Improving Electron Identification in the Endcap of the CMS Electromagnetic Calorimeter

Christopher J Mintram

May 2013

Abstract

One of the most powerful ways the CMS experiment identifies electron candidates is by examining the spatial distribution of energy deposits in the crystals that make up the electromagnetic calorimeter. Currently the shower shape variable $\sigma_{i\eta i\eta}$ used in the forward η region does not separate jets and electrons as well as would be liked. An improved variable is sought to help identify electrons in the endcap electromagnetic calorimeter, so that the data this subdetector records may be of even more use to the collaboration. Various new variables have been tried with the intent of improving on the separation between electron and jet events. It has been concluded that the current variable is likely the most suitable for use without the additional power that can be provided by the preshower. Promise is shown in using the preshower detector in front of the electromagnetic calorimeter, with an isolation variable and a $\sigma_{i\eta i\eta}$ inspired preshower shower shape variable being used in addition to the endcap shower shape variable, improving the background rejection when used in conjunction with the current endcap cuts. The isolation variable can reject 21 % of jet background passing the standard selection criteria while leaving 99 % of the signal untouched.

Contents

1 Introduction

Electrons[1] and photons with high transverse energies are rare in hadron colliders, but are key to many interesting physics processes that are studied including Higgs→ $\gamma\gamma$ and Z'→ee decays. They are however difficult to distinguish from the large proportion of hadronic jets which can fake similar properties when detected, since they are also calorimeter based objects. This results in a large jet background that interferes with the electron/photon analyses and must be reduced. This report will look at how the CMS experiment is used in the detection and study of these areas of physics, particularly focussing on the identification of electrons and the reduction of this hadronic jet background. Z'→ee analysis, which searches for a new resonance in high mass electron pairs, will be used as a benchmark analysis[4].

As of February 2013, the LHC has entered its first long shutdown period, and in this time the CMS collaboration intend to re-evaluate the selection criteria used in various analyses. Among these is the selection criteria used for electrons in the electromagnetic calorimeter (ECAL). Although the central barrel region of the ECAL offers strong separation between electron signal and jet background, with more forward physics expected to occur from 2015 onwards, the selection criteria in the endcap will need to be improved. This report will consider possible candidates for replacing one of the key variables currently used in the endcap ($\sigma_{i\eta i\eta}$), while also looking at any ways of exploiting the high resolution preshower detectors located in front of each endcap to improve the electron selection criteria further, as during this first data-taking period this has been underutilised.

2 The CMS Experiment

The Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) [5] is a large general purpose detector, designed with the intention of searching for the Higgs boson using the predicted Higgs → $\gamma\gamma$ decay channel [10], but also with the hope of finding any other new physics at the TeV energy scale, such as Z' bosons [8] which are believed can decay into electron-positron pairs. Figure 1 shows a general schematic of the CMS detector including the muon chambers, calorimeters and the tracker.

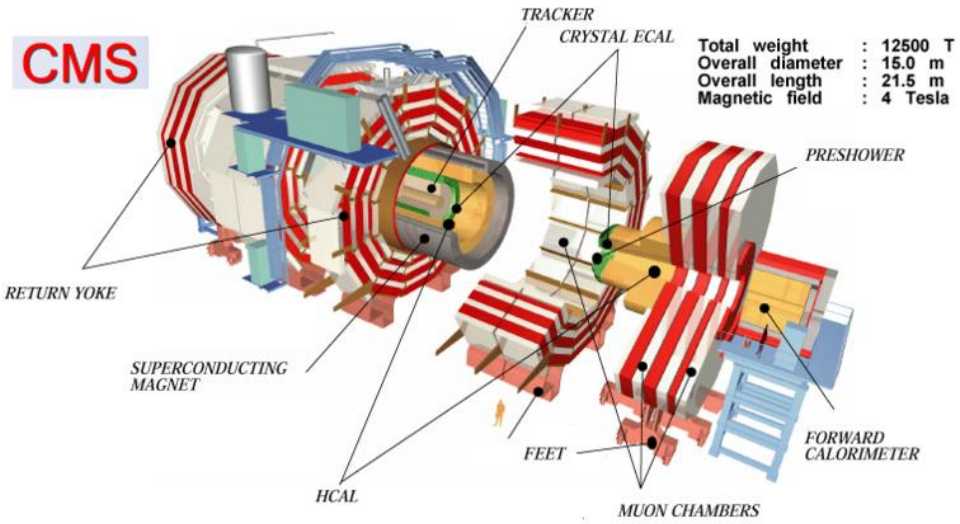

Figure 1: Schematic of the CMS detector.

[1]This report does not distinguish between electrons and positrons, as they have the same properties as far as the calorimeter and preshower detectors are concerned

The detector is built around a solenoid magnet in the form of a cylindrical coil of superconducting cable which produced a magnetic field of approximately 3.8 T (around 100,000 times the 0.4 G field felt due to the Earth's core at its surface) when at its peak operating temperature of 4.5 K. Each part of the detector has a specific function. The tracker, made entirely of silicon, is used to measure the momentum of charged tracks. It is designed to interfere with the particles producing the tracks as little as possible, so that they lose as little energy as possible. The calorimeters are designed to stop the type of particles they have been designed to detect (electrons and photons for the Electromagnetic Calorimeter and hadrons for the Hadron Calorimeter). The outer portion of the detector detects muons, and also helps contain the magnetic field.

CMS uses a right-handed co-ordinate system, with the origin being placed at the nominal collision point. The x-axis points towards the centre of the LHC ring, the y axis points vertically and the z axis points in the anti-clockwise beam direction. In this report $\eta - \phi$ co-ordinates are often used, where ϕ is the radial position around the barrel and η is the pseudorapidity. Pseudorapidity describes the angle of a detected particle relative to the direction of the beam, and is logarithmically related to the tangent of the physical angle θ that is used to describe the distance along the length of the detector.

$$\eta = -ln[\tan(\tfrac{\theta}{2})]$$

2.1 The Electromagnetic Calorimeter

The main component of CMS used to identify and reconstruct electrons and photons is the Electromagnetic Calorimeter (ECAL) [2]. This uses lead tungstate crystals that scintillate when electrons or photons pass through them. Each crystal is attached to a photodetector, which converts the light produced in the crystal by particles interacting with them, to an electrical signal that can be analysed, allowing the energies of the detected particles to be determined [9]. The crystals are made of lead tungstate as they are required to have high density and produce scintillation light in fast, small, well-defined photon showers, to allow for a more precise and compact calorimeter.

In the crystals of the ECAL, the electrons can interact with matter, producing photons via Bremsstrahlung [6]. These photons can then interact and convert into an electron-positron pair. These can then interact producing more photons via Bremsstrahlung, and this chain can continue until the energy is detected by the photodetector at the other end of the crystal. This is known as an electromagnetic shower. The width of the shower is determined by the Moliére radius of the material, and the length is determined by the radiation length of the material. The Molière radius is the radius of a cylinder that is expected to contain 90 % of the energy deposited by an electromagnetic shower. The radiation length describes the distance it takes for an electron to lose all but 1/e of it's energy. Therefore the electrons and photons are showering through the crystal and by the time they reach the other end of the crystal, the energy the initial electron was carrying has been spread over an area described by the Molière radius.

The dimensions of each crystal at the front (inner) face are approximately $2.2\,\mathrm{x}\,2.2\,\mathrm{cm}^2$ in the barrel and $2.6\,\mathrm{x}\,2.6\,\mathrm{cm}^2$ in the endcap, increasing to $3\,\mathrm{x}\,3\,\mathrm{cm}^2$ for the rear (outer) face. The length of the crystals is 23 cm and 22 cm for the barrel and endcap subdetectors respectively. The crystal width is a little over the Molière radius of lead tungstate (2.19 cm) so some of the energy will have spilt over into adjacent crystals, and depending on where the initial electron first interacted with the crystal, most of the energy would normally be contained in the central one or two crystals. The length is corresponds to 25 radiation lengths of lead tungstate (0.89 g cm^{-2}) which means the electrons energy will be fully contained in the crystals.

The electrons and photons only interact with charged matter, so the ECAL uses lead, a material that maximises the number of protons that the electrons and photons can interact with. Jets on the other hand, interact with other hadronic objects, so the HCAL (Hadron Calorimeter) uses

scintillators made from brass, a material that maximises the total number of nucleons the jets can interact with. This is a slower process as the interaction length over which hadrons lose energy in the material is much longer than the radiation length of electromagnetic particles.

The ECAL is split into two main sub-detectors: the barrel, spanning $|\eta| < 1.479$; and the endcaps, which span the η regions of $1.479 < |\eta| < 3.0$. η is almost linear to the position along the detector up to about $\eta = \pm 1.5$ where the linearity breaks down due to the logarithmic relationship between θ and η. Beyond $\eta = \pm 1.5$ are the endcap regions, where the linearity of this relationship starts to breakdown. The ECAL has the crystals in its barrel subdetector arranged in an $\eta - \phi$ geometry. The endcap subdetector is arranged in an x-y geometry, and although it is still possible to parametrise positions within the endcap in terms of η and ϕ, this is less effective than in the barrel. Figure 2 shows the endcaps and barrel of the ECAL along with the preshower detectors.

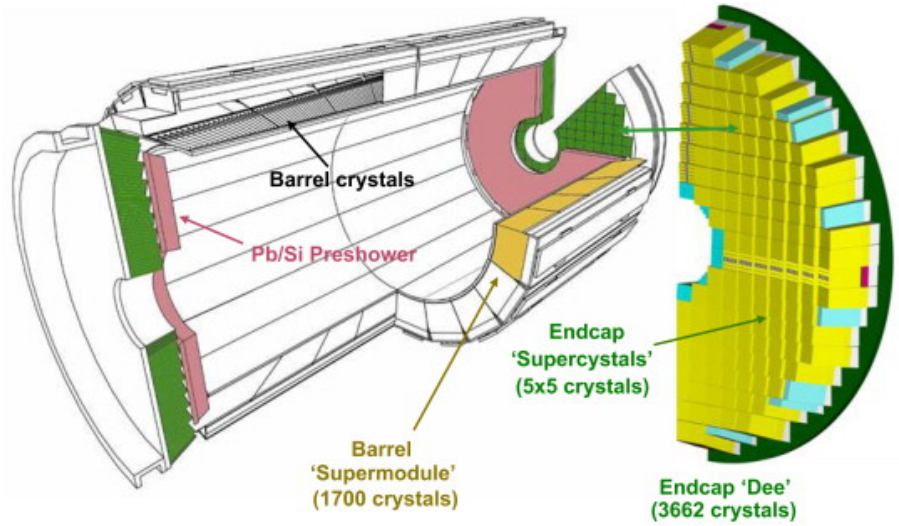

Figure 2: A representation of the ECAL in CMS.

