Electron-based longitudinal weights for the ATLAS EM Barrel Calorimeter and shower isolation studies with an application to the $H \rightarrow ZZ^{(*)} \rightarrow 4e$ analysis

B. Mellado, S. Paganis, W. Quayle, Sau Lan Wu

Physics Department
University of Wisconsin - Madison
Madison, Wisconsin 53706, USA

8 November 2004

Abstract

In this note a full electron-based calculation of longitudinal weights for the Electromagnetic Barrel Calorimeter is performed. Good performance in resolution and linearity for electrons of energies ranging from 10 GeV up to the TeV scale is demonstrated. A general method is proposed which can be applied in multi-lepton final states where detector level information is exploited to discriminate between signal and background. The method is applied to the $H \rightarrow ZZ^{(*)} \rightarrow 4e$ channel.
1 Introduction

ATLAS physics analyses involving processes with electrons or photons in the final state require optimum calorimetric resolution, linearity and accurate knowledge of the electromagnetic energy scale [1], [2]. Understanding of the calorimeter response has been the main goal of ATLAS Liquid Argon Calorimeter Test-Beams [3], [4], [5], [6]. In particular the most recent understanding of the linearity and resolution of the Barrel Calorimeter is described in [7].

The challenge we are facing in ATLAS is the large amount of inactive material upstream the Liquid Argon (LAr) electromagnetic calorimeter (EMC) shown in figure 1. The upstream material causes pre-showering of electrons and photons before they reach the EMC. This results in a calorimetric response which is different for electrons and

![Pseudorapidity](image)

Figure 1: Material thickness distribution (in $X_0$) versus pseudorapidity from the interaction point to the face of the electromagnetic calorimeter in ATLAS. This map is already out of date; in the latest ATLAS design there is slightly more material in the Inner Detector region.
photons. In addition the upstream material causes photon conversion and electron bremsstrahlung which complicate reconstruction. Therefore a calibration of the LAr EMC must take into account the pseudorapidity ($\eta$) dependent distribution of upstream material in ATLAS and the type of particle which is responsible for a shower in the EMC.

In this note an electron-based calculation of longitudinal weights for the ATLAS Electromagnetic Barrel Calorimeter is performed with a new parametrization motivated by recent Test-Beam analyses [8]. This new calculation is compared to the present default Data Challenge 1 (DC1) ATLAS calculation which used Monte-Carlo photon beam samples.

The note is structured as follows: first the Monte Carlo (MC) samples are described, then a study of the ATLAS Electromagnetic Calorimeter longitudinal weights with electrons is presented, together with the approach to combine EMC with Inner Detector (ID) information for optimum performance. Subsequently the performance of the new longitudinal weights for electrons of energies ranging from 10 GeV up to the TeV scale is studied and compared with the present default ATLAS weights. Finally a new analysis method is proposed which can be applied in multi-lepton final states where detector level information is exploited to discriminate between signal and background. The method is specifically applied to the $H \rightarrow ZZ^{(*)} \rightarrow 4e$ channel.
2 Monte-Carlo Sample Production

A large set of MC data samples was generated for the purposes of this analysis. The full set of MC files is presented in table 1.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Run Number</th>
<th>Events</th>
<th>Passed Filter</th>
<th>Gen./Sim. V.</th>
<th>Recon. V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H(130GeV) → 4e</td>
<td>2505</td>
<td>100k</td>
<td>63684</td>
<td>6.5.0/6.5.0</td>
<td>6.5.3</td>
</tr>
<tr>
<td>H(150GeV) → 4e</td>
<td>2506</td>
<td>100k</td>
<td>65000</td>
<td>6.5.0/6.5.0</td>
<td>6.5.3</td>
</tr>
<tr>
<td>H(180GeV) → 4e</td>
<td>2507</td>
<td>100k</td>
<td>71618</td>
<td>6.5.0/6.5.0</td>
<td>6.5.3</td>
</tr>
<tr>
<td>H(200GeV) → 4e</td>
<td>2541</td>
<td>100k</td>
<td>73549</td>
<td>6.5.0/6.5.0</td>
<td>6.5.3</td>
</tr>
<tr>
<td>H(300GeV) → 4e</td>
<td>2542</td>
<td>100k</td>
<td>77241</td>
<td>6.5.0/6.5.0</td>
<td>6.5.3</td>
</tr>
<tr>
<td>pp → ZZ → 4l</td>
<td>2648</td>
<td>12.5M</td>
<td>559707</td>
<td>6.5.0/7.0.2</td>
<td>7.0.2</td>
</tr>
<tr>
<td>gg → Zb → 4e</td>
<td>2658</td>
<td>120M</td>
<td>568k</td>
<td>6.5.0/6.5.0</td>
<td>7.0.2</td>
</tr>
<tr>
<td>gg → tt → 4e</td>
<td>2599</td>
<td>5.58M</td>
<td>43361</td>
<td>7.0.0/6.5.0</td>
<td>6.0.3</td>
</tr>
</tbody>
</table>

