Developing a Search for Dark Matter Direct Production using Razor Variables in Proton-Proton Collisions at 13 TeV

Jared Filseth

Mentors: Harvey Newman, Maria Spiropulu, Javier Duarte, Cristian Pena

Abstract

Dark matter makes up 27% of the mass-energy density of the universe, but still remains elusive today. In recent years, searches for dark matter production in proton-proton (pp) collisions have been carried out at the Large Hadron Collider up to a center of momentum energy of 8 TeV. In this project, we improve and reengineer the analysis methods used on 8 TeV data, and apply them to new 13 TeV data and simulations of new models of dark matter production. These methods involve looking at the so-called razor variables, which are used to quantify the mass scale of hypothetical new particles that may have been produced, as well as the imbalance in transverse momenta in each collision. In order to validate and update these methods, we first measure quantities such as the trigger and lepton efficiencies for simulated events as well as Compact Muon Solenoid (CMS) data. The studies performed in this project aim to improve the methods used to reconstruct the LHC data using the razor variables, and to provide a foundation for the 13 TeV search for dark matter once a sufficient integrated luminosity has been recorded. [1]

Background

While the properties of dark matter are largely unknown, supersymmetric extensions of the standard model provide a potential candidate for dark matter particles (the lightest supersymmetric particle, X). However, as of now, there is no evidence for the existence of supersymmetric particles. According to some supersymmetric theories and alternative dark matter models, high-energy proton-proton (pp) collisions will sometimes result in the production of a top quark/antiquark pair and two dark matter particles.

In recent years, these types of collisions have taken place at the Large Hadron Collider (LHC). Dark matter particles, which are electrically neutral and weakly interacting, cannot be directly observed by the particle detectors at the LHC. Instead, it is predicted that dark matter particles could be detected indirectly in collisions with large missing transverse energy (MET), by observing the distributions of the razor variables R^2 and M_R, which are calculated from the momenta of the outgoing particles of the collisions. In previous years, pp collisions have been produced at center-of-momentum energies up to 8 TeV[1].

In 2015, the LHC began producing pp collisions at 13 TeV, providing a greater potential source of evidence for supersymmetric and dark matter particles, covering a larger range of such particle masses than ever before. In the following, we test the methods that are to be used in the search for direct dark matter production at 13 TeV. The aim of these studies is to improve the methods, and provide a foundation for the searches for dark matter in

CMS using the razor variables, during the present data taking run (known as LHC Run 2) that is underway at the time of this writing.

The main standard model background to the dark matter model described previously is the production of a top-antitop quark (ttbar) pair plus extra jets of particles. [1] In these events, the top quarks decay predominately to W bosons accompanied by an anti-bottom quark. The W bosons then decay into either a lepton and a neutrino, or 2 quarks. This mechanism is shown in Figure 1a. However, in the dark matter model, this collision will occasionally result in the additional production of a Φ mediator particle, which produces 2 dark matter X particles. This hypothesized event is shown in Figure 1b.

As of yet, it is unknown whether any of these dark matter events are present in LHC ttbar events, since their presence would be difficult to detect, as there are far more background ttbar events, making filtering the signal difficult. However, there are certain variables, called "razor variables"[1], which are related to the mass scale of the hypothesized new heavy particles and the missing transverse energy of the events, and are expected to help discriminate dark matter signal events from ttbar background events. These variables are illustrated in Figure 2, which shows the normalized distributions of the razor variables M_R and R^2, calculated from Monte Carlo (MC) generated background and signal simulated events. As shown, the presence of dark matter collision events at the LHC would be evidenced by M_R or R^2 distributions in CMS data which deviate significantly from the expected background distribution In particular, a distribution with an unexpectedly large number of high M_R and R^2 events could signal the presence of dark matter particles.

In order to perform a search for new physics, the data is usually divided into different categories with different signal and background compositions. Analyzing these categories separately would determine if the dark matter signal is present across multiple categories with different backgrounds. For this search, we use a leptonic category, in which the event contains at least one lepton, and a hadronic category, in which the event contains no leptons. For the hadronic analysis, which consists of events without electrons or muons, we validated the jet reconstruction algorithms to ensure that the reconstructions match the generator-level jets[1] with sufficient accuracy, within the statistical and systematic errors. Additionally, for the hadronic analysis, we measured the efficiency in the simulation of a new trigger algorithm, which filters out events with an R^2 value less than 0.36. In the leptonic category, which includes collisions involving electrons and muons, the tag and probe method was used to determine the amount of mis-modeling of the lepton triggers in the MC sample and correct the lepton efficiency of the selection by using a scale factor.

[1] In particle physics experiments, the Monte Carlo simulation procedure includes two steps. The four momenta of the final state particles are produced by "Generator" programs, and the generated particles are then propagated through a detailed representation of the particle detection apparatus. This results in simulated detector responses in digitized form which are then analyzed in the same manner as actual data from the experiment.

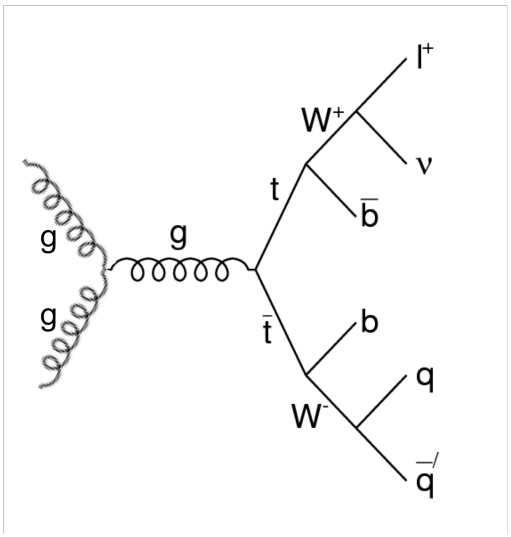

Figure 1a. [2] Feynman diagram for typical background ttbar collision. Two gluons (from the protons) collide, producing a pair of top and antitop quarks. These quarks proceed to decay into W bosons and bottom quarks. The W bosons may then either decay into quark-antiquark pairs (hadronic events), or into a lepton and a neutrino (leptonic events).

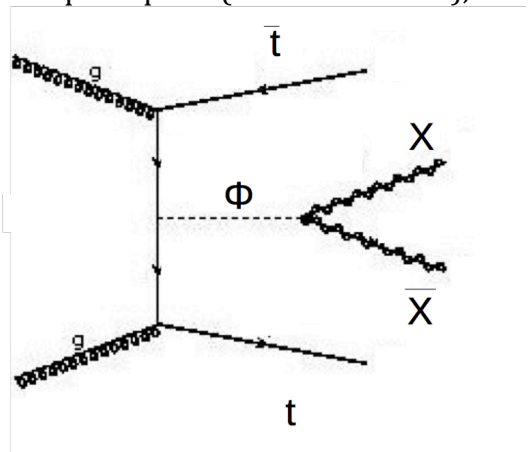

Figure 1b. [3] Feynman diagram for dark matter event. This collision is almost identical to the one shown in Figure 1a (the decay products of the t quarks are omitted), except for the production of 2 dark matter X particles.

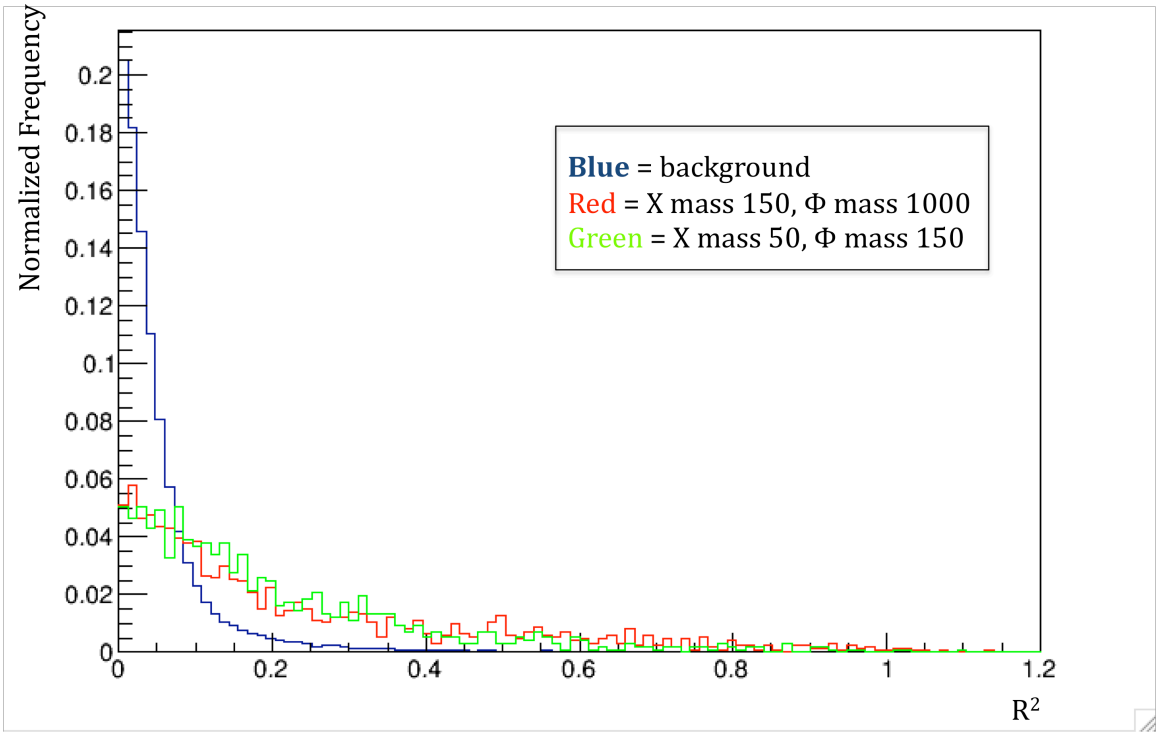

Figure 2a. This histogram shows the normalized R^2 distributions for MC simulated dark matter signal and background events, and demonstrates the difference between dark matter signal distributions, shown in red and green for different masses, and the ttbar background distributions, shown in blue. The signal events tend to have higher R^2 values regardless of the masses of the dark matter and mediator particles, so the presence of high R^2 events in LHC collisions would be indicative of the production of dark matter particles.