2.2 Preshower Detectors

The ECAL also contains high resolution preshower detectors that sit in front of each endcap in the η regions of $1.653 < |\eta| < 2.6$ [7]. Instead of crystals, the preshower is made up of two planes of lead followed by silicon sensors. When a photon passes through the lead layer, it creates an electromagnetic shower of electron-positron pairs that the silicon sensors then detect and measure.

Whereas the ECAL crystals are 2.2 cm wide, the silicon strips in the preshower detector are only 1.9 mm wide making it easier to distinguish between two or more separate photons which may be hitting the same crystal in the detector. This is particularly useful for detecting the short-lived neutral pion particles that can be produced in collisions and then decay into two closely-spaced lower energy photons that may hit the same crystals in the detector which mimic high-energy electrons and could be mistaken for those from the Z'→ee decay channel.

In the preshower, each detector in the preshower has an overall area of $63 \times 63\,\mathrm{mm}^2$ and is made of 32 silicon strips. Each strip measures $1.9 \times 61\,\mathrm{mm}^2$, giving each individual sensor a working area of $61 \times 61\,\mathrm{mm}^2$. These are oriented so that the strips run parallel to the x axis in the first plane, and parallel to the y axis is the second. This gives strong x resolution when compared to its y resolution in the first plane, and strong y resolution when compared to its x resolution in the second.

3 Electron Identification

Electron candidates are constructed out of deposits of energy in the ECAL in groupings of crystals called superclusters and associated tracks. Electrons are normally identified by looking at these and applying various cuts that are used to distinguish them from jets. In Z'→ee analysis, cuts on the shape of the electromagnetic shower (shower shape), the track of the particle, H/E (hadronic energy over electronic energy) and isolation(which looks at what is going on around the event rather than the event itself) [1]. In the endcap, the shower shape variable is currently less well developed than other cuts, and this project focuses on improving it.

Representations of electron event superclusters in the barrel and endcap of the ECAL are shown in Figure 3. The barrel electron's position is described in $\eta - \phi$ co-ordinates as explained earlier, while the endcap electron's location is described by the x-y geometry used by CMS.

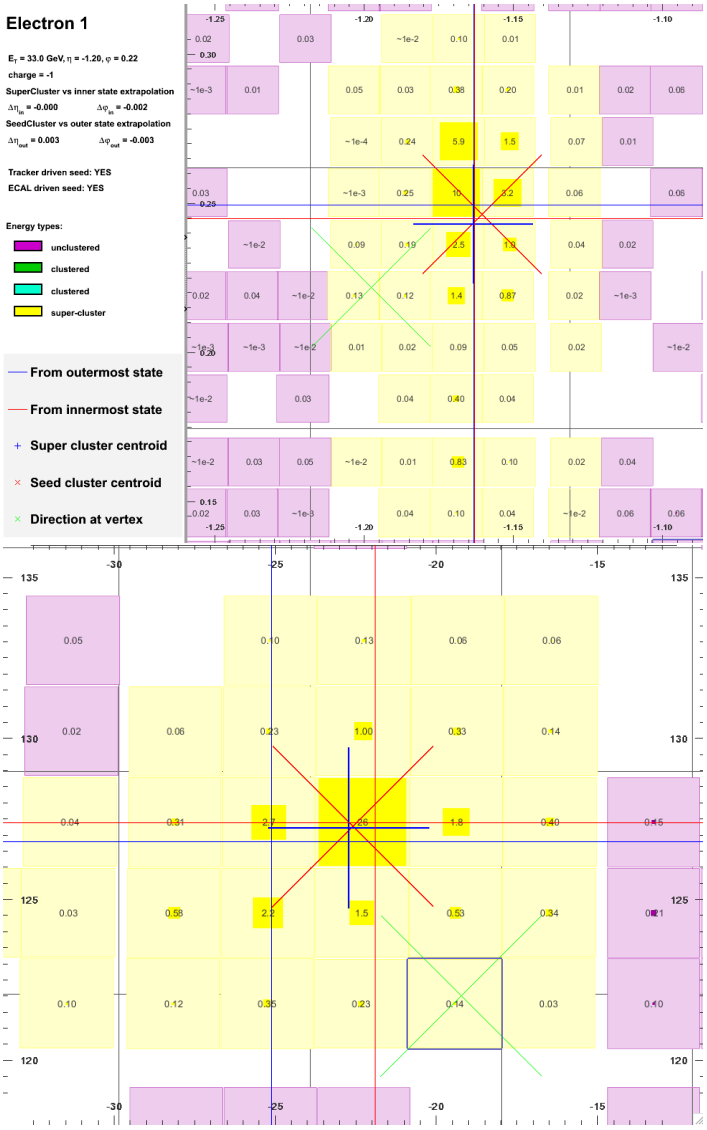

Figure 3: Close-up views of a barrel(above) and endcap(below) electron from a Z→ee candidate. Each box represents a single crystal and the numbers in the boxes represent the transverse energy recorded in the crystal.

The electrons are bent by the magnetic field in the detector. This causes them to bend in ϕ. As the electrons pass through the tracker, they can interact with the electromagnetic fields of its atoms in the same way as they do in the crystals of the ECAL releasing photons via the Bremsstrahlung process. These photons are neutral and so follow the initial trajectory of the electron as they will not be bent by the field. The photons still interact with the atoms in the tracker, and when they do, they pair convert to an electron-positron pair. These can then also undergo Bremsstrahlung if they interact with the matter in the tracker. This can result in the original electron becoming a spread of photons and electrons in ϕ although it does not occur for all electrons. This has hampered the used of shower shape variables in the endcap, as any shower shape cut must allow for additional spread in ϕ due to this premature electromagnetic shower. The x-y geometry of the endcap can't do this directly, which is one of the reasons it struggles to replicate the usefulness of the barrel.

Shower shape variables consider the spread of energy from the original point of impact (the seed crystal) to the surrounding crystals. Electromagnetic objects typically spread in a distinct shape described by the Molière radius, unlike jets which have more diffuse energy spreads. Because the electrons are bending, the shower shape is not a clean spread equally in η and ϕ. The direction of the magnetic field results in an extended spread in ϕ with little change in η. In the barrel of the ECAL, this is easier to spot and easier to cut against than in the endcaps mainly because of the x-y endcap geometry making the spread in ϕ more difficult to identify.

Another important variable type for electron identification is isolation variables. Unlike shower shape variables, that look at the electron candidate directly, isolation variables try to determine whether there are any other hits close to the candidate. If there are, this implies that there are likely to be other particles showering with the candidate, suggesting that it is part of a jet, rather than a solitary electron. This can be a powerful method for removing the jet background.

Unfortunately, data acquired from the endcap has been exploited far less than data from the barrel, and has therefore been less prominent in the search for data for the Higgs$\rightarrow \gamma\gamma$ and Z'$\rightarrow$ee decay channels. This is partly down to the methods currently used exploiting the $\eta - \phi$ geometry in the barrel and are less effective in the x-y geometry of the endcap. The difference in geometries makes it hard to directly use the barrel variables in the endcap as the x-y geometry cannot be used as easily to map the $\eta - \phi$ spread as in the barrel, making using them generally less efficient as they are in the barrel. Figure 4 shows the fraction of preselected jets passing the high energy electron ID (HEEP) [3]). This shows that the fraction of jets passing the HEEP ID can be up to 10 times higher in the high η region of the endcap than in the barrel. Therefore additional rejection is needed and current ID variables are not sufficient to achieve this. This project aims to rectify this by introducing new ID variables both calorimeter and preshower based.

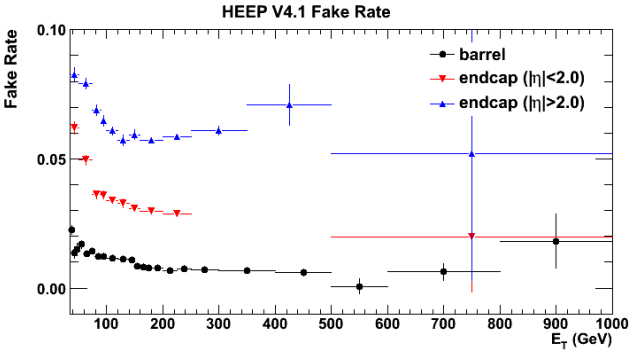

Figure 4: Fake rates for electron detection in the barrel and endcap regions(above). Event distribution for endcap-endcap dielectron production channel(below).

The preshower detector's job is to aid in electron identification in the endcap of the ECAL. It does this by providing high x-y resolution which helps distinguish between hit sources. As previously stated, the preshower is made of silicon strips 1.9 mm wide, rather than the 2.2 cm and 2.6 cm wide crystals used in the barrel and endcap respectively, making it good at distinguishing between closely spaced particles. However as they are found in strips which are longer than the width of the crystals, they can only have this strong resolution in one direction, resulting in the need for a second, orthogonally rotated preshower plane in order to improve the chances of recognising multiple particles in close proximity.

4 Method

The initial aim of this project was to improve electron identification in the endcap by either finding a replacement shower shape variable, or to improve on the current variable used. In order to achieve this, the separation between the electron signal and mainly jet background was studied. The main factor that was considered was how efficient the variable was for accepting the signal, and how well it rejected the background.