The generation definition for the standard Higgs and background samples (pp → ZZ → 4l) is summarized below:


- Simulation using Atlsim 01-03-08 (Geant 3.21).

- Cuts in the particle level filter at $|\eta| \leq 2.7$, $p_\perp > 4$ GeV (for 4 electrons).

- Cuts in the simulation at $|\eta| \leq 2.7$, (for 4 electrons).

- Reconstruction using Athena 6.5.3 for signal, and 7.0.2 for background (for $t\bar{t}$ see below). The standard $e/\gamma$ job options for low luminosity were used. Electronic noise in the CALO was switched off. No pile-up was added.

- For the 4$\mu$ final state the Data Challenge 1 (DC1) samples were used [10].

Samples were also generated at Next-to-Leading Order (NLO) for Higgs and background ($t\bar{t}$):

- Generation in Athena 7.0.0, using Herwig [11] and MC@NLO [12], and filtering in the final state using a modified MultiLeptonFilter.

- Simulation using Atlsim 01-03-08 (Geant 3.21).

- Cuts in the particle level filter $|\eta| \leq 2.7$, $p_\perp > 4$ GeV (for 4 electrons).
• Cuts in the detector simulation at $|\eta| \leq 2.7$, (for 4 electrons).

• Reconstruction using Athena 6.0.3, with the $e/\gamma$ eg9.lumi02.603.job options (Standard $e/\gamma$ job options, low luminosity).

The EMShowerBuilder algorithm in Athena was modified so that the low luminosity shower cuts in EM identification are applied.
3 Longitudinal Weights for the LAr EM Calorimeter

In this section an electron-based calculation of the LAr EM Barrel calorimeter longitudinal weights is performed for $|\eta| < 1.37$. For the End-Cap calorimeter (EMEC) an overall $\eta$ dependent scale correction is applied, since the simulation of the detector response did not take into account the high voltage variation with $\eta$ in offline releases prior to 8.0.0.

An electron-based recalculcation of the LAr EM calorimeter weights was deemed essential for the following reasons:

- The present DC1 weights within the full ATLAS reconstruction software were obtained using beams of photons.

- Energy weights including corrections due to the presence of upstream material were applied at the cell level prior to clustering. This is problematic because corrections are not universal but particle dependent (for example different for electrons, photons and early versus late converted photons).

- Test-Beam data analyses using electron beams of fixed energy suggest modifications of the present calibration parametrizations of the calorimeter energy response. [7,8].

In the region $|\eta| > 0.8$, the upstream material thickness in ATLAS increases, putting into question the requirement of the LAr linearity of better than 0.5%. This requirement is particularly important for the $W$-mass measurement and $H \rightarrow \gamma\gamma$ analyses. To more quantitatively motivate a recalculation of the longitudinal weights, we make use of a LAr Toy MC which was originally setup for Test-Beam studies [13].