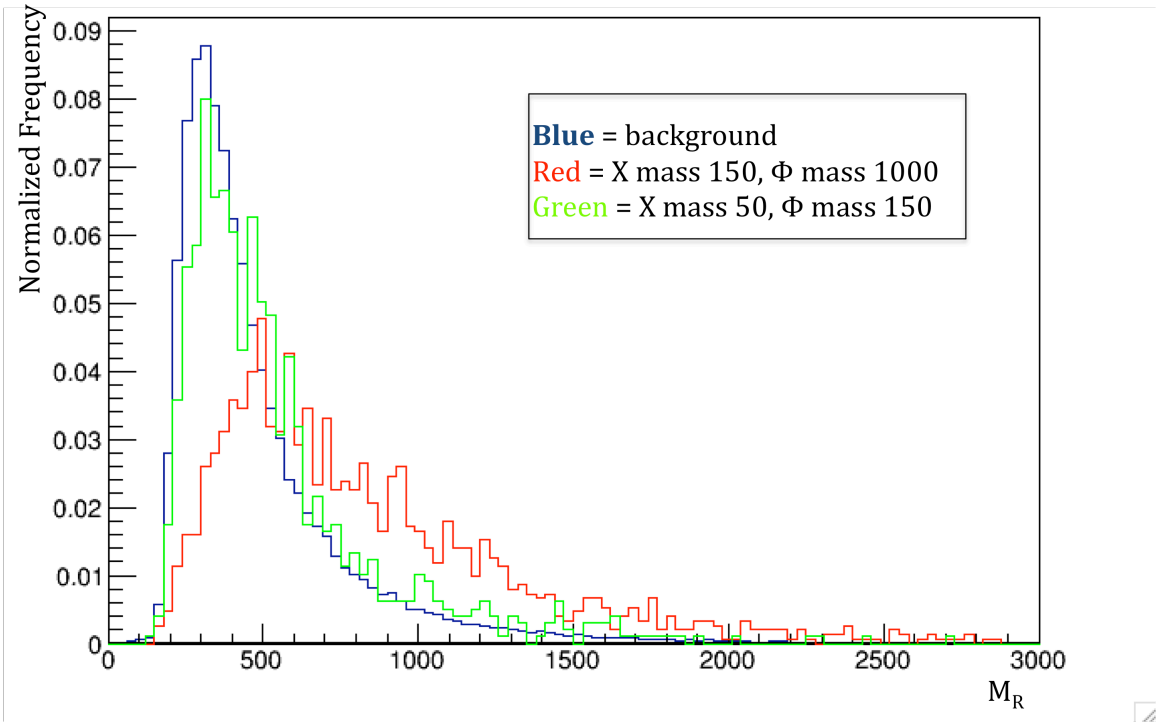

Figure 2b. This histogram shows the normalized M_R distribution for MC simulated dark matter signal events, shown in red and green for different masses, and the ttbar background events, shown in blue. For low masses of X and Φ, there is not much difference between the background and signal distributions, which would make it difficult to determine the presence of a dark matter signal from the M_R distribution in these cases. However, for higher masses of X and Φ, the M_R distribution peaks at higher values, so a large proportion of high-M_R events would be evidence for a dark matter signal.

Hadronic Jet Reconstruction Analysis

At the LHC, pp collisions may produce a top-antitop quark pair (ttbar events), which then further decay into other particles, resulting in the production of jets[2], as well as leptons. During the reconstruction phase, the generated pp collision events are recorded by various particle detectors and reconstructed online, giving an estimate of the identities and momenta of the particles produced in the collision.

To test the algorithms used to perform the reconstructions, MC generated dark matter ttbar collision events are used. These dark matter events involve 2 hypothetical undiscovered particles: a dark matter candidate (X) and a mediator particle (Φ). As the masses of the X and Φ particles are unknown, 4 different batches of MC generated events were used, each with a different combination of masses of X and Φ. The reconstruction algorithms were used on the generated events, and the reconstructed events were compared to the generated events. Specifically, the accuracy of the jet reconstruction was tested for the 4 models.

Each reconstructed jet is represented by a four-vector containing energy, transverse momentum (momentum in the plane perpendicular to the axis on which the particles travelled prior to collision), and two angle quantities (η and φ), which determine the direction in which the jet travels: φ is the azimuthal angle, which measures the angle of the jet in the plane perpendicular to the collision axis, and η is the pseudorapidity, which is related to the angle of the jet relative to the collision axis. To test the reconstructions, the reconstructed jets were matched up to the MC generated jets by angles, and then the accuracy of the reconstructions was tested by comparing each reconstructed jet's transverse momentum to that of the matching generated jet.

The results are shown in Figures 3-6 below. These all generally suggest a slight bias for smaller reconstructed transverse momenta, but no bias for any particular angle. Thus, the 13 TeV reconstruction algorithms seem to work well for supersymmetric models of high-energy pp collisions, though improvement may be possible and further study is needed.

[2] Quarks and antiquarks produced in a collision are "confined" [4] such that they do not appear as stable particles in nature. Instead they pull additional quark-antiquark pairs from the vacuum, resulting in "jets" of particles in the final state. Gluons produced in the collision each decay to a quark-antiquark pair, resulting in two final state jets.

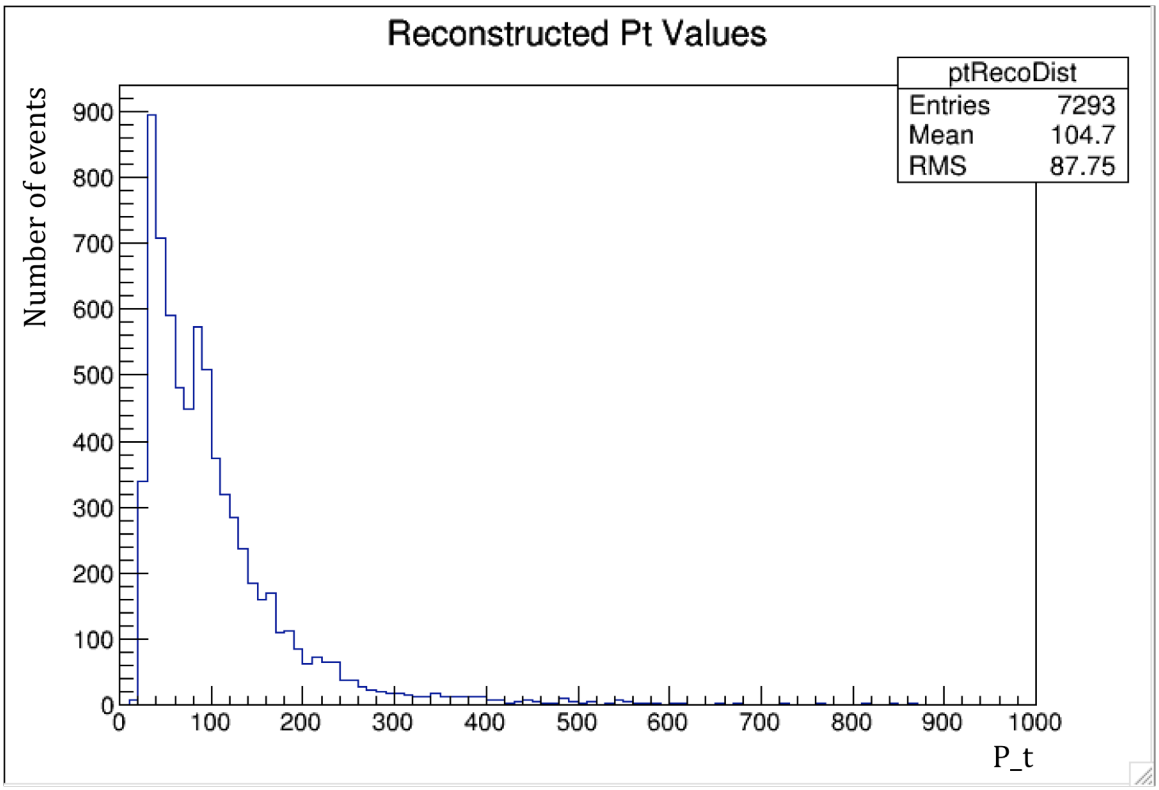

Figure 3a. This histogram shows the distribution of reconstructed P_t values across all events for masses X=150, Φ=1000. The distribution is largely clustered towards small P_t values, but there are a few very large values with P_t close to 1000. This allows for an analysis of the effectiveness of the reconstruction methods by comparison to the distribution of generated P_t values. One odd feature of this graph is the peak around 100 GeV; however, in order to explain this feature, it is necessary to consider the original distribution of generated values to see whether this peak could reasonably be expected to be present.