Two data sets were studied. One dominated by electron signal, and one by jet background. The signal sample was composed of Z→ee events, all of which had one electron located in the barrel (the "tag" electron), and one located in the endcap (the "probe" electron). The events were selected by loose double electron triggers with the full set of data included. The tag electron was required to pass the full "HEEP" selection used by CMS to select high energy electrons [3].

The background sample was mainly composed of jet events. They had one "probe" electron located in the endcap and no corresponding "tag" electron. This rejected Z events, however there was still some contamination (of the order of 20 %) of real electrons, mainly from W boson decays. This used a loose single γ trigger and only included a proportion of the data. This is because the jet background is substantially larger than the signal. When operational, the LHC collides 20 million bunches of particles per second. Collecting all the events from this would require far more computing power than available, so triggers are used to ensure that only around 300 or so "interesting" events are recorded and stored per second. These loose single γ triggers would record several thousand background events alone, and as these are only required for background studies, only a small fraction of events that pass them are saved for data analysis(roughly corresponding to less than 1 event per second).

For these two samples, the efficiency of the signal sample will be plotted against the rejection of the background sample, and where deemed important, other parameters will be compared. In this report, the methods used to create potential new variables in both the endcap of the ECAL and in the preshower detectors will be discussed and any approaches that should be followed up in the future will be outlined. Depending on the needs of the particular analysis, the events may be required to pass the full existing selection cuts first. If this is the case it will be specified.

Once a suitable variable was found it was tested against a signal sample of events derived from a simulation that used the Monte Carlo method to ensure that it was correctly modelled.

5 Calorimeter Based Variables

5.1 SigmaiEtaiEta ($\sigma_{i\eta i\eta}$)

The current shower shape variable used in the endcap is sigmaiEtaiEta ($\sigma_{i\eta i\eta}$). It measures the root mean square of the position of energy deposited by the electrons in units of crystals over a 5x5 block of crystals centred on the seed crystal.

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\Sigma_i(\eta_i - \bar{\eta})^2\omega_i^2}{\Sigma_i\omega_i^2}} \text{ where } \omega_i^2 = 4.7 + ln(\frac{E_i}{E_{5x5}}) \text{ for the endcap of the ECAL}$$

Each 'i' identifies an individual crystal in the 5x5 block surrounding the seed crystal. Only the 5x5 block is necessary as it will usually contain 99 % of the impacting electron's energy. η is the location of each crystal in the detector. This was also used in the barrel, but a combination of two better shower shape variables are now used there (e1x5Over5x5 and e2x5Over5x5 which take the sum of energy in the 1x5 or 2x5 strips of crystals around the seed crystal as a ratio of the energy of the 5x5 block). The aim is to come up with a similar new variable to be used in the endcap. The current cut used, $\sigma_{i\eta i\eta} < 0.03$ represents that the shower of electromagnetic objects usually spreads half a crystal from the seed crystal. For historical reasons, this is measured in units of crystals, but then scaled by 0.0447. Using the current accepted cut value 65 % of background events are still picked up, and although 93 % of actual electrons are accepted. It would be preferable to reduce the proportion of background events being accepted by the cut. Figure 5 shows the distribution of $\sigma_{i\eta i\eta}$ for both the signal and background enriched samples, and indicates the cut point of $\sigma_{i\eta i\eta} < 0.03$. Figure 6 shows how much of the signal is accepted compared to the background rejected at different cut points when it is the only cut being applied (cuts include all values below the cut point).

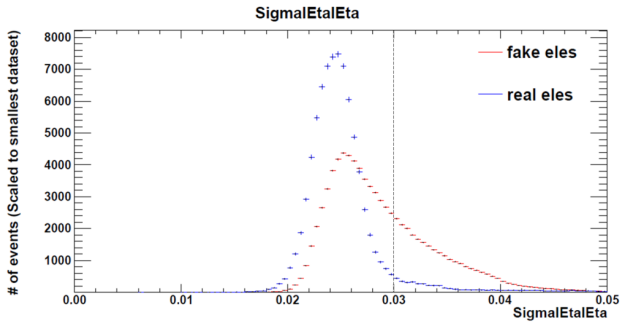

Figure 5: $\sigma_{i\eta i\eta}$ distribution for endcap events in the signal and background enriched samples.

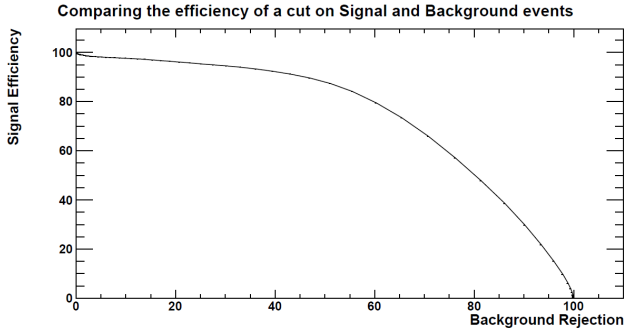

Figure 6: A comparison of the percentage of the signal sample passing and the percentage of the background sample being rejected by $\sigma_{i\eta i\eta}$ when no other cuts have been applied.

The e1x5Over5x5 and e2x5Over5x5 variables used in the barrel are shown in Figure 7. 90 % of the energy should be contained in a cylinder with a radius of one Molière radius. The additional spread in ϕ, due to the electrons undergoing Bremsstrahlung, causes most of the energy to be in the central strips in η, rather than the central crystal, which would be expected if the electron was not undergoing Bremsstrahlung. The two variables are designed to work in tandem as there are two possible cases of where the spread occurs. Either the electron is detected in the centre of the seed crystal, in which case 80 % of its energy should be located in the 1x5 strip. Otherwise the electron has hit near the edge of a crystal, and is spread between the two strips, which will contain 95 % of the energy.

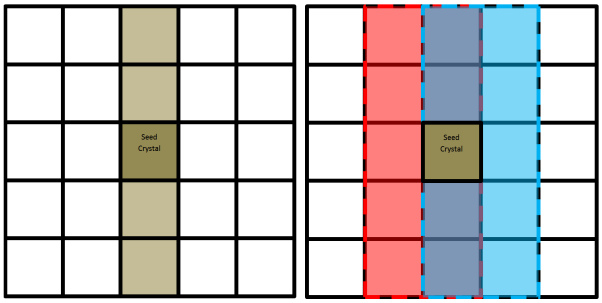

Figure 7: e1x5Over5x5(left) and e2x5Over5x5(right). The energy of one set of shaded cells is summed and divided by the total energy in the 5x5 group of crystals. In e2x5Over5x5 the larger of the two sums of energies is used.

Using these variables in the geometry of the endcap in place of $\sigma_{i\eta i\eta}$ requires cut points of > 0.44 and > 0.73 respectively to get the same 93 % efficiency as $\sigma_{i\eta i\eta}$ when cutting on electron events, however 92.2 % (for e1x5Over5x5) and 92.6 % (for e2x5Over5x5) of the background events also pass the cut. Clearly neither of these are suitable solutions and alternate variables were tried with the intention that one of the variables tried would prove to be more effective than $\sigma_{i\eta i\eta}$.

5.2 New Variables

The first new approach attempted in the endcap was to choose and count the energy for particular crystals in the 5x5 square centred on the seed crystal to attempt to collect a large proportion of the energy from real events, but a significantly smaller amount for fake events (in a similar manner to the e1x5Over5x5 and e2x5Over5x5 variables). It was not difficult to get either strong signal efficiency or powerful background rejection, however getting a variable that achieved both was never achieved with this method. The shapes tried using this method were chosen by visually scanning many individual events in both samples, looking for patterns in energy spreads and trying to find shapes that would work better when applied to the signal sample than the background.

A strong signal efficiency with as tight a cut as possible was best achieved by taking the largest amount of energy in a 4x4 square within the 5x5 block of crystals surrounding the seed crystal as a ratio of the 5x5 block(the four possible squares this could be are shown in Figure 8 and the distribution it creates is shown in Figure 9). A tight cut achieved a very strong efficiency for real events with 93 % of real events passing a cut of >0.95 (given the maximum value of this should be 1) when no other cuts were applied. However it worked poorly on fake events with 89 % passing the same cut. Other cut points were tried but at no point could a cut achieve better signal efficiency and better background rejection than $\sigma_{i\eta i\eta}$ (as shown in Figure 10). None of the attempted collections using this method showed enough distinction between the real events and the background enriched sample.

In a similar vein, other shapes such as a diamond (Figure 11) and a cross yielded similar results with not enough distinctions being drawn between real and fake Z' $\rightarrow$ ee events(its distribution is shown in Figure 12). For the diamond shape, to get the 93 % efficiency level that $\sigma_{i\eta i\eta}$ manages, the cut point was >0.83, however 95 % of the background sample was accepted. Figure 12 shows the distribution the diamond shape produces.

The energy scale's zero point is not at 0 eV, but is at the level of the background noise picked up when there is no activity in the ECAL (approximately 40 MeV, although this fluctuates). As such, it is possible for the energies to be recorded as negative. If these are considered then it is possible for a variable such as e4x4Over5x5 to be higher than one, which doesn't appear to

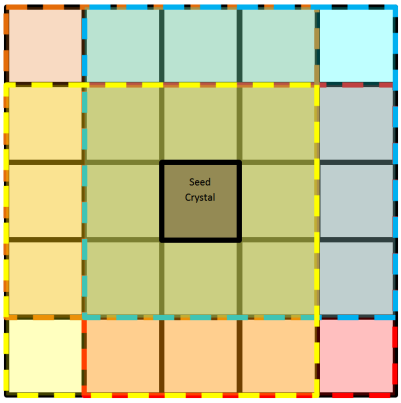

Figure 8: The possible collections of energy used in the e4x4Over5x5 variable.