3.1 Overview of LAr Toy MC Studies

The intrinsic linearity and resolution of the LAr EM calorimeter (EMC) is achieved for EM showers which originate and are fully contained in the calorimeter. This is an ideal case since the EMC is positioned inside a cryostat with an Aluminum inside wall of 8 cm which is the equivalent of 0.9 $X_0$ at normal incidence. As a result, the LAr EMC weight given as

$$w = \frac{E_{\text{LAr}} + E_{\text{Ph}}}{E_{\text{LAr}}}$$

(i.e., the inverse of the sampling fraction) acquires a small energy dependence. This is demonstrated in figure 2 (left), through a Geant 4 Toy Monte Carlo Model in which electrons of varying energies are sent toward a 10 cm thick block of Aluminum behind which a presampler and a LAr EMC volume have been positioned [13]. Increasing the block thickness (shown in the right plot, figure 2) causes an increase of the energy
dependence. The effect is largest at low beam energies. This is mainly due to the shower commencing before the EMC. For lower energies, there is a smaller shower-max depth, leading to the sharing of the energy by a higher multiplicity of soft particles which are more likely to be trapped in the Lead absorber. It must be stressed that this energy dependence remains after the energy lost upstream of the EMC is corrected. In ATLAS, a 25 cm thickness is representative of the central region of the detector.

As an example, a 40 cm thick Al block is used in figure 3 where both electrons (denoted by the filled circles) and photons have been used. To demonstrate the effect of the upstream material, one can examine the ≈ 5% of non-interacting photons (i.e., photons that traverse the Al block or the presampler without interacting) and see that the corresponding weights behave as in the case of Figure 2 (left) for electrons and very little upstream material. This indicates that while in the ideal (i.e., no-upstream material) case photons and electrons lead to the same EMC response, the addition of upstream material changes the picture and probably necessitates a different treatment of photon-induced clusters. Study of these effects for different amounts of upstream material with photon and electron samples is planned during the ATLAS 2004 Combined Test Beam (CTB04).

3.2 EM Barrel Calorimeter Longitudinal Weight Calculation

The official LAr Calorimeter calibration strategy is in preparation [14]. In the ATLAS offline software, longitudinal weight corrections are currently applied at the cell level while EM cluster information is typically available at the cluster level for fixed cluster sizes at the ntuple level. The present nominal ATLAS weighting scheme in the reconstruction software, is to weigh linearly the energy of every cell as a function of the
Figure 3: LAr EMC weight \((w = (E_{\text{LAr}} + E_{\text{Pb}})/E_{\text{LAr}})\) dependence on beam energy for electrons (filled circles) and photon (rectangles) beams going through 40 cm of Al. The photons which do not interact with the Al or the presampler are also shown (triangles).

Pseudorapidity \(\eta\):

\[
E_{i,\text{rec}}^{\text{cell}} = W_i^{\text{cell}}(\eta) E_i^{\text{cell}}
\]

where \(i\) is one of the presampler (PS), strips (S1), middle (S2), and back (S3), longitudinal samplings of the LAr Barrel (EMB) and End-Cap (EMEC) calorimeters. The weights \(W\) depend on \(\eta\) because of the \(\eta\) dependence of the material thickness upstream of the calorimeter (ID, supports, cryostat, cables, etc) as shown in figure 1. In the absence of upstream material the weights (for a fully contained shower) depend only on the geometry of the cell of the LAr calorimeter, so for example for the EMB there are two average (intrinsic) weights: one for \(\eta < 0.8\) and one for \(\eta > 0.8\). For the EMEC the situation is complicated due to variations of the gap width and the High Voltage setting. The presence of upstream material destroys the universality of the LAr response on electrons and photons. This requires different corrections depending on the particle hypothesis at the cluster level, therefore it is essential to have available the uncorrected (or corrected up to a known constant factor) cell information. The correct approach [14] is to decouple the upstream material effects and leakages from the intrinsic universal calorimeter response. A summary of possible steps concerning the calibration of the EMB after electronics calibration follows:

- Apply an average constant intrinsic weight at the cell measured energy. This weight is the ratio between the deposited energy and visible energy in this cell. It provides an initial guess of the true deposited energy in the cell.
• Apply cell-level corrections (LAr purity, temperature variation, intercalibration weights, etc).
• Perform clustering.
• Perform particle identification using the ID and Calorimeter.
• Apply \( \eta \) and particle identification dependent weights to the cells consisting the cluster.
• Recalculate the cluster energy for particular cluster sizes/algorithms.
• Apply lateral leakage corrections which depend on \( \eta \) and cluster size or algorithm.
• Apply longitudinal leakage corrections.