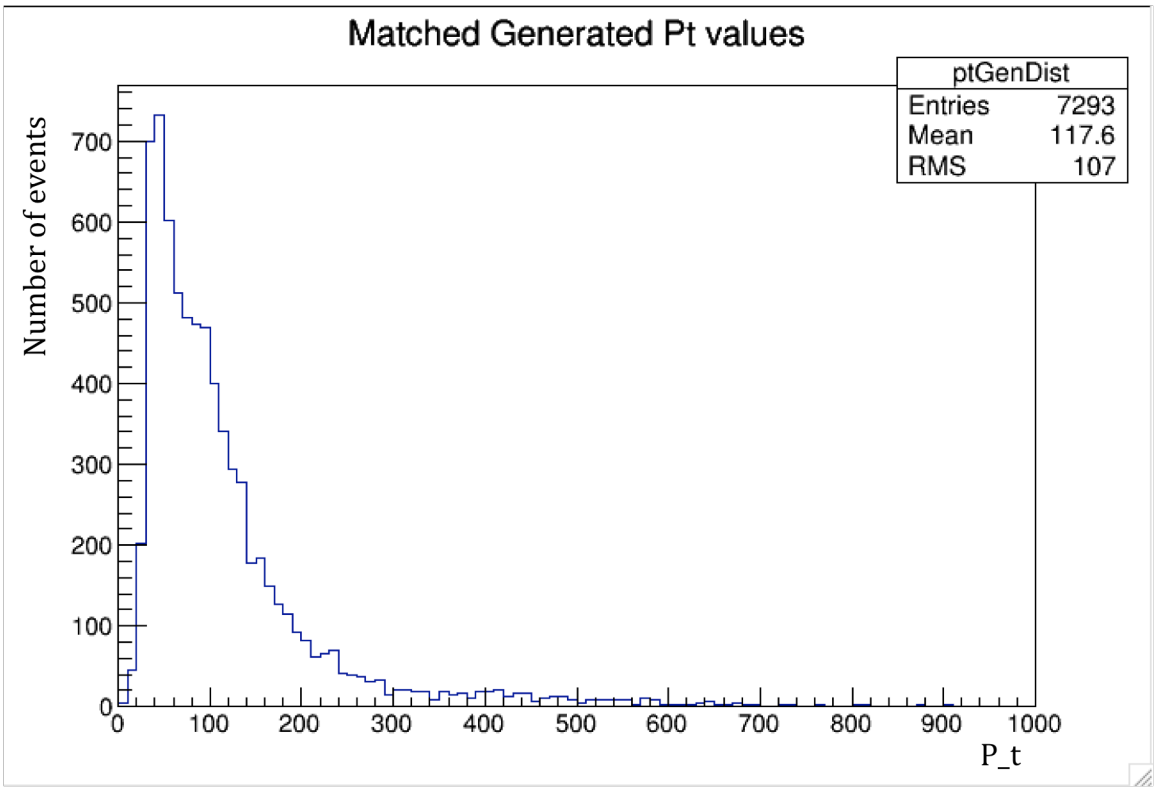

Figure 3b. This histogram shows the distribution of the generated P_t values which were matched to a reconstructed value across all events for masses X=150, Φ=1000. The distribution is largely clustered towards small P_t values, but there are a few very large values with P_t close to 1000. This allows for an analysis of the effectiveness of the reconstruction methods by comparison to the distribution of reconstructed P_t values. Notably, the peak around 100 GeV that was present in Figure 3a is not present here, but there is a shallow plateau around that area. Thus, the peak in Figure 3a is not indicative of any sort of error or bias in the reconstruction algorithm, as there is an unexpectedly high concentration generated values around 100 GeV, so it would be expected to have a similar feature in Figure 3a. The differences between the 2 graphs are likely due to resolution, as they have mostly similar features and there are no clear biases.

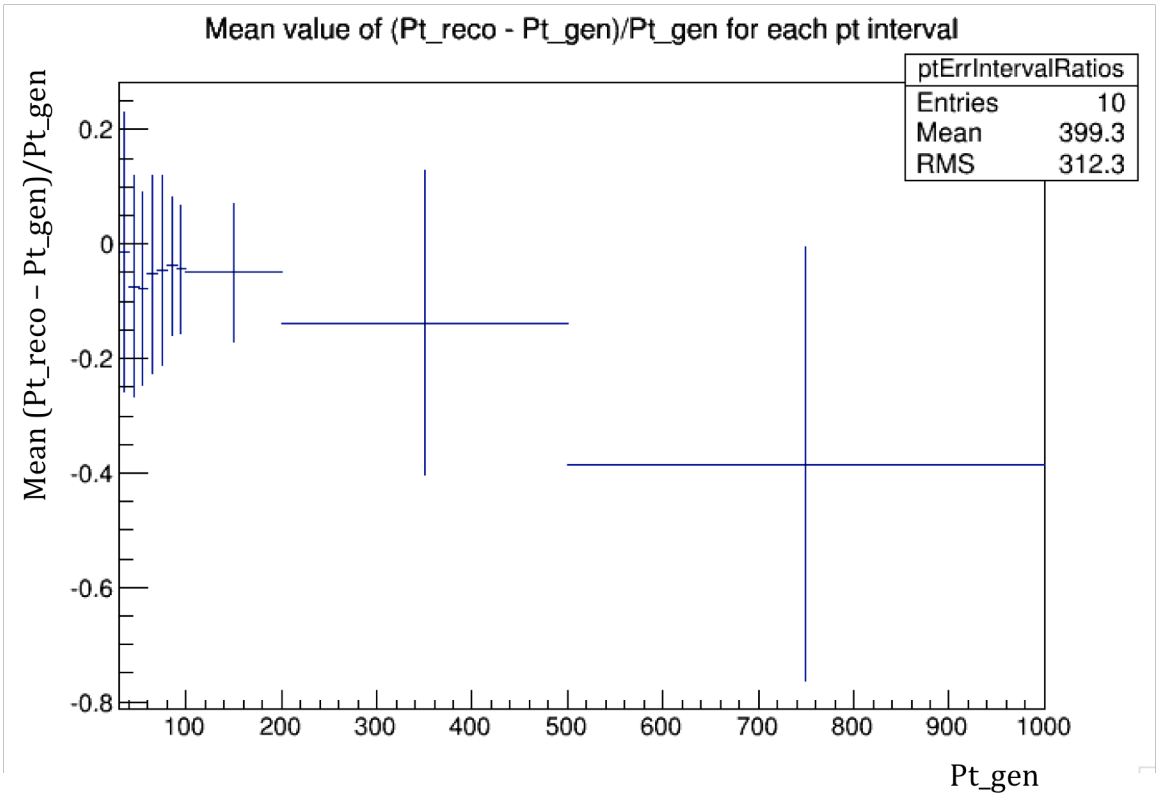

Figure 3c. This histogram shows the resolutions in the reconstructed transverse momentum with masses of 150 and 1000 GeV for X and Φ, respectively. On the x-axis, this histogram shows the transverse momentum of the matched generated jets (in GeV); that is, $P_{T\_reco} - P_{T\_gen}$, where P_{T_reco} is the transverse momentum of the reconstructions, and P_{T_gen} is the transverse momentum of the generated jets. The y-axis shows the mean resolution. The error bars all pass through 0, suggesting little evidence for bias. However, since the mean is slightly negative, the reconstructions on average have smaller magnitudes than the generated jets. Similar results were obtained for the other combinations of X and Φ masses. The bins in this histogram were chosen so that the number of events in each bin were roughly similar.

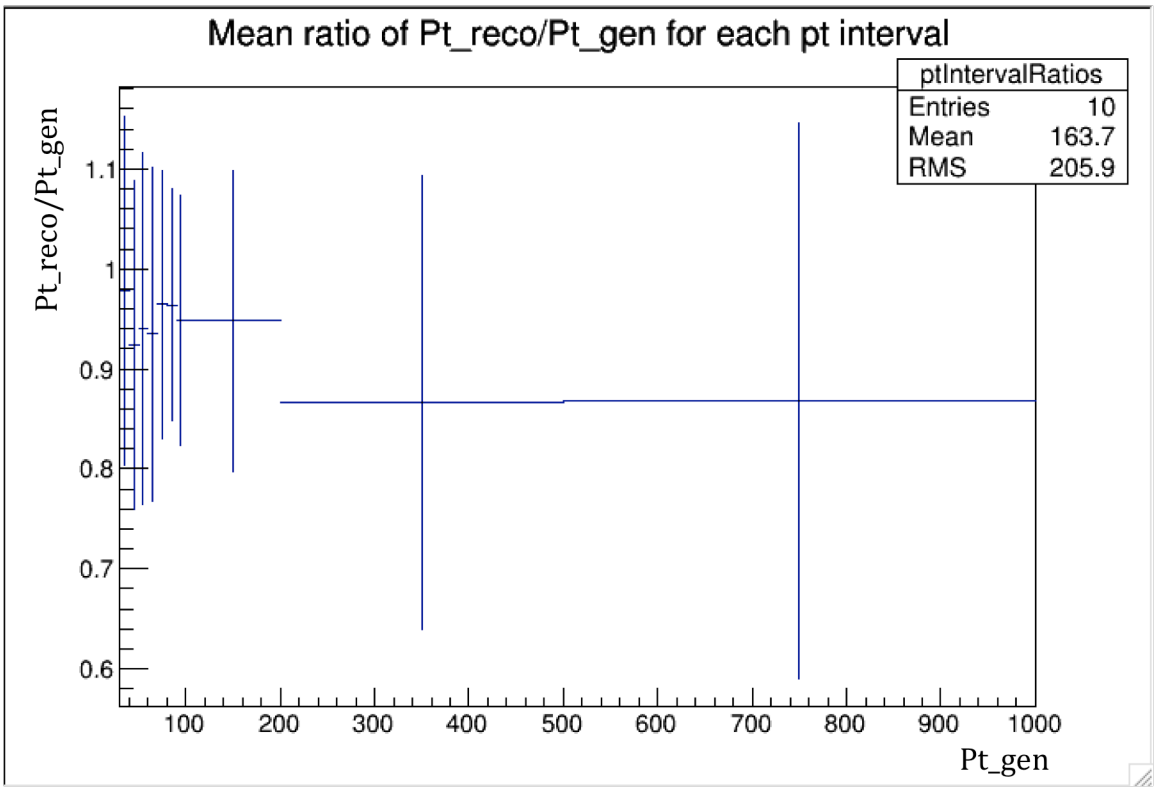

Figure 4a. This graph shows the mean ratios of the reconstructed transverse momentum to the generated transverse momentum for each generated transverse momentum interval, where the masses of X and Φ are 150 and 500, respectively. The x-axis shows the generated transverse momentum in GeV, and the y-axis shows the ratio. The results are fairly inconclusive from this graph due to the large errors; however, they do not appear to show any severe bias, as all ratios are larger than 0.85, and accounting for the error bars, there is little to no apparent bias. Similar results were obtained for the MC batches with the masses of X and Φ as 50 and 500 respectively, and 50 and 150 respectively. None of these batches show significant deviations from 1 or trends in the ratio, and hence no evidence for bias.