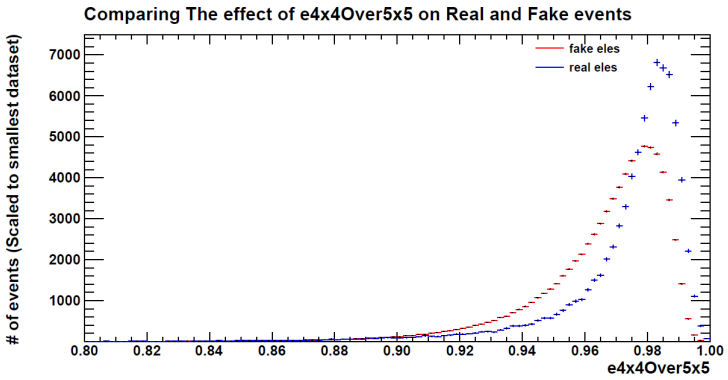

Figure 9: e4x4Over5x5 distribution for endcap events in the signal and background enriched samples.

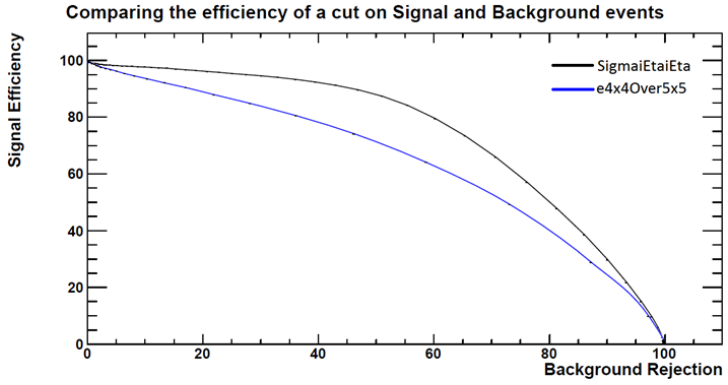

Figure 10: Comparing the Signal Efficiencies of $\sigma_{i\eta i\eta}$ and e4x4Over5x5 against background rejection when no other cuts have been applied.

make sense. To eliminate this, any crystals with negative recorded energies were ignored. As it is expected that at least some of the low energy deposits are background, adding a minimum energy threshold, such as 1 % of the energy of the seed crystal, was tried. This provided no additional discrimination, so has not been pursued further.

Increasing the amount of crystals to the 7x7 block centred on the seed crystal was attempted to

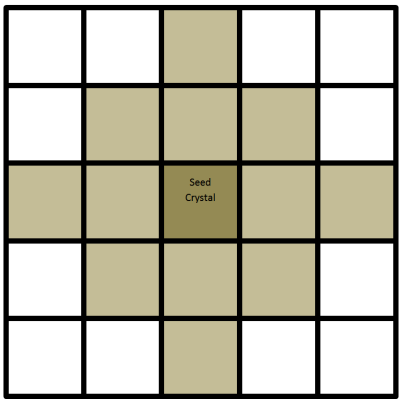

Figure 11: The 5x5 diamond shape attempted.

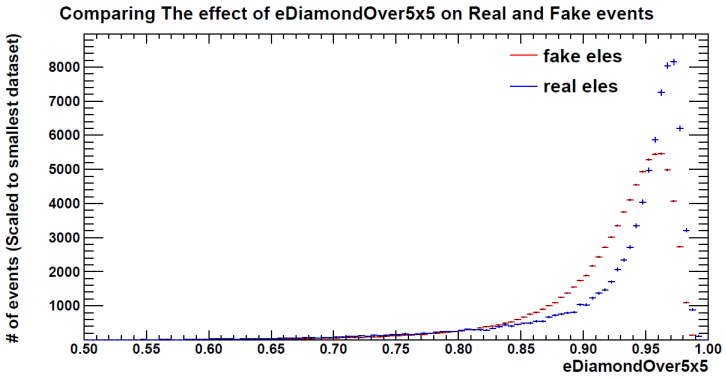

Figure 12: The 5x5 diamond shape's distribution for the signal and background samples.

see if any additional distinctions could be made from the larger spread. Outside of the 5x5 block surrounding the seed crystal, the data is "zero suppressed". In order to reduce bandwidth, noise is not read out unless it is significantly higher than the rest of the background. This should not affect the analysis, but is worth noting. Variables based off the proportion of energy in a shape in a ratio with the total in the 7x7 block surrounding the seed included a 5x5 and a diamond shape. These still did not give enough distinction between Z'→ee electron candidates and background events.

The next approach was to try shapes that only incorporated proportions of the energies in some crystals. The idea behind this was that it would give a better representation of how the energy is spreading. The chosen shapes for this were circles or various radii. Circles with a radius of 2 crystals to 4.5 crystals were created and tested by taking a ratio of the different sized circles in the same way as for the energy collections within the 5x5 and 7x7 block surrounding the seed crystal. The energy of each crystal that was taken was proportional to the amount of the crystal that was inside the circle's circumference. This estimate of the energy wasn't necessarily precise as it assumed that the energy in the crystal is picked up evenly over the surface of the crystal. The use of circles worked on the assumption that in the endcap, the energy spreads equally in every direction. This is known not to be true as the electrons are spreading in ϕ due to Bremsstrahlung and a more accurate description of the spread would be an oval oriented with relation to the η and ϕ co-ordinates. Although this gave less overlap than some of the inclusions of whole energies, there was still too little distinction between the real and background events. For example, the circle of radius 2 divided by the circle with a radius of 2.5 was cut for >0.87 to accept 93 % of the signal but also accepted 95 % of the background enriched sample.

Although these methods haven't yielded the results hoped, it should not be regarded as a failure. Thanks to these attempts, the collaboration now knows that these methods are very unlikely to work for the endcap, even if a shape that has been missed by this project is thought up. It will save time for researchers in the future who now have a basis to work from should they follow up these attempts.

5.3 Changing Geometries

It was thought that attempting to change the co-ordinate system might yet yield a better variable. Converting from the x-y geometry used in the endcap to the $\eta - \phi$ geometry used in the barrel may allow for a variable that is independent of its position in the endcap allowing for a universal variable in the endcap rather than a set of different co-ordinate dependent variables, or even a variable that works well in both the endcap and the barrel. By converting the geometry to the system used in the barrel, it could be possible to use the e2x5Over5x5 and e1x5Over5x5 variables used for the shower shape in the barrel. This may give fairly crude results depending on assumptions made, particularly if the use of the assumption that the energy in the crystal is spread evenly over its surface continues to be made.

To test whether this idea was worth pursuing, a trial using eight distinct sections oriented at separations of 45^o in phi around the detector (rather than a continuous rotation around the 360^o of the detector), each with a shower shape variable that is a 45^o rotation of the previous section, has been attempted and shows promise, giving better distinction between real and fake events than many of the standard shapes. A rough idea of the shapes involved is shown in Figure 13.

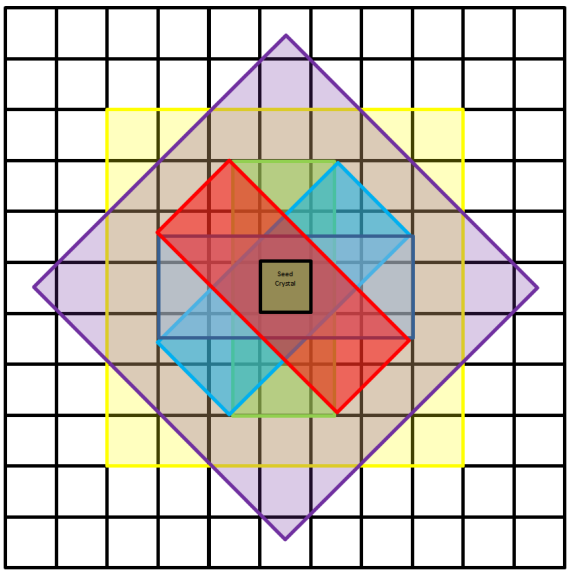

Figure 13: The possible shower shapes taken into account by "e2x5Rotated".

The current version of this accepts 93 % of the actual signal, but 91 % of background events above 0.62. Although this performed better than some of the earlier attempts in the x-y geometry, it was still nowhere near as effective as $\sigma_{i\eta i\eta}$. The cut point was varied to see if perhaps cutting for a lower signal efficiency would result in it exceeding $\sigma_{i\eta i\eta}$, but at no point does it cut more background and accept more signal than $\sigma_{i\eta i\eta}$ (as shown in Figure 15). At this point it was decided that the mathematical and technical work required for computing the full 360^o continuous rotation, would be unlikely to gain enough for the amount of work required.

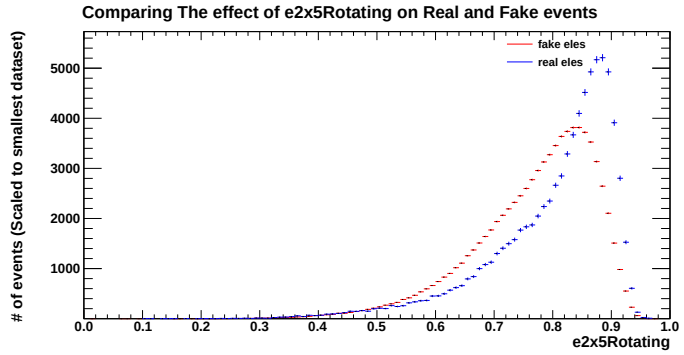

Figure 14: The "e2x5Rotated" Distributions.

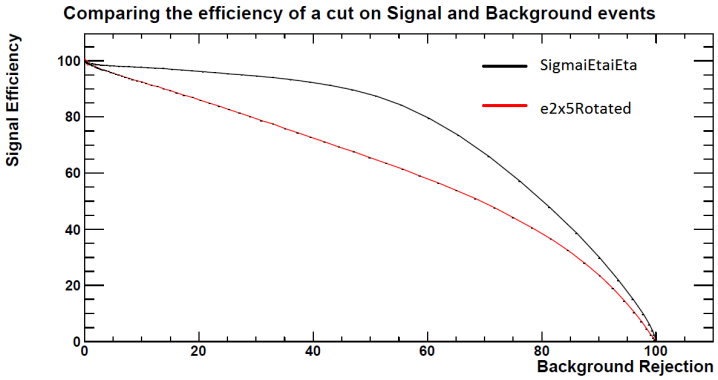

Figure 15: Comparing the change in signal efficiency for e2x5Rotated and $\sigma_{i\eta i\eta}$ when no other cuts are in effect.