There are more corrections that need to be applied (for example \( \eta \), \( \phi \) modulations or corrections for the intrinsic non-linearity of the LAr calorimeter, which are smaller when compared to the weight corrections) but these are out of the scope of this note.

For the calculation of the longitudinal weights we follow a modified parametrization for the energy reconstruction [8]:

\[
E_{\text{rec}} = \lambda (b + W_0 E_{\text{pres}} + E_1 + E_2 + W_3 E_3)
\]

where \( b \) is an offset with the units of energy, motivated by recent electron Test-Beam analyses. This offset was found to optimize simultaneously electron energy linearity and resolution. The weight \( W_3 \) for the third sampling is supposed to be correcting for longitudinal leakage. For energies below 50 GeV the energy of the back sampling is not used (noise dominated).

The longitudinal weights (including the offset) can be obtained by fitting electron beams of known energy or \( Z \rightarrow ee \) events where the electron energy or the \( Z \) mass are used. For \( |\eta| < 1.37 \) we use equation 2. For \( |\eta| \geq 1.37 \) we do not uncorrect but instead we extract an overall factor (typically \( \simeq 1.01 - 1.02 \)). A full calibration of the EMC for \( |\eta| < 2.5 \) will be performed with Data Challenge 2 (DC2) for offline releases starting from 9.0.0. In these releases the Geant-4 simulation will be used and the default calibration will not include corrections for energy lost upstream the calorimeter.

### 3.3 Longitudinal weights after uncorrecting DC1 weights

The MC samples used in the present analysis still use \( \eta \) dependent weights which have been applied at the cell level during reconstruction. However cell energies are not available at the ntuple level (CBNT) and the easiest option at the moment is to uncorrect the already applied weights at the cluster level.

The longitudinal weight calculation approach used in this note consists of the following steps:
- Remove the existing DC1 cell-level weights: based on the $\eta$ of the EM cluster, we divide the cluster energy by the corresponding known weight. This procedure ignores the fact that since the cluster size is $3 \times 7$ (second sampling units), a cluster consists of 3 $\eta$ bins, each with a different weight;
- perform calibration with electrons from $Z \to ee$, using the parametrization 2 for $|\eta| < 1.37$;
- use the electron-photon block of the CBNT, i.e. fixed cluster size with 3 bins in $\eta$ and 7 bins in $\phi$;
- truncate non-gaussian tails by keeping clusters $3\sigma$ around the peak of the distribution;
- perform minimization of a function for all clusters belonging to a specific $\eta$ bin. Here we minimize $\sum \frac{(E_{\text{rec}} - E_{\text{true}})^2}{\sigma_{EMB}}$, where $E_{\text{rec}}$ the reconstructed energy, $E_{\text{true}}$ the true electron energy and $\sigma_{EMB}$ is a simple calorimeter resolution parametrization.

The resulting longitudinal weights are shown in figure 4.

![Graphs showing longitudinal weights](image)

Figure 4: Electron-based longitudinal weights as calculated from the method described in Section 3.3.
3.4 Longitudinal weights using uncorrected LAr cells

The longitudinal weights were also calculated using electron beam samples of fixed energies (10, 50 and 100 GeV) without prior application of upstream material dependent weights on the cells \[15\]. In these samples, there was only an $\eta$ dependent cell-level scale correction applied. This scale weight is the same for all samplings (including the PS) and is shown in fig. 5.

![Intrinsic Barrel Scale Factor (LC)](image)

Figure 5: Overall $\eta$ dependent scale factor applied to all LAr EMC samplings.

These samples were subsequently fitted as follows:

- Perform calibration with fixed energy electrons, using the parametrization 2 for $|\eta| < 1.37$.
- use cluster energy using all cluster cells (i.e. no specific fixed size);
- truncate non-gaussian tails by keeping clusters $3\sigma$ around the peak of the distribution;
- perform minimization of a function for all clusters belonging to a specific $\eta$ bin. Here we minimize $\sum \frac{(E_{\text{rec}} - E_{\text{true}})^2}{\sigma_{EMB}^2}$, where $E_{\text{rec}}$ the reconstructed energy, $E_{\text{true}}$ the true electron energy and $\sigma_{EMB}$ is a simple calorimeter resolution parametrization.