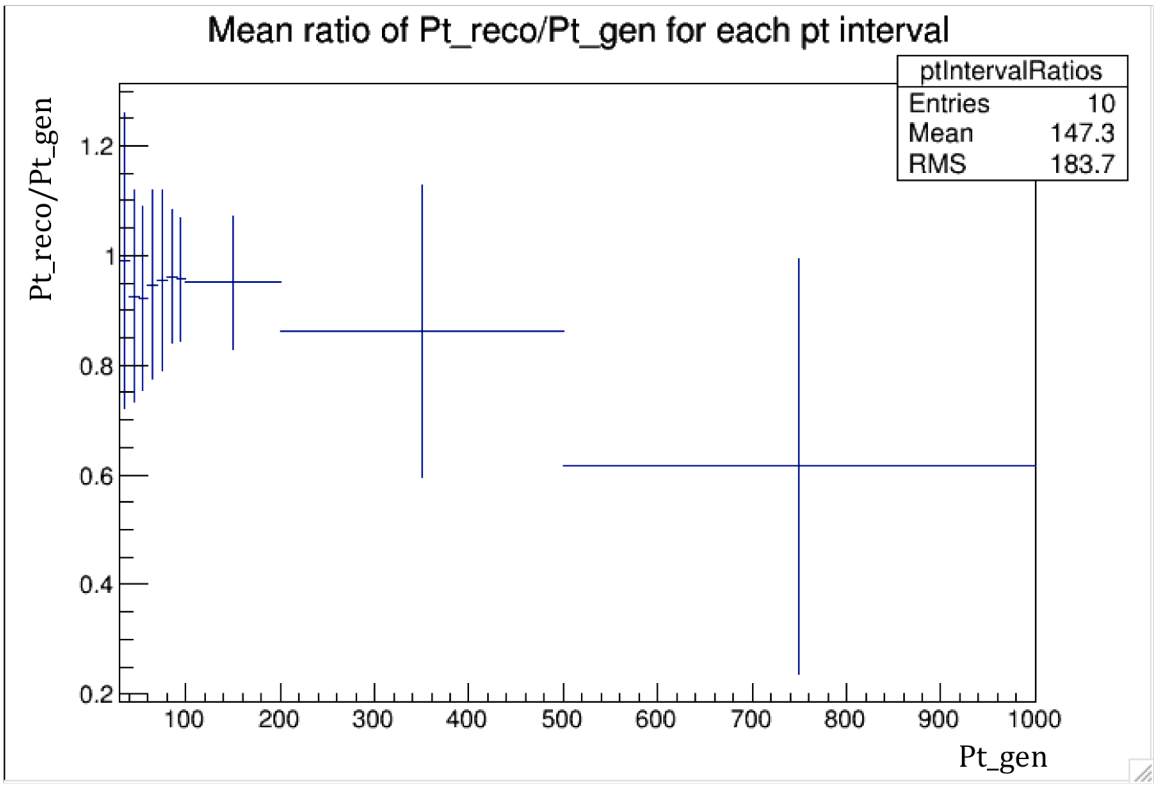

Figure 4b. Similar to Figure 4a, this graph shows the mean ratios of the reconstructed transverse momentum to the generated transverse momentum for each generated transverse momentum interval, where the masses of X and Φ are 150 and 1000, respectively. The x-axis shows the generated transverse momentum in GeV, and the y-axis shows the ratio. While the graph shows a possible bias at larger transverse momenta, with a mean around 0.6 in the rightmost bin, the results include 1 within one standard deviation. While there is no significant evidence for a bias at present, the graph indicates that further study using more data is important.

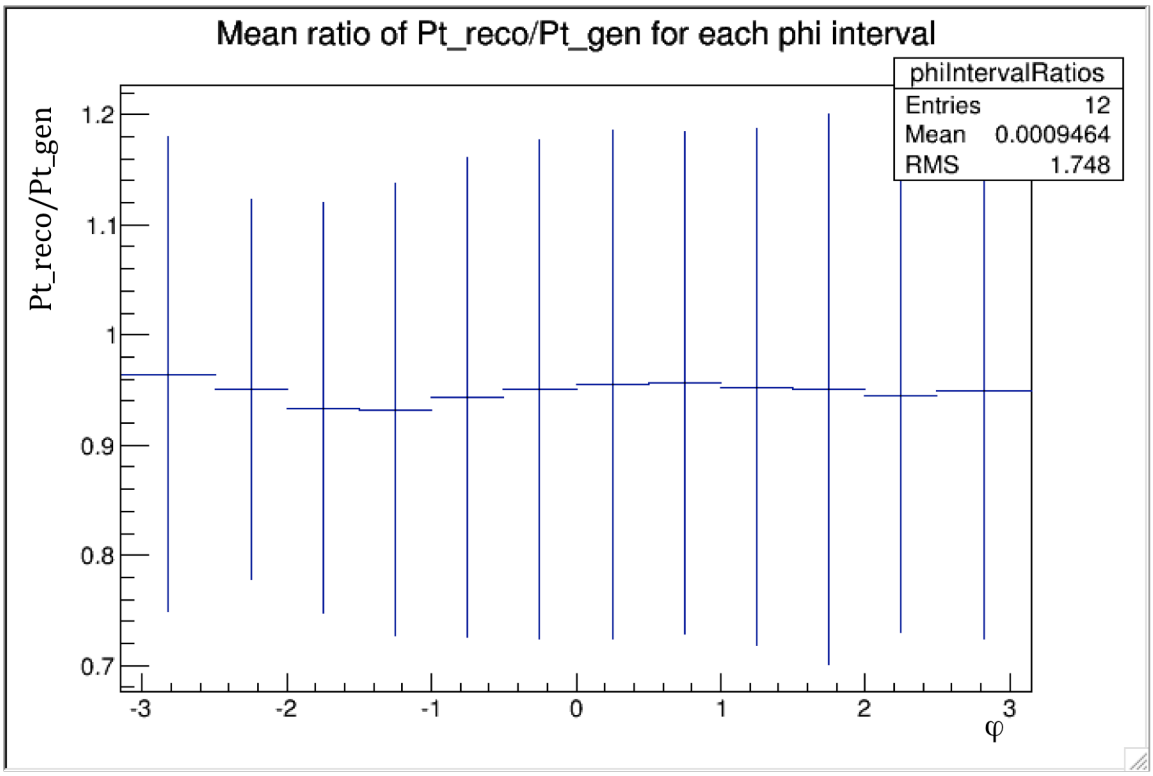

Figure 5. This graph shows the mean ratio of the reconstructed transverse momentum to the generated transverse momentum as a function of the angle φ (which ranges from −π to π), where the masses of the particles X and Φ are 150 and 1000, respectively. The x-axis contains the angle φ, and the y axis contains the mean ratios. Although the mean ratio is less than 1 across all angles, suggesting a potential very small bias, there is no significant bias towards any particular angle. Similar results were obtained for all other combinations of X and Φ particle masses.

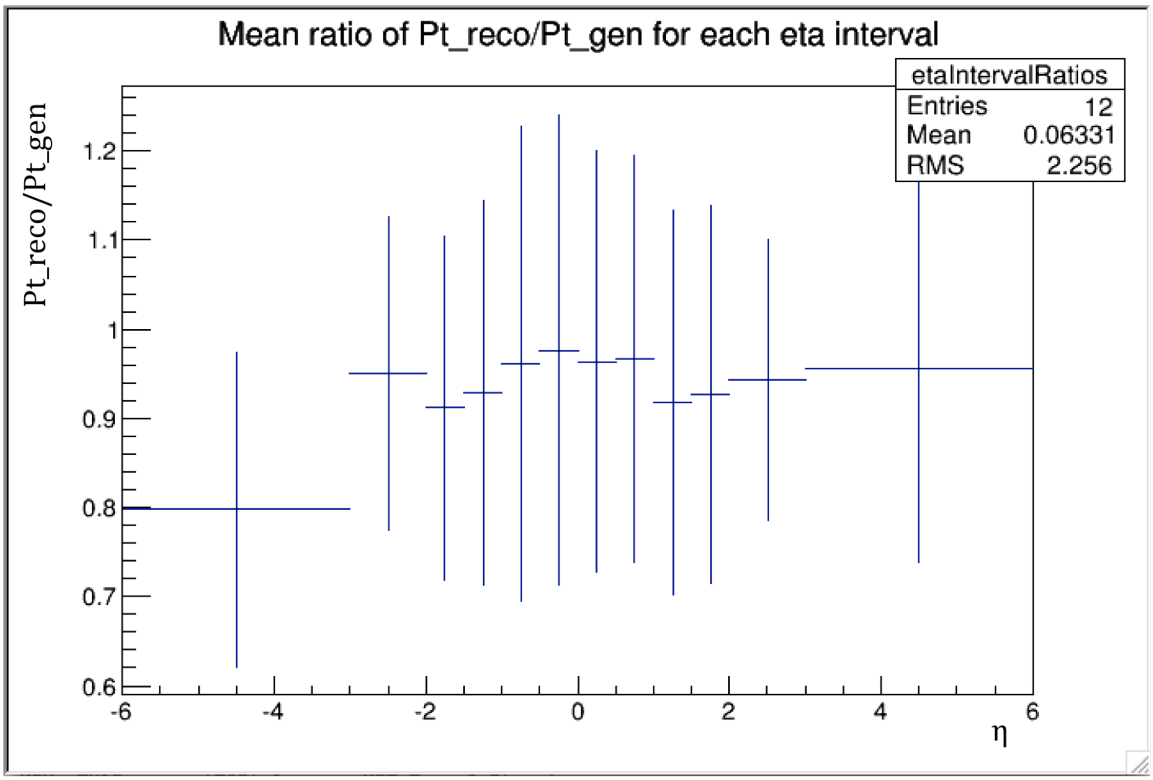

Figure 6. This graph shows the mean ratio of the reconstructed transverse momentum to the generated transverse momentum as a function of the angle η (which ranges from $-\infty$ to ∞, but is generally cut off at around ±3 due to the CMS detector acceptance), where the masses of the particles X and Φ are 150 and 1000, respectively. The x-axis contains the angle η, and the y axis contains the mean ratios. Although the mean ratio is less than 1 across all angles, suggesting a potential very small bias, there does not appear to be any bias towards any particular angle. Similar results were obtained for all other combinations of X and Φ particle masses.