5.4 Improving on $\sigma_{i\eta i\eta}$

As no new shower shape variable was deemed promising, it was decided to return to $\sigma_{i\eta i\eta}$ and see if there were any ways that it could be improved as a variable. The first thing considered was whether the current cut point was giving the best possible compromise between signal efficiency and background rejection. The values of signal efficiency and background rejection for the two data samples were compared for values either side of the current cut of $\sigma_{i\eta i\eta}$ <0.03. It was found that using a cut of $\sigma_{i\eta i\eta}$ <0.0285, 3 % signal could be sacrificed to reject an additional 10 % of the background-enriched sample. This was originally considered as a viable change, however further study showed that it did not work as well with the new variables created that utilised the preshower as the current cut point does, and was therefore abandoned.

An attempt was made to improve of $\sigma_{i\eta i\eta}$ by changing the area over which the spread was considered. Areas of 7x7, 9x9, 11x11 and an alternate calculation of the 5x5 were tried. The 5x5 version was created in order to confirm the code worked. If it did then it should perform near identically to $\sigma_{i\eta i\eta}$. Of the different sizes, only the 7x7 and 5x5 variables proved to be as powerful as $\sigma_{i\eta i\eta}$ at differentiating between the samples, and even then the 7x7 area $\sigma_{i\eta i\eta}$ was only as powerful for lower signal efficiencies than the current signal efficiency obtained for $\sigma_{i\eta i\eta}$. The 5x5 area performed just as well as $\sigma_{i\eta i\eta}$ but that is to be expected given that they are interchangeable ways of calculating the same thing(just with a different scaling of the final output). The performance of each of these variables is shown in Figure 16. $\sigma_{i\eta i\eta}$ is barely visible because of the performance of the 5x5 matching it almost perfectly.

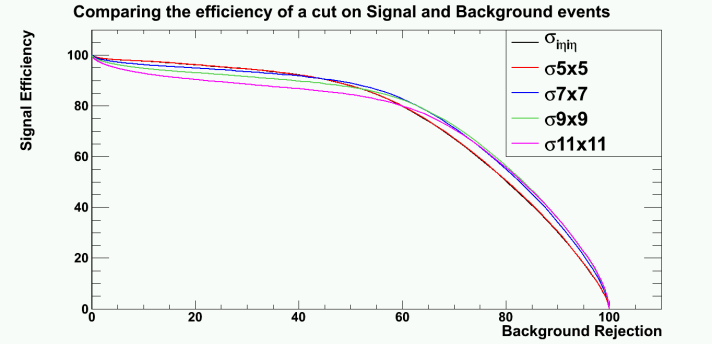

Figure 16: Comparing signal efficiency and background rejection for the "σ" functions acting over different areas when no other cuts are in effect.

There were concerns that the 5x5 region in x-y was not containing the full $\eta - \phi$ regions necessary. By expanding the area and adding η and ϕ limits, an area which may better show the spread was created. Although not a particularly elegant or accurate fix, it did try to provide a variable closer to that which was originally so powerful in the barrel of the ECAL. The best of the variables of this matched was over an area of 4 increments in η and 4 increments in ϕ. It matched the performance of $\sigma_{i\eta i\eta}$ when no other cuts were being used.

6 Preshower Based Variables

As previously stated, the preshower detector has so far been underutilised. Unlike the endcap, data in the preshower is not split up by a matrix of crystals, but by the strips which make up the individual sensors.

6.1 Preshower Shower Shape Variables

Initially, new shower shape variables were created over cones or radius 0.1 and 0.3 in the $\eta - \phi$ geometry with the hope that one would provide a reasonable addition to the full set of endcap selection cuts.

The first attempt at using it was inspired by $\sigma_{i\eta i\eta}$, and used the same idea over a cone of radius 0.1 in η with a weighting function based solely on the energy of the clusters rather than including the logarithmic weighting that is used in calculating $\sigma_{i\eta i\eta}$ for the endcap. This variable will be referred to as σ_{ES}.

$$\sigma_{ES} = \sqrt{\frac{\Sigma_i(\eta_i - \bar{\eta})^2 E_i^2}{\Sigma_i E_i^2}}$$

As a starting point, it performed reasonably well. Its performance when paired with the current cut value and the proposed tighter cut value for $\sigma_{i\eta i\eta}$ are detailed in Table 1.

Table 1: σ_{ES} performance comparison

$\sigma_{i\eta i\eta}$ cut	σ_{ES} cut	Signal Efficiency	Background Rejection
< 0.03	< 0.032	92.5 %	32.5 %
< 0.0285	< 0.03	92.7 %	31.1 %

σ_{ES} clearly performs better when paired with the current cut value used by $\sigma_{i\eta i\eta}$ than the tighter cut point tried. It is more effective to add σ_{ES} to the selection criteria, if a higher rejection

is needed than to tighten the $\sigma_{i\eta i\eta}$ cut. This is shown in Figure 17.

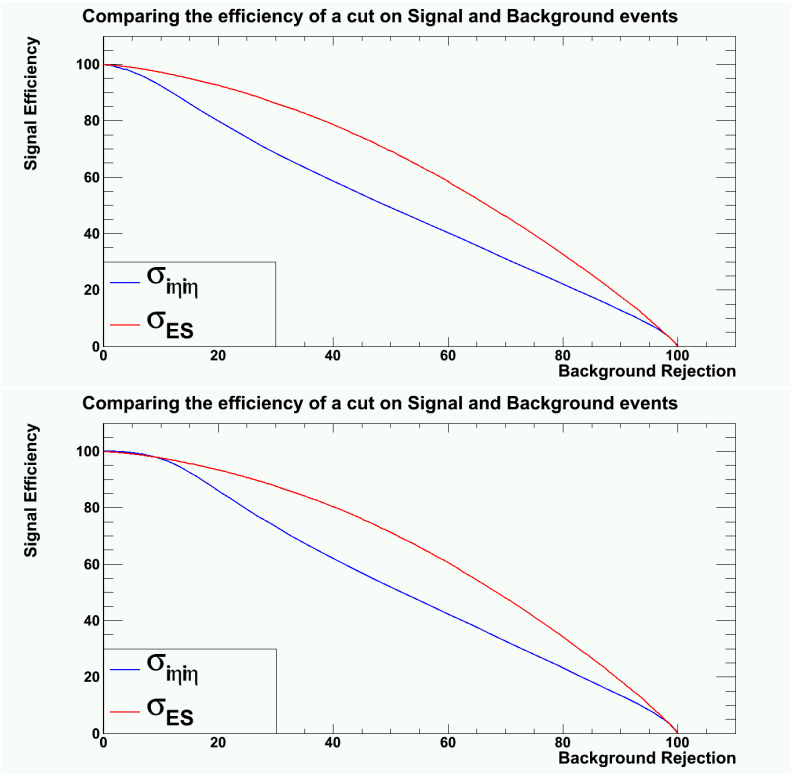

Figure 17: Comparing how much additional cut power is gained by tightening the standard $\sigma_{i\eta i\eta}$ cut point and incorporating the σ_{ES} variable on top of the normal $\sigma_{i\eta i\eta}$ cut(above) and how much power σ_{ES} can provide when used instead of $\sigma_{i\eta i\eta}$ when higher background rejection is preferred to high signal efficiency (below).

Various possible ways on improving this were tried. As the preshower is made up of two planes, the signal efficiency-background rejection curves were compared for each plane alone and for both together. As shown in Figure 18, it was found that using the variable over both planes was most effective. Considering this further, it is obvious as to why. The planes have different orientations, and hence better x or y resolution. Because of this, some events that may be in the $\eta - \phi$ cone are ignored as the hit position recorded is in the centre of the strip rather than the edge of the strip, or included as the centre of the strip is in the cone, but the actual hit is outside it. The number of hits per cluster were also cut on, in the hope that electron events might have more hits per cluster on average than jets. No cut of the number of hits per cluster was the most effective, cutting more background in exchange for the same amount of signal as shown in Figure 19.

Another method used was to consider different sized and shaped collections of energies, similar to those tried in the endcap and barrel. Shapes tried included 1x5 and 2x5 strips as a proportion of a 5x5 area, where a 1x1 square in $\eta - \phi$ being an area of 0.02x0.02. The other variable of this type was to use circles in x-y with different radii. These were found to have little power as the geometry of the detector for such a thing to work is not optimal.

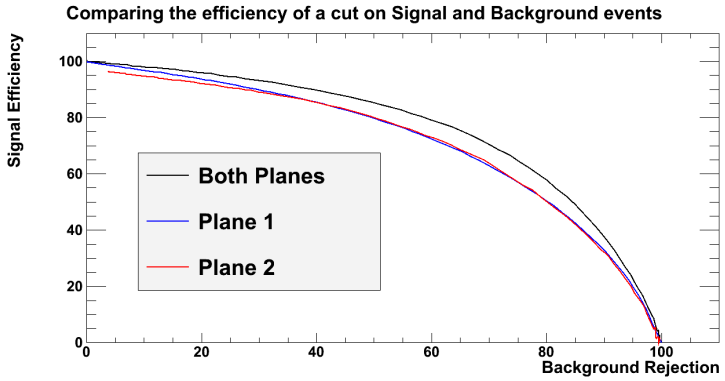

Figure 18: Comparing the signal efficiency - background rejection curves for σ_{ES} when data from each plane was considered separately and then together.

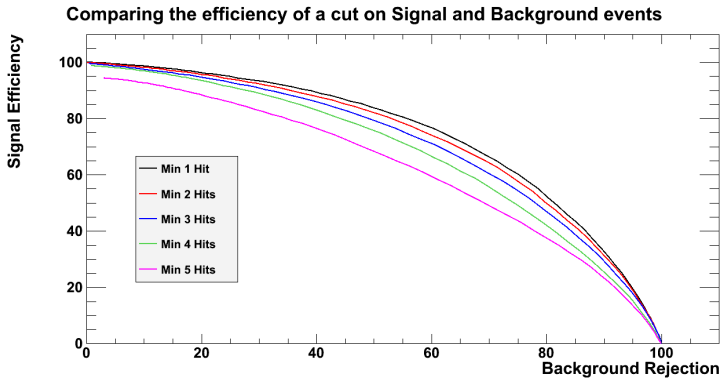

Figure 19: Comparing the signal efficiency - background rejection curves for σ_{ES} with a cut on minimum number of "hits" per cluster, showing that no cut (at least 1 hit per cluster) is the optimum set-up.