The extracted overall scale parameter $\lambda$ is shown in figure 6. The distribution follows the upstream material thickness as a function of the pseudorapidity. The small increase at $\eta = 0$ is due to the gap between the two parts of the barrel calorimeter.
Figure 6: Overall energy scale fitted parameter $\lambda$. The distribution follows the upstream material thickness as a function of the pseudorapidity. The small increase at $\eta = 0$ is due to the gap between the two parts of the barrel calorimeter.

Longitudinal weights obtained with fixed energy electron beams presented here, lead in general to small improvement in resolution and linearity with respect to the method used in section 3.3. However we currently do not have independent uncorrected samples to perform detailed studies so we proceed with the weights obtained from the standard MC samples as described in section 3.3.

### 3.5 Calorimeter-Inner Detector combination

At low electron energies (roughly below 50 GeV) the inner detector has comparable or better electron energy resolution than the EMB/EMEC calorimeters. In this section we assume a well calibrated ID (can be confirmed with $Z \rightarrow \mu\mu$ samples). In this energy regime we would like to improve the electron energy reconstruction by combining the information from the two detectors. The combination is problematic due to the presence of upstream material: the electrons radiate photons and the reconstructed $p_{\perp}$ by the ID is typically smaller and has a very long low energy tail.

In order to properly combine the ID and EMB/EMEC information we select tracks satisfying the following conditions:

- Tracks which do not radiate or radiate a soft photon.
- Tracks which radiate late (i.e. after the ID and before the EMB/EMEC).
The discriminating variable used in this analysis is $E/p$ where $E$ is the energy reconstructed by the calorimeter and $p$ the momentum reconstructed by the ID. For calibrated (for energy scale) ID and EMB/EMEC, a track/cluster with $E/p \approx 1$ satisfies the conditions above. This is demonstrated in figures 7 and 8 where the $p_\perp$ resolution is plotted for the EMB/EMEC (solid line) and the ID (dashed line). In the case of figure 7 a cut of $E/p > 1.03$ is applied to demonstrate that a combination between the two detectors is not possible. For $p_\perp < 60$ GeV, this corresponds to $\approx 65\%$ of the electrons. In figure 8 a cut of $0.98 < E/p < 1.02$ is applied and the ID $p_\perp$ resolution becomes better than the calorimeter one, while at the same time the energy scales for both detectors agree. For $p_\perp < 60$ GeV, this corresponds to $\approx 35\%$ of the electrons. It must be stressed here that the $E/p$ variable cannot be used prior to a proper EMB/EMEC calibration as described in the previous section, and the performance of the combination described here is critically dependent on the knowledge of the calorimeter energy scale.

![Graph](image)

Figure 7: $p_\perp$ resolution for the EMB/EMEC (solid line) and the ID (dashed line) for $E/p > 1.03$ and $p_\perp < 60$ GeV. This corresponds to $\approx 65\%$ of the electrons.

In the analysis reported in this note a combination of the energy of tracks/clusters is performed for $0.98 < E/p < 1.02$. The combined energy is simply the weighted mean of the two independent measurements, where the weight is obtained by a parametrization of the energy resolution. The following parametrizations were used:

$$\sigma_{\text{CAL}} = 0.1\sqrt{E_{\text{CAL}}}$$

$$\sigma_{p_\perp,\text{ID}}/p_\perp = 0.015 \oplus 0.0004p_\perp$$

Several different parametrizations were used for the ID without a large effect in the analysis.
One way to extract the weighted energy is to use [16]:

$$W_{ID} = 0.5 \operatorname{erf}(x - 1.5)/\sigma_{ID}^2 \quad W_{CAL} = (1 - 0.5 \operatorname{erf}(x - 1.5))/\sigma_{CAL}^2,$$

where $x$ is defined as:

$$x = (E_{CAL} - E_{ID})/\sqrt{\sigma_{CAL}^2 + \sigma_{ID}^2},$$

and finally the weighted energy is:

$$E_{Weighted} = (E_{ID}W_{ID} + E_{CAL}W_{CAL})/(W_{ID} + W_{CAL})$$

Alternatively one can extract the weighted energy from [17]:

$$E_{Weighted} = (E_{ID}\sigma_{CAL}^2 + E_{CAL}\sigma_{ID}^2)/(\sigma_{ID}^2 + \sigma_{CAL}^2).$$