Hadronic Trigger Turn-On Analysis

In particle physics, the "trigger" consists of a set of decisions taken in real-time, using specialized electronics and general purpose computer servers, to select which events are to be stored for further offline processing and analysis. At the CMS detector, there are two levels of triggers [5], which record the presence of certain types of particles produced in the collisions: Level 1 (L1) triggers; and high level (HLT) triggers, which are the focus of this analysis.

The high level trigger works by performing a fast reconstruction of the jets and leptons, then assessing whether the reconstructed event matches certain criteria. If so, the detector records the event. In this manner, it is possible to only look at events that meet certain criteria.

For the Run 1 razor search for dark matter production, which was conducted on the 8 TeV data collected in 2012, the Caltech/CMS group used an analysis with triggers dedicated to other purposes (searches for supersymmetric particles), but later applied an $R^2 > 0.5$ requirement to the data. [1] This selection was sufficiently sensitive, so a dedicated $R^2 > 0.36$ dark matter trigger was created for Run 2.

To estimate the selection efficiency of the trigger, the algorithm was used on MC simulated events with at least 2 jets of transverse momentum greater than or equal to 80 GeV. The results are shown in Figure 7 below, and indicate that the trigger algorithm is sufficiently sensitive: it excludes most events with low R^2 values but does not exclude events with higher R^2 values, and maintains this property across different M_R selections.

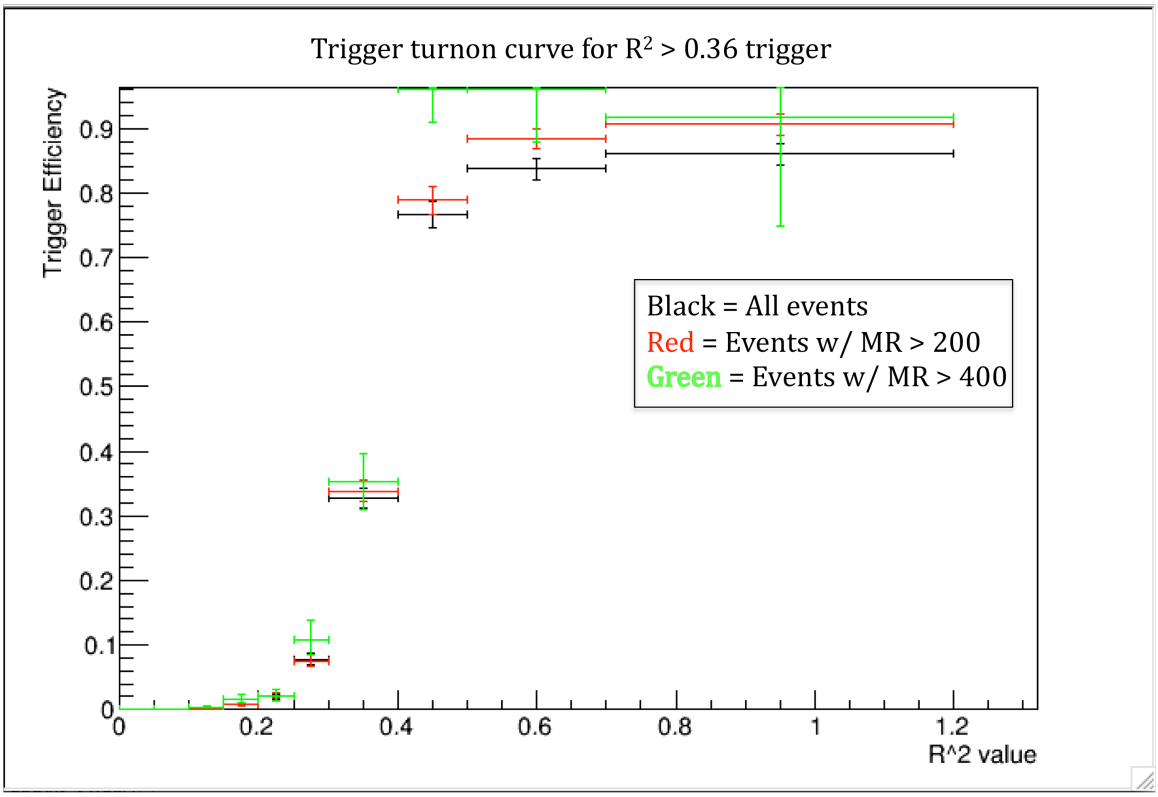

Figure 7. This histogram shows the trigger efficiency; that is, the ratio of the number of events with a given R^2 value that were accepted by the trigger to the total number of events with that R^2 value. The trigger efficiency is near 1 for high values of R^2 regardless of the M_R cut. However, a high M_R cut produced a curve with an efficiency closer to 100% at high R^2 values. Overall, this graph illustrates that the trigger is able to effectively filter out the low R^2 values, and therefore preferentially select the dark matter signals as desired.

Tag and Probe Analysis

In order to correctly simulate the single-electron and double-electron background for the MC simulations, it is necessary to know the scale factor of the lepton triggers; that is, the ratio of the trigger efficiencies of the MC simulations to the actual LHC data. This is done by calculating the efficiencies for the lepton triggers for LHC events and for MC simulated events, then taking the ratio of these 2 efficiencies. [6]

In order to calculate the lepton trigger efficiencies, the tag and probe method is used for muons and electrons. The tag and probe method uses events with at least two muons. Then, for each muon pair, 1 muon is selected as the "tag" and the other the "probe". The tag muon needs to pass the tight identification (ID) and isolation filters, which verifies that it is relatively isolated (and thus unlikely to be a jet or some other product of a collision). Thus, since the probe is known to be an muon, this provides an unbiased sample of muons, and the efficiency of the high level trigger muon selection can be tested by finding the proportion of the probe muons which passed the trigger. [6,7,8] These proportions are shown in Figure 8. A similar process was done for electrons, which is shown in Figure 9. The scale factors are shown in more detail in figures 8c and 9c.

A similar approach was taken to find the efficiency of the Tight ID/Isolation filters themselves; the proportion of the probe muons that passed the filters was measured and plotted in figure 10. The proportion of the probe electrons that passed the Tight ID/Isolation electron filters is plotted in figure 11. The scale factors are shown in more detail in figures 10c and 11c.

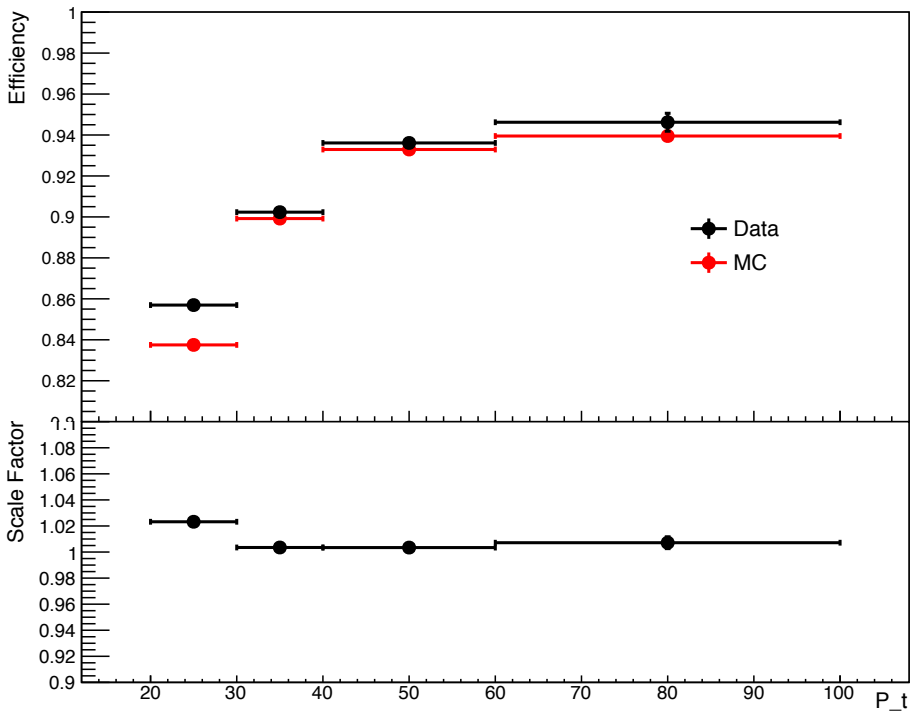

Figure 8a. This figure shows the efficiency of the muon HLT filters for the data and MC simulations, as well as the resulting scale factor, as a function of the transverse momentum of the muons. The efficiency tends to be noticeably lower at lower transverse momenta; however, the scale factor is slightly higher, but still very close to 1.