6.2 Preshower Isolation Variable

Even greater success has been found with the creation of an isolation variable that would compliment the selection criteria currently in use. The first attempt considered the total hits in a cone of radius 0.1 and 0.3 in $\eta - \phi$. The distribution for the number of hits in the cone of radius 0.3 is shown in Figure 20. There was no clear way to cut a suitable amount of background for an acceptable reduction in signal efficiency.

It was noted that the majority of the events for the signal sample were in the peak of the graph, but due to the large contribution from the tail of the background enriched sample, it could not given a decent separation of the two samples for high signal efficiencies.

An isolation variable should ignore the electron, and focus on any remaining hits. By ignoring a central strip in η for the cone of radius 0.3, which would ignore any hits which had an η value 0.05 higher or lower than that of the centre of the supercluster, the large amounts of events in the peak if the signal were discounted, but the long tail of the background enriched sample remained. The "veto strip" was chosen to be a strip in η rather than ϕ to allow for the electron candidate to spread any amount in ϕ, while limiting its spread in η, that is known to occur because of Bremsstrahlung. This allowed for a cut that gave high efficiency with 99.1 % of signal being kept for a loss of 17 % of the background sample on top of the signal and background cut from the standard endcap selection criteria.

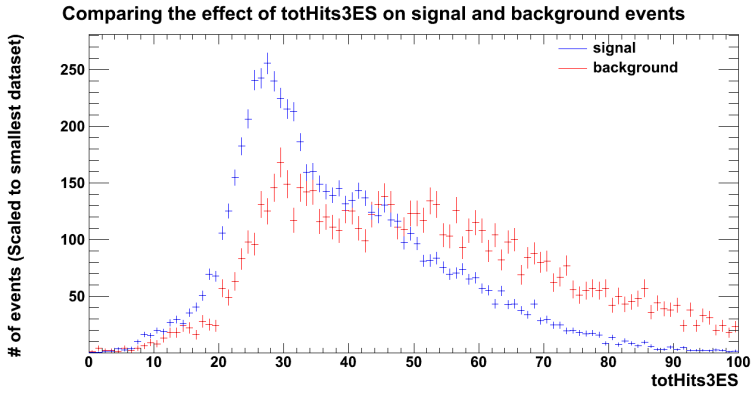

Figure 20: Distribution for the total number of hits in a cone of 0.3 in $\eta - \phi$.

Different sized strips in η were tried to find the best possible variable. Of the six sizes tried, a strip of which vetoed up to 0.02 in η either side of the supercluster's η was determined to be the most powerful as it could give the largest background rejection (rejecting 21 % of the remaining background) for the acceptable limit of 1 % loss in signal after all other cuts as seen in Figure 21. Whether total hits, or total clusters of hits was more suitable was also checked, with total hits being confirmed as giving better results. The distribution for the isolation variable with a strip width of 0.02, "psIsol2ES", is shown in Figure 22.

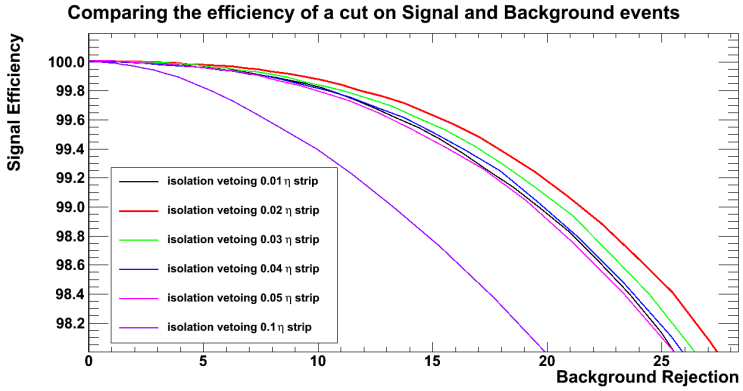

Figure 21: Comparing the signal efficiency - background rejection curves for the different sized isolation strips tried.

The peak of this isolation variable is at 0, showing that in most cases the electron's footprint is being removed. However, there is still evidence of a tail, implying that there is still some of the footprint left in some of the signal events. Further work should try to reduce this even more.

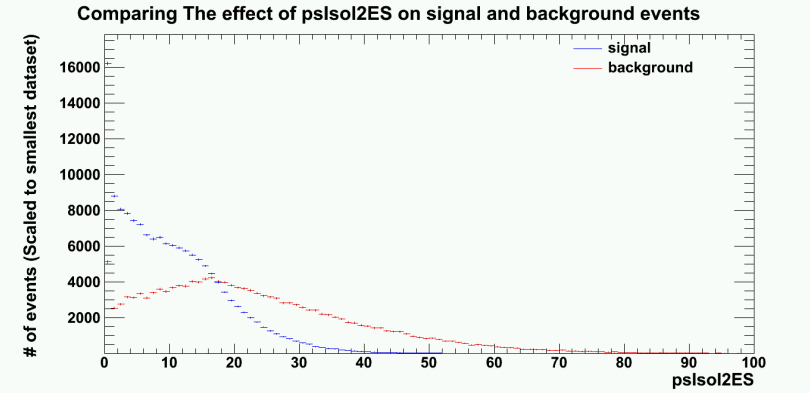

Figure 22: The distribution of the number of hits in 0.3 excluding the hits in the central η strip of ± 0.02.

6.3 Testing the Preshower Isolation Variable

Having created a variable that appears to give strong separation between the signal and background samples, it was important to ensure that the variable was consistent in detector η, detector ϕ, E_t (the calorimeter measured energy x $\sin(\theta_{trk})$ where θ_{trk} is the polar angle of the electrons track measured at the inner layer of the tracker and then extrapolated to the interaction vertex) and the number of interaction vertices. Testing against detector η, the variable was found to be more powerful at cutting events near the centre of the endcaps rather than the edges, with far more events (both signal and background being accepted for lower values of absolute η). This did not mean that it lost an unacceptable amount of signal efficiency for higher absolute η, with a minimum of 96 % of signal still being accepted for $\eta = \pm 2.5$. The proportion of events that have already passed all other selection criteria that are also passing this cut for any given absolute value of η within the bounds of the endcap are shown in Figure 23.

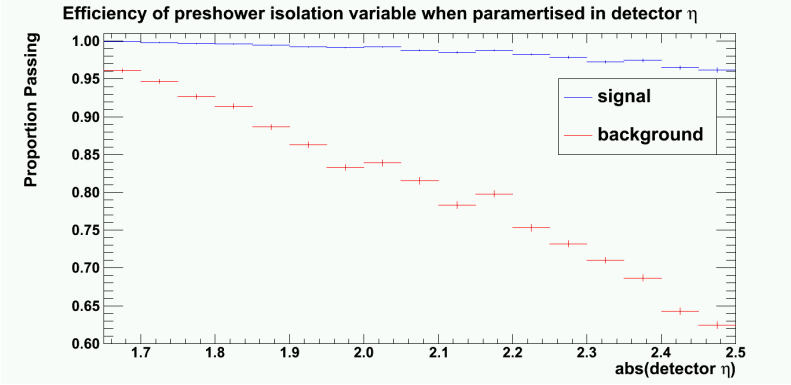

Figure 23: Efficiency of the preshower isolation cut at any absolute value of detector η for the signal and background samples.

Against detector ϕ, the variable works well with little variation in signal efficiency, with a minimum of 98 % being accepted at any point. The background efficiency does mainly vary between 70 % and 85 %, however this is not a problem as the signal efficiency remains reasonably stable for any ϕ position in the detector. One curiosity (picked up from Figure 24) was that at $\phi = \pm\pi$ there was an increase in background efficiency. Further investigation found that this was far more prominent in the "negative" end of the detector, where $\eta < 0$. This is most likely due to a slight fault in the detector, with little data being picked up in this region, perhaps because of "dead"

crystals or strips that are known to be located here, rather than an effect of the variable (which should have no asymmetries in η or ϕ).

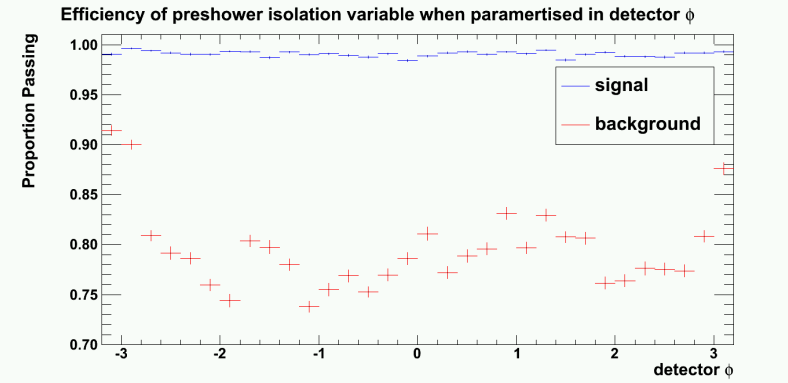

Figure 24: Efficiency of the preshower isolation cut at any value of detector ϕ for the signal and background samples.

A plot showing how the variable affected data in x-y co-ordinates was created(Figure 25). There was a noticeable trend where whenever ϕ lined up with the x or y axis (at $\phi = \pm\pi, \pm\pi/2, 0$) the electron footprint removal was not effective, particularly for strips that were 60 mm apart. This effect was more prominent in the signal sample than in the background sample, but was still present for the background. This is thought to be due to the structure of the preshower. In the preshower, each sensors' active are measures 61x61 mm^2 and made up of 32 strips, each 1.9 mm wide. These strips are oriented 90^o differently between the two planes, meaning that when one plane has strong η resolution, the other can only tell which of the strips the hit is located in. This meant that the η strip being vetoed by the variable would reject electron candidates even if they should pass the isolation cut.