Both methods give similar results.
4 Longitudinal Weight Performance

The performance of the electron-based longitudinal weights described in section 3 was tested using independent electron samples. In this section we first test the performance of the weights on low energy electrons and $H \rightarrow ZZ'(s) \rightarrow e^+ e^- e^+ e^-$ samples. The new weights were also used in a recent $Z' \rightarrow ee$ analysis for $Z'$ masses of 200 MeV, 1.0 TeV and 1.5 TeV [18]. The high energy electron resolution and linearity from $Z' \rightarrow ee$ is presented here and demonstrates the good performance of the new longitudinal weights.

4.1 Low energy electrons and Higgs samples

Comparison between application of our longitudinal weights with the default in ATLAS software is shown in figure 9 where at least the linearity of electrons coming from $H \rightarrow ZZ'(s) \rightarrow e^+ e^- e^+ e^-$ decays using the full ATLAS simulation and reconstruction is improved when compared with the current standard EMC weights. It is also apparent that the default photon based weights applied to electrons gives $\approx 0.8\%$ lower energy scale.

![Graphs showing electron linearity and resolution comparison](image)

Figure 9: Electron linearity (left) and resolution (right) for the standard EMC longitudinal weights (filled circles) and the new weights including a better treatment of upstream material (squares). Electrons come from $H \rightarrow ZZ'(s) \rightarrow e^+ e^- e^+ e^-$ samples.

The $H \rightarrow 4e$ invariant mass distributions for Higgs masses of 130, 150, 180, 200 and 300 GeV are shown in figures 10. In addition to the calibration applied to the EM clusters a constrained fit was applied. A comparison between the result of the present analysis and TDR is shown in figure 11. There is a deterioration in the mass resolution...
for Higgs masses below the ZZ threshold which comes from the larger Bremsstrahlung tails from electrons due to the increase of the upstream material.

![Graphs showing ZZ invariant mass distributions for Higgs masses of 130, 150, 180, 200 and 300 GeV.]

Figure 10: $H \rightarrow 4e$ invariant mass distributions for Higgs masses of 130, 150, 180, 200 and 300 GeV.

4.2 High energy electrons from $Z'$ samples

The longitudinal weights calculated in section 3 were used in a recent $Z' \rightarrow ee$ analysis [18]. In searches for $Z'$ with masses at the TeV scale, the quality of electron energy
Figure 11: $H \to 4e$ invariant mass widths for the present analysis (filled circles) and the TDR (triangles) for Higgs masses of 130, 150, 180, 200 and 300 GeV. Note that there is no data point for the TDR and a Higgs mass of 300 GeV.

resolution and linearity at very high energies is crucial. Such analyses present an excellent test-bench for EMB/EMEC calibrations, which typically use lower energy electrons, because it is important to know if the calibration constants are preserved for very high electron energies.

In figure 12 the electron resolution

$$R = (E_{\text{true}} - E_{\text{recon}})/E_{\text{true}}$$

is plotted for electrons coming from $Z' \to ee$ decays for a mass of 1.5 TeV. The solid line shows the result using the default ATLAS software longitudinal weights, the dashed and dotted-dashed lines show the result of our new weights for the EMB alone (blue dotted-dashed) and for the full EMB/EMEC calibration (green dashed). From the figures it is clear that the energy scale is well preserved to the ppm level while the resolution is significantly improved (by almost a factor of 2). Further tests at electron energies of 200 GeV and 1.0 TeV show similar behaviour: the energy scale is at the ppm level and the resolution is significantly improved [18].
Figure 12: Electron resolution \((E_{\text{true}} - E_{\text{recon}})/E_{\text{true}}\) plotted for electrons coming from \(Z' \rightarrow ee\) decays for a \(Z'\) mass of 1.5 TeV. The result using the default weights is shown by the black solid line, the result using the longitudinal weights calculated in this note is shown by the green dotted histogram. The result without EMEC corrections is shown by the blue dashed histogram.
5 Optimization: Classification of Electromagnetic Objects

In this section a new method aiming in increasing the signal (Higgs in this case) significance by exploiting the available electromagnetic (EM) shower information for multi-lepton final states is proposed. In particular the method is applied to the $H \rightarrow 4e$ channel using an algorithm performing a classification of the 4 EM LAr calorimeter objects.