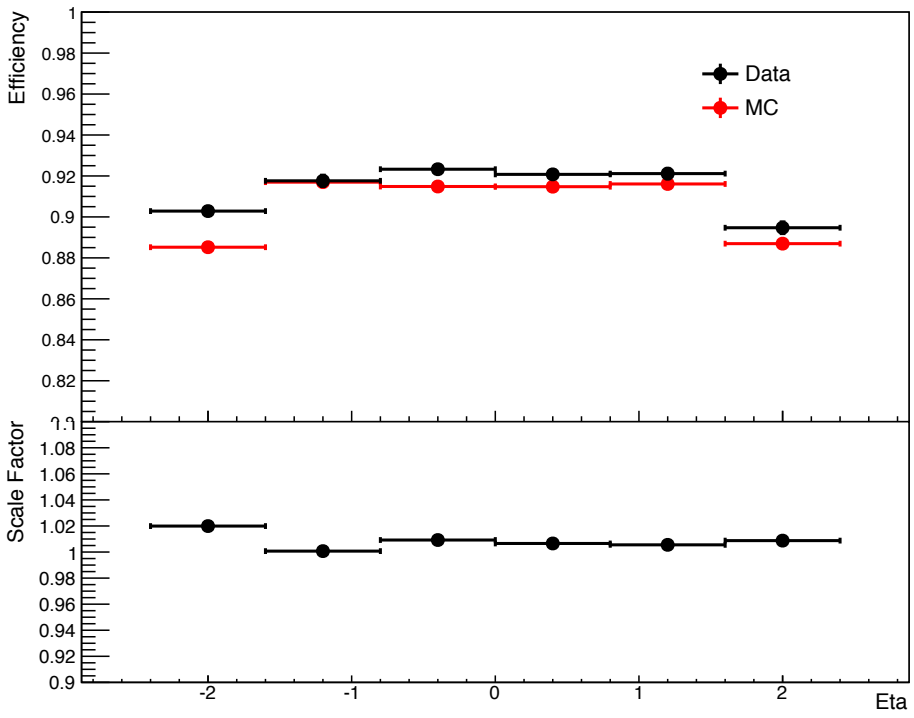

Figure 8b. This figure shows the efficiency of the muon HLT filters for the data and MC simulations, as well as the resulting scale factor, as a function of the angle η. The efficiency tends to be noticeably lower at high magnitudes of η, which correspond to the endcap of the CMS detector, rather than the barrel.

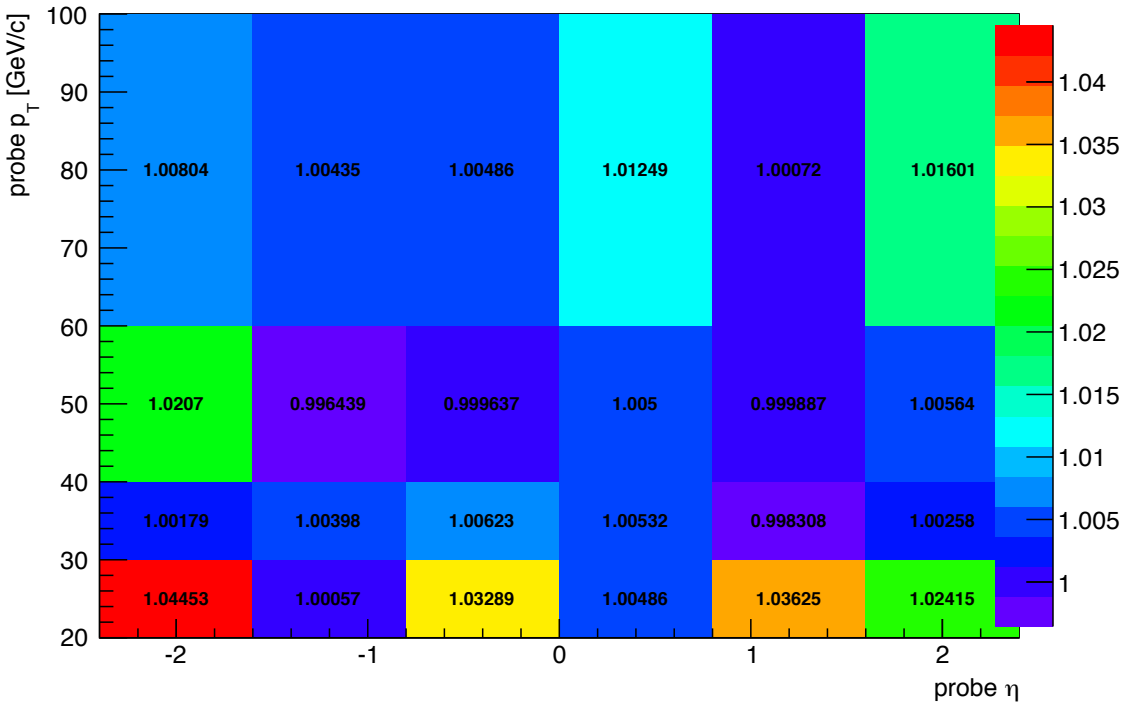

Figure 8c. This 2d graph shows the muon HLT filter scale factors as a function of the transverse momentum and the angle η. The scale factors are roughly constant across all values of P_T and η, though they are slightly higher at the endcaps (high magnitudes of η).

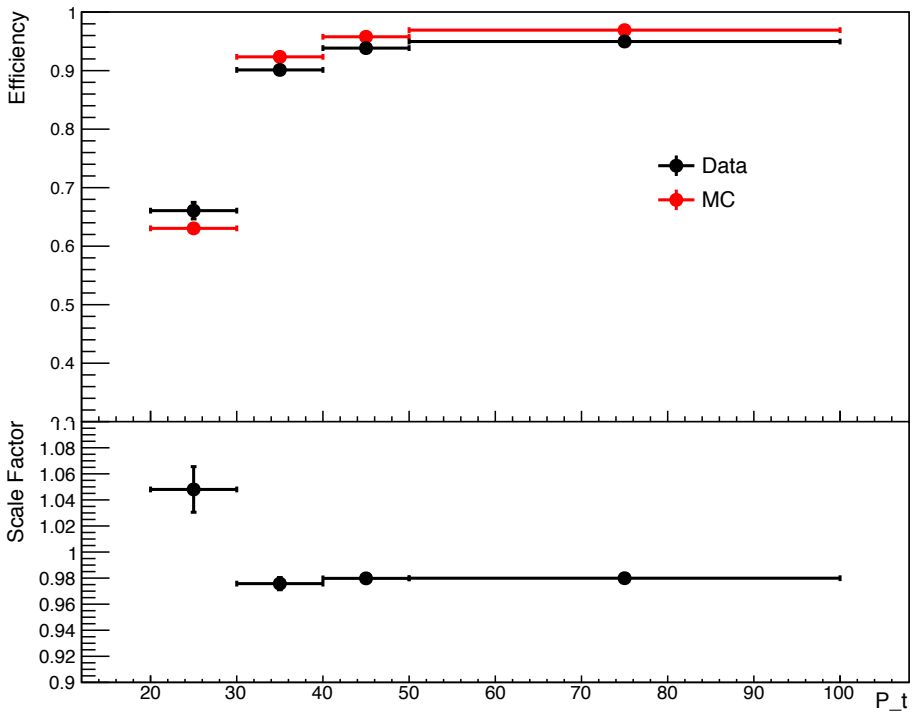

Figure 9a. This figure shows the efficiency of the electron HLT filters for the data and MC simulations, as well as the resulting scale factor, as a function of the transverse momentum of the electrons. The efficiency tends to be noticeably lower at lower transverse momenta; however, the scale factor is higher, but still close to 1.

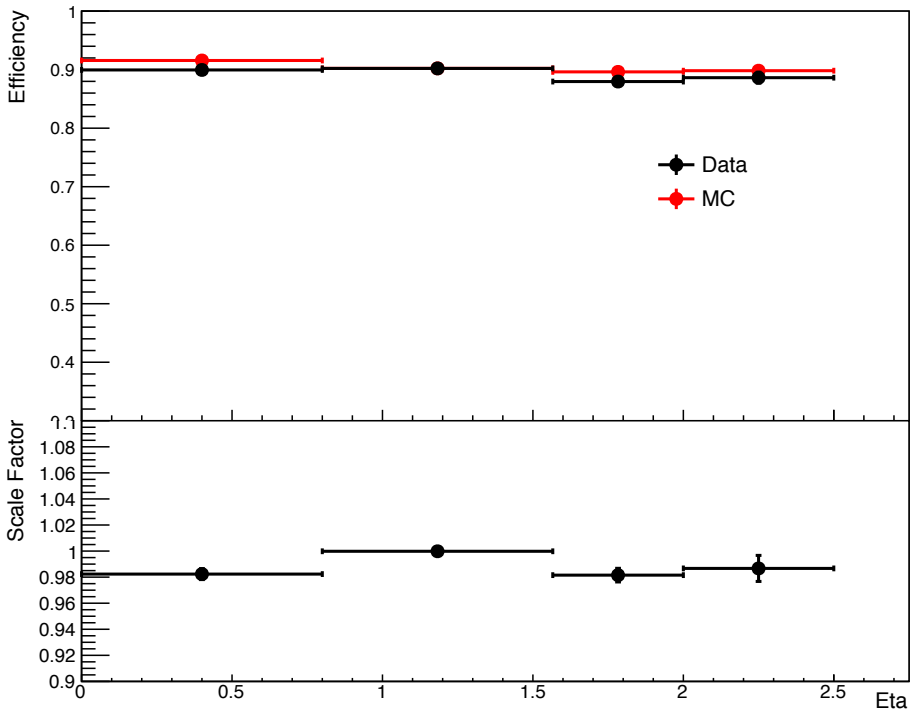

Figure 9b. This figure shows the efficiency of the electron HLT filters for the data and MC simulations, as well as the resulting scale factor, as a function of the angle η. Unlike the muon filter, there does not appear to be much difference in either the efficiency or the scale factor between the endcap and the barrel in the electron filter.

