	P_T ranges	Without any cut : no. of Signal(N_s) (cut 0)	With only tight recons. cut : no. of Signal(N_s) (cut1)	With recons.+isolation cut: no. of Signal(N_s) (Cut 2)	Efficiency after Cut1	Efficiency after Cut2
1	0-20	54858	53501	52968	0.975	0.965
2	20-30	148307	142784	140487	0.962	0.947
3	30-40	329268	311223	307709	0.945	0.934
4	40-50	395199	374229	369996	0.946	0.936
5	50-60	97902	92630	92402	0.946	0.943
6	60-80	40352	36978	38069	0.916	0.943
				Average	0.948	0.945

So, from the fitting, we also get the same efficiency. That is, the result from method 1 is satisfied with the result of method 2.

7 Conclusion

7.1 Summary of Findings and Contributions

- The objective of this project was to measure muon detection efficiency using the Tag and Probe method using CMS data.
- By measuring efficiency, we can evaluate detectors' capabilities and identify areas of improvement, cross-section measurement, and look for new physics.

7.2 Observations

- The efficiency of the loose ID cut was the highest, and that of the tight ID cut was the lowest.
- Efficiency after the isolation cut was not affected too much (slight change).
- Efficiency was slightly higher for the higher value of eta.
- Background contribution in the low pt range was significant, but in the higher pt range, it was low.

References

- [1] Allan Jales1,Thomas Gaehtgens. *Muon efficiency studies using Tag and Probe method*. 2020. URL: <https://espace.cern.ch/lip/pub/docs/LIP-STUDENTS-20-18.pdf>.
- [2] *Efficiency Studies using the Tag and Probe Method*. URL: <https://cms-opendata-workshop.github.io/LIP-STUDENTS-20-18workshop-lesson-tagandprobe/aio/index.html>.
- [3] Pedro J. Fernández Manteca. *Muon Identification and Isolation efficiencies on Run II data with the CMS experiment*. URL: https://indico.cern.ch/event/681549/contributions/2956802/attachments/1661977/2663023/Poster_TnP-2.pdf.
- [4] Rodrigo Campello. *Muon efficiency for Z boson with the tag and probe method using CMS open data*. 2021. URL: https://indico.lip.pt/event/936/contributions/3642/attachments/2990/4592/muon_cms_openData.pdf.
- [5] *Tag and Probe*. URL: http://nuhep.northwestern.edu/~ksung/ForYihan/tutorial_tnp.html.
- [6] *The CMS experiment at the CERN LHC*. URL: <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004>.