The basic idea of the method is to optimize the way the final state objects are selected based on detector information. Currently detector based selection criteria are applied on the basis of a single EM object. So for example, in the case of electron isolation, cuts were optimized using special samples (e.g. dijet events) in order to obtain a certain efficiency ($\approx 90\%$) and jet rejection. Thus, for multi-electron final states like the $H \rightarrow 4e$ channel the overall efficiency is $\varepsilon \approx 0.9^4$ and the background rejection quite large. However the background rejection may still be very high by only requiring 3 isolated electrons and 1 non or loosely isolated. Our aim in this section is to demonstrate that this is the case in the $H \rightarrow 4e$ channel where 50–60% of the signal can be recovered without significant increase of the reducible background.

The “3+1” method can be applied in different channels and different searches. It should be emphasized that each case requires a separate study and detector selection criteria optimization.

5.1 Electron Isolation for $H \rightarrow 4e$

Our starting point is the very low electron efficiency ($\approx 10\%$) for $H(130) \rightarrow 4e$ [19], [20]. In particular at the EM isolation stage the signal efficiency is 0.56 for a Higgs mass of 130 GeV. In figure 13 the fraction of isolated electrons for $H \rightarrow 4e$ is shown as a function of $p_\perp$. It is clear that loss of efficiency occurs for $p_\perp < 20$ GeV. The same distribution for $Zb\bar{b} \rightarrow 4e$ and $t\bar{t} \rightarrow 4e$ is shown at figures 14 and 15. From these figures it is clear that the background involves electrons which are non-isolated. The degree of non-isolation increases at low $p_\perp$. The difference between signal and background is due to electrons coming from the b-lines for $Zb\bar{b}$ and $t\bar{t}$.

It is interesting to study the stages at which calorimeter clusters are classified as being isolated electron/photon ($e/\gamma$) objects. The first stage is the shower leakage in the hadronic calorimeter and the second stage is cuts in the second sampling of the EMB/EMEC (see appendix). The ratio of efficiencies for these two stages versus $p_\perp$ is shown in figures 16, 17 and 18 for signal, $t\bar{t}$ and $Zb\bar{b}$ backgrounds respectively. A striking conclusion from these plots is that most electrons coming from the signal are cut at the first two samplings of the calorimeter after passing the hadronic leakage cut, while electrons from the background are cut at the hadronic leakage stage. An explanation
Figure 13: Fraction of isolated electrons for $H \rightarrow 4e$.

Figure 14: Fraction of isolated electrons for $Zb\bar{b} \rightarrow 4e$.  

19
Figure 15: Fraction of isolated electrons for $t\bar{t} \rightarrow 4e$.

is that electrons from the b-quark lines are most likely to be surrounded by hadronic debris than signal electrons. The electrons from background are then longitudinally deeper, while the signal electrons are contained but they may fail the shower cuts due to large lateral fluctuations.

To better examine the nature of the electrons for signal and background we define a simple flag we call “HiggsEMweight” which characterizes the full 4e final state (i.e. the Higgs candidate) as follows:

- Isolated electrons: HiggsEMweight=1.
- Electron matched to track: HiggsEMweight=0.1.
- Electron not in the EMB/EMEC crack region: HiggsEMweight=0.01.

As an example, a Higgs candidate with 4 isolated electrons, 3 of them matched to tracks and none pointing to the crack would give HiggsEMweight=4.34. The HiggsEMweight distribution for $H(130) \rightarrow 4e$ and $t\bar{t} \rightarrow 4e$ is shown in figure 22, where we see that most of the signal loss involves candidates with 3 isolated electrons.