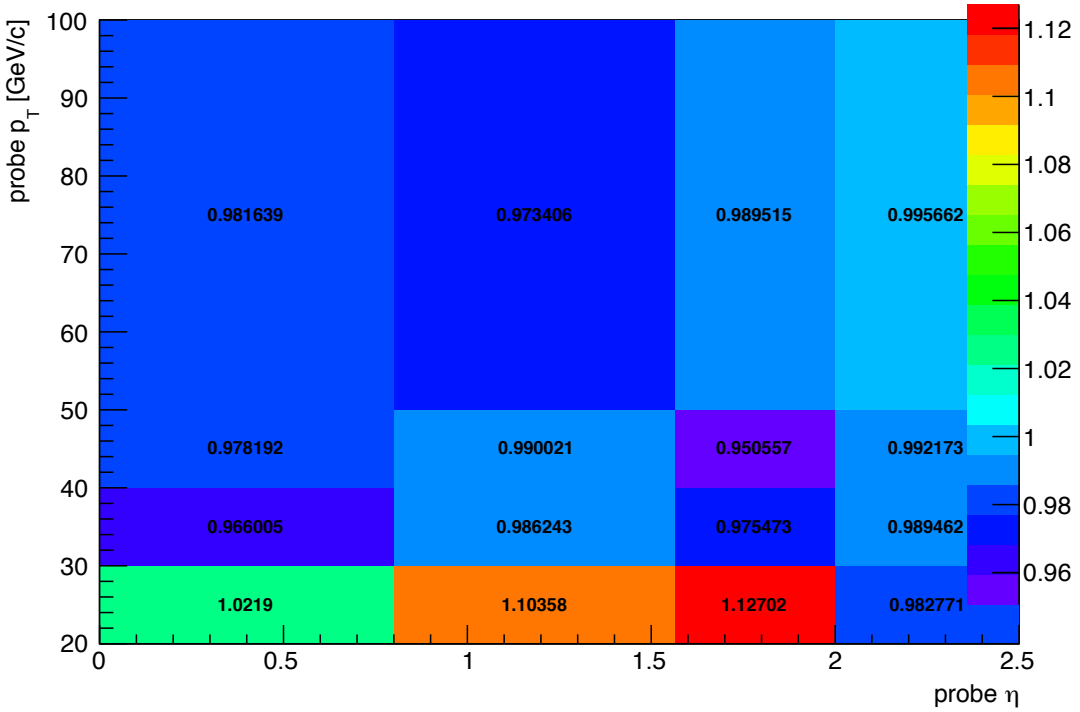

Figure 9c. This 2d graph shows the electron HLT filter scale factors as a function of the transverse momentum and the angle η. The scale factors are roughly constant across all values of P_T and η.

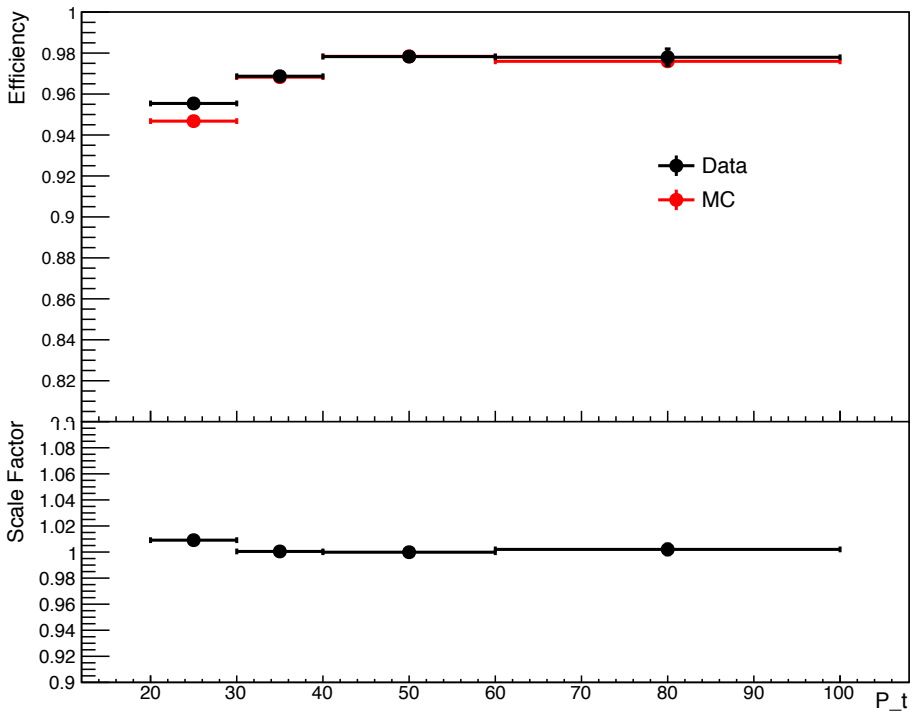

Figure 10a. This figure shows the efficiency of the tight ID/isolation muon filters for the data and MC simulation, as well as the resulting scale factor, as a function of the transverse momentum of the muon. The scale factor is calculated by dividing the efficiency for the data by the efficiency of the MC simulation. This scale factor appears to be very close to 1, but is slightly larger, particularly at lower transverse momenta.

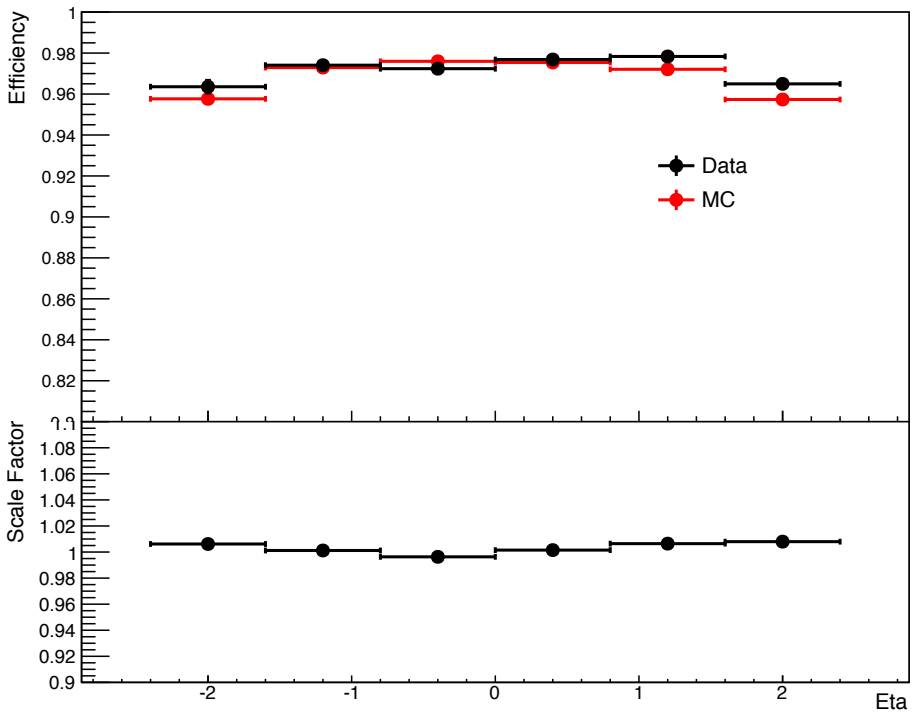

Figure 10b. This graph shows the efficiency of the tight ID/isolation muon filters for the MC simulations and data, and the resulting scale factor as a function of the angle η. There does not appear to be a large difference in the scale factors across values of η, though the efficiency tends to be slightly lower for both the data and the MC simulations at high values of η, which correspond to the endcap rather than the barrel of the Compact Muon Solenoid (CMS) detector.

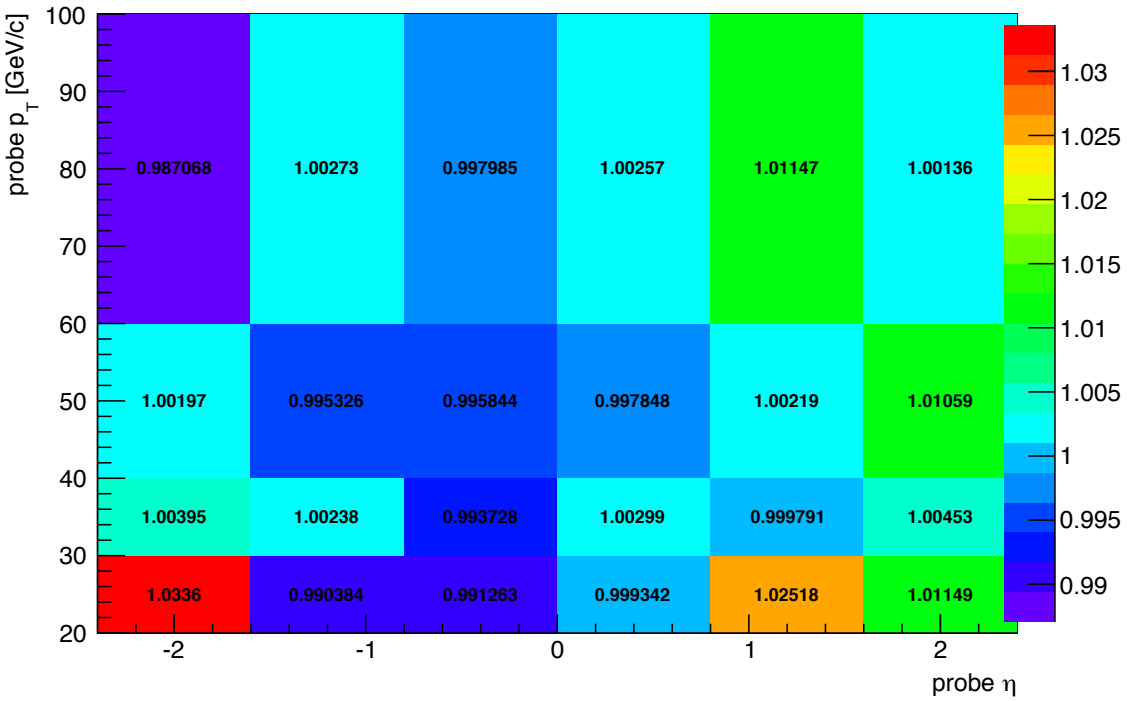

Figure 10c. This 2d graph shows the muon tight ID/isolation scale factors as a function of the transverse momentum and the angle η. The scale factors are roughly constant across all values of P_T and η, though they are slightly higher at the endcaps (high magnitudes of η).