The most obvious solution would be to change the width of the veto strip, however Figure 21 shows that increasing the strip width would reduce the power of the variable. This could be combined with an increase in initial area, however the radius of 0.3 in $\eta - \phi$ already uses about half the η range of the detector and 1/5 of the possible ϕ values and increasing this further is likely to result in overlap problems, where other electrons are being picked up in the non-isolated area.

The first solution tried to remove the effect was to modify the variable so that it would only use the plane with good η resolution for electrons in the regions where it proved to be a problem. For the ϕ regions where of $\pi/4 < |\phi| < 3\pi/4$ the plane with strong y resolution was used, and the other plane was used for everywhere else. This allowed for a tightened cut point. It did however cut less events with only 19 % of events being cut. This makes sense as the signal electrons being cut are from the tail of the distribution. These are likely caused by edge affects, which the discrete areas are not attempting to fix. It also meant that the preshower wasn't being used to its fullest as half the area of each plane were not being used.

A second attempted fix was to veto any strip that is even partially in the veto region. This was achieved by calculating the x and y values of the edges of the hit regions when the resolution was worst (i.e. when η lined up parallel to x or y). If the hit was detected in the first plane (where the y resolution was poor), the maximum and minimum y values of the hit location were calculated and used to find the maximum or minimum η values of the strip. If either of these was found to be in the "η vetoing" region, it would be rejected. This proved to be even less effective, with only 17.5 % of the remaining background events being rejected for a 1 % signal loss. More interestingly, it also didn't manage to remove the curious detector structure effect that it was designed with the intent of removing. It did however make the gaps when no electron footprint were being found

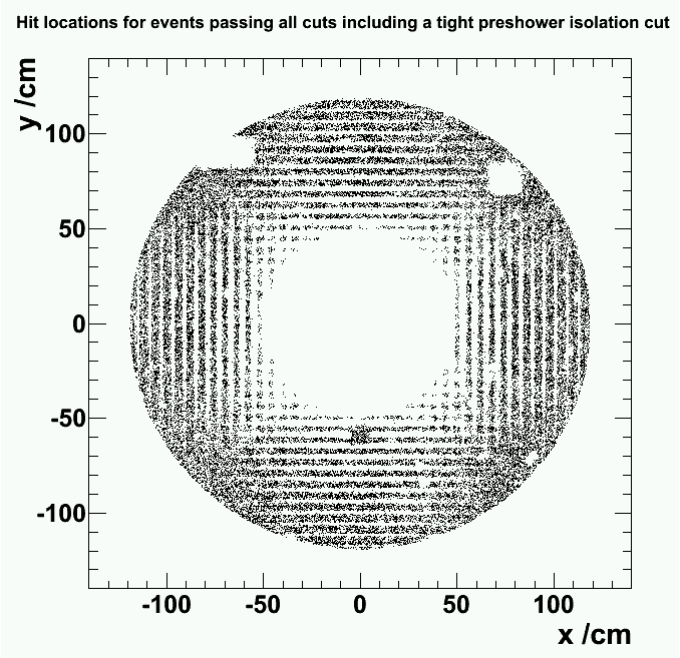

Figure 25: Positions of hits still accepted after the preshower isolation cut (tighter than when actually used for emphasis) has been applied to the positive endcap.

thinner, particularly for the one in the varying x direction towards the outer edge of the detector. No suggestion as to why this has failed to completely remove this quirk has been put forward so far.

Another possible future attempt to fix it which could yield an even more powerful variable is to try associating clusters between the two planes of the preshower in order to get a more precise lock on the position of the hit within the strips the hits are found in. This is a far more complicated solution and as such has yet to be attempted.

From these attempts, it was determined that for the time being this variable was as strong as it could be and would be taken to be the best available variable. Further tests were run to confirm that it could perform consistently against other parameters.

The signal efficiency when tested against E_t was less consistent than against ϕ, but still accepted at least 95 % of the signal sample for energies up to around 140 GeV. Above this threshold, the sample of Z→ee events did not have enough data to give accurate results (as seen by the increasing error bars in Figure 26). The background efficiency was also higher for low energies, with over 95 % of events with $E_t < 40\,GeV$ passing the isolation cut. However, this quickly dropped, and by $E_t = 120\,GeV$ less than 80 % of events were passing the cut.

When it was tested against the number of interaction vertices, the proportion of signal events passing dropped steadily from 100 % to 98 % between 2 and 25 vertices before a slight increase and bump as it increased into a region with far less data. For the points where there is a reasonable amount of data, the efficiency is as expected. The background on the other hand, behaved a little less predictably. When there were less interaction vertices, a larger proportion of events passed. This looks to be because of there being less events in this region. When more than 5 interaction vertices were present, the background efficiency was reasonably stable, decaying similarly to the signal ,except less smoothly and consistently until it starts to be even more sporadic when over 30 interactions occurred (this again looks like the lack of events plays an important factor). All this is shown in Figure 27.

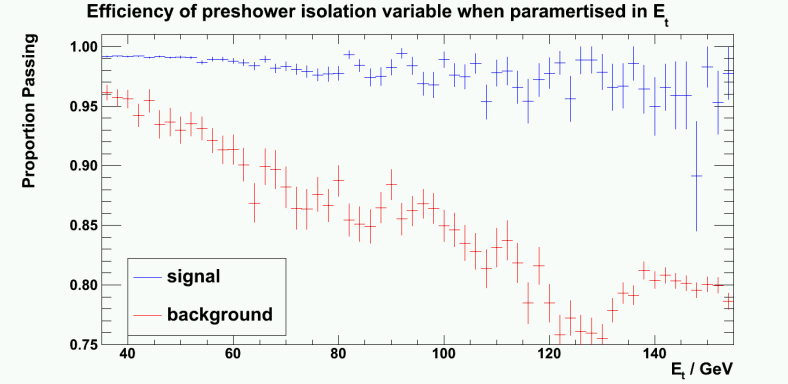

Figure 26: Efficiency of the preshower isolation cut at possible values of E_t up to 150 GeV. for the signal and background samples.

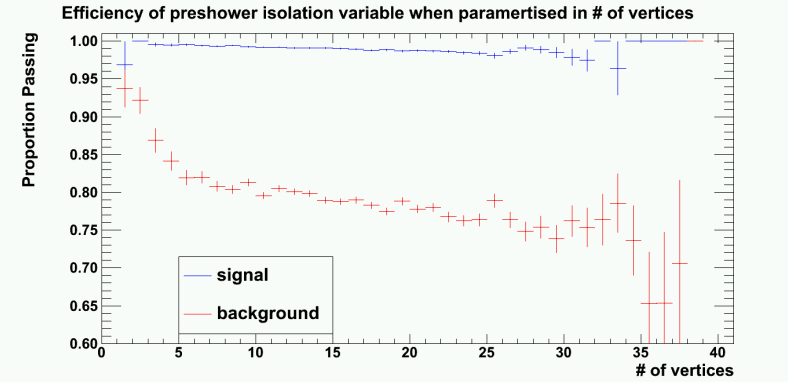

Figure 27: Efficiency of the preshower isolation cut at any number of interaction vertices for the signal and background samples.

It was important to ensure that the isolation variable works equally well for both high energy electrons as well as for the low energy Z→ee electrons that were studied (which has a peak signal energy of 45 GeV). An additional cut on E_t was imposed, removing all events where $E_t < 100$ GeV. This cut most of the signal, but few background events. In detector η, the signal efficiency above 100 GeV decayed faster from the edge to the centre of the endcap, however it can be put down to lack of data (only 4500 of the 330 000 events that passed all the standard cuts were left) as much as to weaker performance for higher energy ranges. The background rejection did not change much for high energies compared to any energies (above 35 GeV which is one of the standard cuts). This was down to the trigger selections, which resulted in a large number of higher energy fake electrons.

Once the variable was tuned, it was tried on data obtained from Monte Carlo simulations. A maximum mass cut of 120 GeV was also imposed on both the data and the Monte Carlo simulation data in order to compare only around the Z-peak. If this had not been applied a more substantial difference between the efficiencies would be seen. Including this cut gave strong agreement between the data and Monte Carlo simulation, with only 0.6 % less Monte Carlo simulated events passing the cut than data events. Figure 28 shows how closely the distribution for the simulated isolation variable compares with the distribution of the signal data sample.

It was also checked against E_t and Figure 29 shows how its efficiency against E_t compares to the signal data sample.

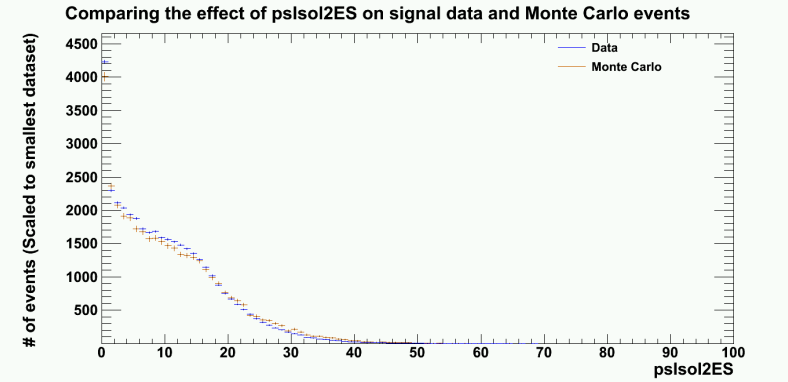

Figure 28: Efficiency of the preshower isolation cut at any number of interaction vertices for the signal and background samples.