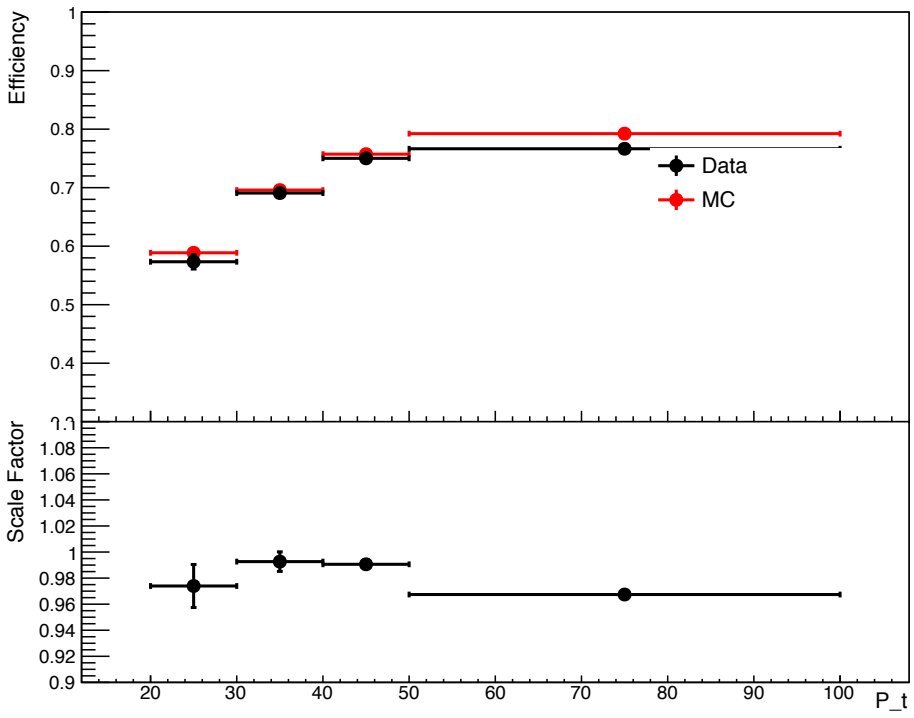

Figure 11a. This figure shows the efficiency of the tight ID/isolation electron filter for the data and MC simulation, as well as the resulting scale factor, as a function of the transverse momentum of the electron. The scale factor is calculated by dividing the efficiency for the data by the efficiency of the MC simulation. This scale factor appears to be very close to 1, but is slightly smaller at low and high transverse momenta.

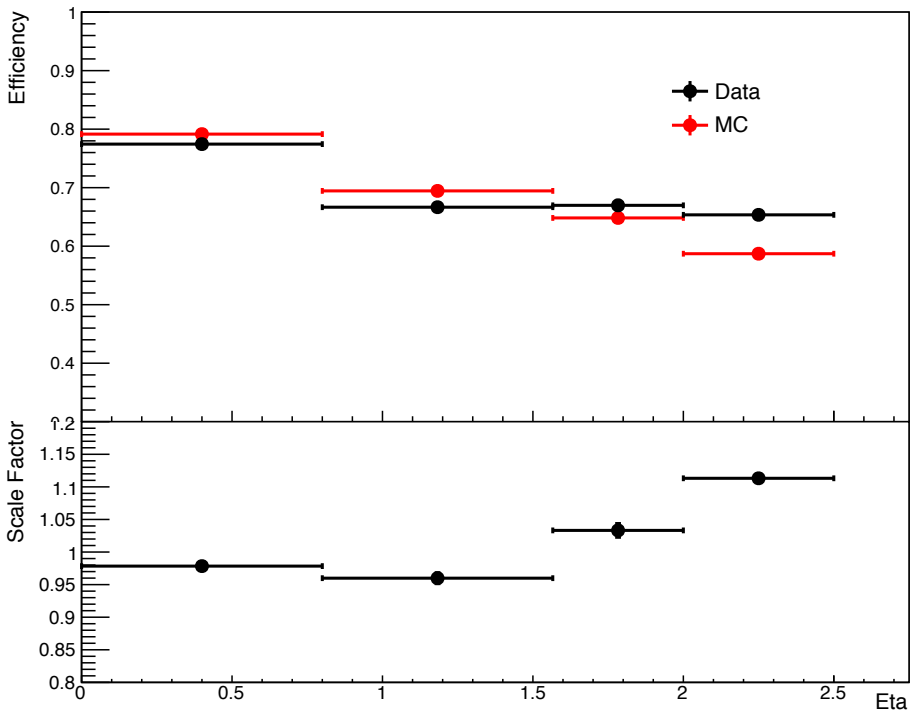

Figure 11b. This graph shows the efficiency of the tight ID/isolation electron filter for the MC simulations and data, and the resulting scale factor as a function of positive values of the angle η. The efficiency tends to be slightly lower for both the data and the MC simulations at high values of η, which correspond to the endcap rather than the barrel of CMS detector. However, the scale factor tends to be higher at the endcap.

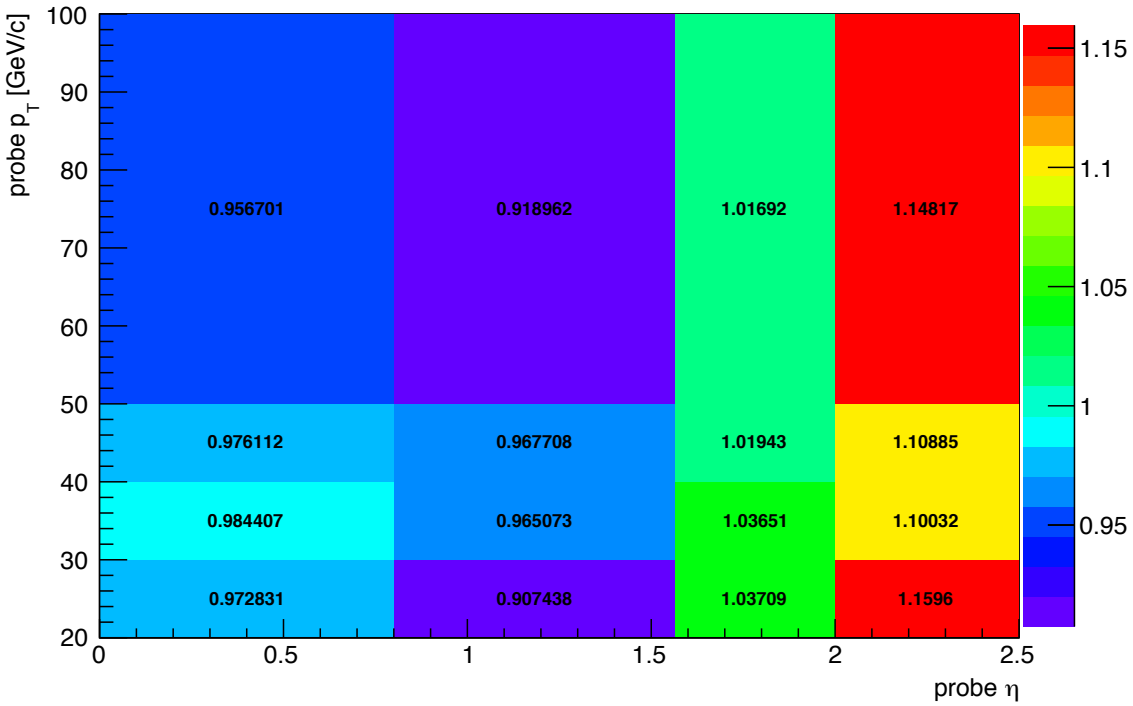

Figure 11c. This 2d graph shows the electron tight ID/isolation scale factors as a function of the transverse momentum and the angle η. The scale factors are somewhat higher at the endcap (high η) than in the barrel (lower η).

References

1. CMS Collaboration, *Search for dark matter direct production using razor variables in events with two or more jets in pp collisions at 8 TeV* (2015), CMS-PAS-EXO-14-004

2. Original image retrieved from https://inspirehep.net/record/871832/plots

3. Original image retrieved from T. DuPree et. al., "Simplified Models for tt-bar + Dark Matter". LHC DM Forum.

4. B. Andersson *et al.*, "Parton Fragmentation and String Dynamics", *Phys. Rep.* 97(2–3), 31–145 (1983).

5. Found on CERN website: http://cms.web.cern.ch/news/triggering-and-data-acquisition

6. CMS NOTE AN-10-464, "Electron and muon efficiency measurements in 2010 proton-proton dataset."

7. CMS NOTE AN-11-097, "Lepton efficiencies for the inclusive *W* cross section measurement with 36.1 pb^{-1}."

8. CMS NOTE AN-12-311, "Electron Efficiency Measurement Using the Radiative Z → $e^+e^-\gamma$ Sample."

Acknowledgments

I would like to acknowledge Professors Harvey Newman and Maria Spiropulu for offering me this project and for providing resources and information on particle physics and the project, and for having meetings and talks, which allowed me to share my research, receive feedback, and hear about others' projects.

I would also like to acknowledge Javier Duarte and Cristian Pena, who co-mentored me throughout the program, taught me the necessary skills and concepts for the research, and looked over and helped me improve my work.

Also, I would like to acknowledge Si Xie, who helped me with the trigger turn-on curves, and gave me feedback throughout the project.

Finally, I would like to acknowledge Dorian Kcira, who provided necessary technical support and access to the Caltech server on which the data is stored.