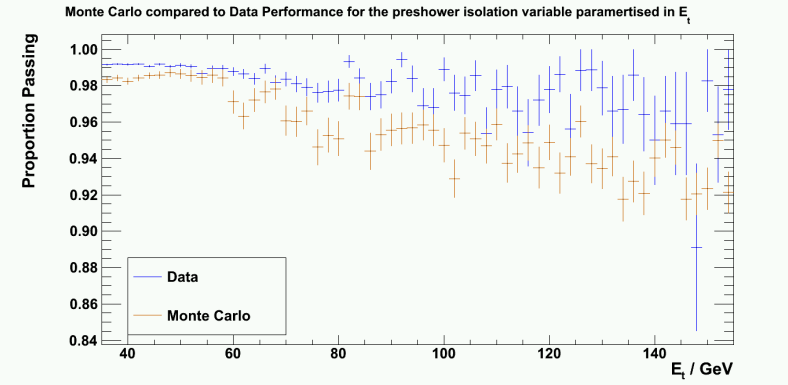

Figure 29: Efficiency of the preshower isolation cut at any number of interaction vertices for the signal and background samples.

To put in context how powerful this variable was, table 2 compares it to two of the cuts currently used for electron identification in the endcap of the ECAL. For consistency all cuts, including the preshower isolation were applied, except the cut being tested. This is known as (n-1) cutting.

Table 2: Endcap cut performance comparison using (n-1) cut method for all energies, and then energies above 100 GeV.

Cut	Signal Efficiency	Background Rejection	Signal Efficiency($E_t > 100$)	Background Rejection($E_t > 100$)
$\sigma_{i\eta i\eta}$	99.1 %	4.1 %	93.6 %	27.4 %
Calorimeter Isolation	75.9 %	32.9 %	93.2 %	29.7 %
Preshower Isolation	99.1 %	20.9 %	96.9 %	22.5 %

The preshower isolation variable performs far better than $\sigma_{i\eta i\eta}$, cutting five times more background for a very similar reduction in signal. It also compares favourably to the calorimeter isolation variable used (EM+Had Depth 1). Even though it cuts around 2/3 of the background, it does so for a far lower reduction in signal. At higher energies, the preshower isolation background

rejection is slightly lower than both $\sigma_{i\eta i\eta}$ and ECAL isolation, however it also keeps an extra 3 % of signal. These fact show that it is a powerful variable for electron identification.

7 Implications for Z' Analysis

The endcap-endcap dielectron production channel has a large jet background which has so far not been reduced sufficiently. The fake rate of electron detection in the endcap was, in places, five times higher for the high $|\eta|$ endcap regions than for the barrel region. Because of this the Z' analysis has not be fully conducted for the endcap-endcap channel.

The Z' analysis, which measures the high mass Drell-Yan spectrum, was run for the endcap-endcap channel with the addition of the preshower isolation variable, to see how much additional power it gave. Figure 30 shows the improvement in fake rate achieved.

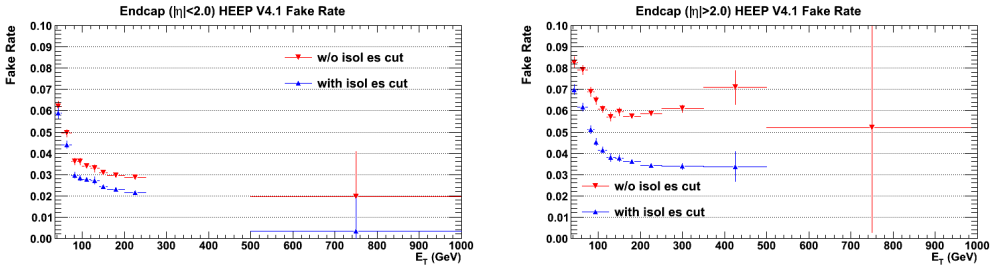

Figure 30: Comparing fake rates for the $|\eta| < 2.0$ region of the endcap (left) and the $|\eta| > 2.0$ region of the endcap (right) with and without the preshower isolation cut.

The cut is reducing the fake rate to 2/3 of its value without the cut, and even further in places. It might still be possible to reduce this further, but for now this is sufficient improvement. Unlike many ID variables, this is more powerful for the high η region, reducing the fake rate considerably more than in the low η region.

The first two graphs of Figure 31 compares the event distributions with and without the cut. It is difficult to see the differences between the two graphs, so the third graph shows the proportion of background remaining after applying the cut to the endcap-endcap channel for this Z'→ee analysis.

At lower energies, the performance of this variable looks less powerful than it is. Below 400 GeV, the background is dominated by the γ/Z background. This is indistinguishable from the signal, and therefore cannot be reduced without cutting signal events. This results in only a small fraction of background being cut. At higher energies, the jet background dominates, and more than 45 % can be rejected by the preshower isolation cut.

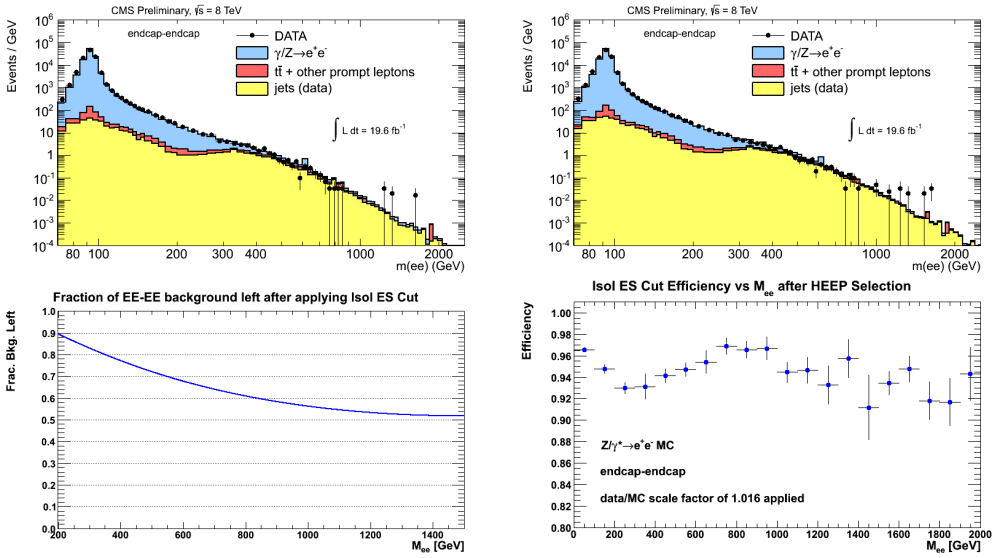

Figure 31: Comparing the event distributions with (top left) and without (top right) the preshower isolation cut. As the difference isn't especially clear, the bottom graph gives a comparison of the amount of additional signal (bottom left) and background (bottom right) being cut.

8 Conclusions

This project attempted to find a shower shape variable that could be used in place of $\sigma_{i\eta i\eta}$ in the endcap of the ECAL. It also tried to find a way to utilise the preshower detectors for electron identification in the endcaps.

The shower shape variable used in the endcap ($\sigma_{i\eta i\eta}$) still doesn't give as much distinction between electron candidates and background events as would be liked and none of the various variable tried in the endcap were shown to be any better. It is likely that $\sigma_{i\eta i\eta}$ is the most suitable shower shape variable for use in the endcap of the ECAL without the assistance of the preshower.

Using the preshower detector to support electron identification in the ECAL, an additional shower shape variable, focussing on the spread in the preshower was developed. A variation on $\sigma_{i\eta i\eta}$ could be applied in addition to $\sigma_{i\eta i\eta}$ and the other selection cuts to reduce the amount of background further than editing the cut point of the current $\sigma_{i\eta i\eta}$ variable or changing it any obvious way. It is unlikely to work as well as it can in its current form since it suffers from poor resolution in x or y. Further work should be attempted to tweak and hone it into a usable variable for electron identification.

It was also found that an isolation variable that can be applied on top of all other endcap selection cuts, and any potential preshower shower shape variables resulting in a significant improvement in background rejection, rejecting 21 % of the remaining background events for a loss of less than 1 % of the signal. Using this variable improved electron identification in the endcap. This variable is set to be added to the standard selection criteria for use by those studying Z'→ee endcap electrons in the ECAL of the CMS detector. As a benchmark, in the endcap-endcap channel it rejected around 45 % of the background for 5 % efficiency loss at 1 TeV. It was also shown to outperform, or perform comparably with existing ID variables in the endcap.

Acknowledgements

Firstly, Big thank yous go to my supervisors, Sam, David and Dave, for all the hard work they've put in to helping me with this project, particularly Sam, who has taught me a lot about both Physics and computing, as well as helping me with my project as a whole. Thank you to Tom, whom I shared a desk with for most of this period, for helping me when I get stuck every 5 minutes, as well as for putting up with my LaTeX induced rants. Thank you to the LHC team, for creating the engineering marvel that produced all the data I've studied. Thank you to Alex, Ben, Declan, Nicole and Rhys, all who have help keep me sane (relatively speaking), and to Hugo for allowing me to unleash my insanity upon him. Thank you to Kevin and the rest of the UKLO, who have provided me with a place to stay as well as general support. Thank you to my friends and family back home, all who have provided me with great support. And of course thank you to Stefano, Steve and the rest of the staff at Southampton University for making this course possible.

References

[1] Electron reconstruction and identification at $\sqrt(s) = 7$ TeV. Technical Report CMS-PAS-EGM-10-004, CERN, Geneva, 2010.

[2] Performance and operation of the CMS electromagnetic calorimeter. *Journal of Instrumentation*, 5(03):T03010, 2010.

[3] B. Clerbaux et al. Search for High Mass Resonances Decaying to Electron Pairs at 8 TeV. *CMS Note*, CMS AN-2012/171, 2012.

[4] Serguei Chatrchyan et al. Search for heavy narrow dilepton resonances in pp collisions at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV. *Phys.Lett.*, B720:63–82, 2013.

[5] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[6] The Particle Data Group. The Review of Particle Physics. *Phys. Rev. D*, 86(010001), 2012.

[7] Kai-Yi Kao. Alignment of the CMS Preshower Detector. Technical Report CMS-CR-2012-025. CERN-CMS-CR-2012-025, CERN, Geneva, Jan 2012.

[8] A. Leike. The phenomenology of extra neutral gauge bosons. *Physics Reports*, 317(34):143 – 250, 1999.

[9] David A. Petyt. Anomalous APD signals in the CMS Electromagnetic Calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 695(0):293 – 295, 2012. New Developments in Photodetection NDIP11.

[10] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, 2008.