In general, in order to obtain an optimum signal reconstruction we have to find a test statistic. The simplest choice is the probability of a non-isolated electron to come from the Higgs decay (i.e. a signal electron), given a $p_{\perp}$:

$$p(e_{H \rightarrow 4e} | p_{\perp}) = \frac{\epsilon_{e(H)} g(p_{\perp}|e_{H \rightarrow 4e})}{\epsilon_{e(H)} g(p_{\perp}|e_{H \rightarrow 4e}) + \epsilon_{e(H')} g(p_{\perp}|e_{H' \rightarrow 4e})}$$  \hspace{1cm} (4)$$

where:

$g(p_{\perp}|e_{H \rightarrow 4e})$: the probability density function (PDF) for a non-isolated electron with
Figure 16: Fraction of events failing the Middle sampling cuts over the events failing the leakage in hadronic calorimeter cuts for Higgs mass of 130 GeV.

Figure 17: Fraction of events failing the Middle sampling cuts over the events failing the leakage in hadronic calorimeter cuts for $t\bar{t} \to 4e$. 

21
Figure 18: Fraction of events failing the Middle sampling cuts over the events failing the leakage in hadronic calorimeter cuts for $Z\ell\ell \rightarrow 4e$.

Figure 19: Fraction of events failing the Strips sampling cuts over the events failing the leakage in hadronic calorimeter cuts for Higgs mass of 130 GeV.
Figure 20: Fraction of events failing the Strips sampling cuts over the events failing the leakage in hadronic calorimeter cuts for $t\bar{t} \rightarrow 4e$.

Figure 21: Fraction of events failing the Strips sampling cuts over the events failing the leakage in hadronic calorimeter cuts for $Zb\bar{b} \rightarrow 4e$. 
Figure 22: HiggsEMweight distribution for $H(130) \rightarrow 4e$ and $t\bar{t} \rightarrow 4e$. A cut is typically placed at HiggsEMweight > 4 which corresponds to events with 4 isolated electrons. From this distribution it is seen that most of the signal loss involves candidates with 3 isolated electrons.
certain $p_\perp$ to come from signal; 
$g(p_\perp|e_{4\ell-3e})$: the PDF for a non-isolated electron with certain $p_\perp$ to come from background; 
$\epsilon_e(H)$, $\epsilon_e(t\bar{t})$: electron efficiencies for signal and $t\bar{t}$ background. 
Obviously the integral of the test statistic 4 in a certain $p_\perp$ range (bin) is the signal purity. 

Therefore it is important to study the signal purity for the 3 isolated electron candidates. The purity defined as $S/(S+B)$ where $S$ the signal and $B$ the background is shown in figure 23 without Higgs mass window cut. These plots show a significant signal content for the 3 isolated electron sample. Relaxing isolation for more than one electrons leads to a dramatic reduction of the signal purity $\sim 10^{-4}$ which motivates a $\sim 10^4$ background rejection.

![Purity plots](image_url)

Figure 23: Purity $(H(130) \rightarrow 4e)/(H(130) \rightarrow 4e + t\bar{t} \rightarrow 4e)$ (left) and $(H(130) \rightarrow 4e)/(H(130) \rightarrow 4e + Z\ell\bar{\ell} \rightarrow 4e)$ (right) versus $p_\perp$ for candidates with 3 isolated electrons. Purity after cuts including (excluding) the Strips sampling is shown in full triangles (circles). No Higgs mass window cut was applied.

### 5.2 Optimized $H \rightarrow 4e$ Analysis

We next apply a simple algorithm which takes into account the shower shape information used in the electron isolation criteria. In particular:

- We accept candidates with 3 isolated electrons and 1 non-isolated electron. All 4 electrons are required to have an associated matched track in the ID. The non-isolated electron is required to pass the hadronic leakage stage (here one can go tighter and require to pass the Middle sampling cuts). For this sample we must optimize the normalized track impact parameter cut.

- We accept candidates with 4 isolated electrons requiring at least 3 of these electrons to have an associated matched track in the ID. For this sample we use the
standard normalized track impact parameter cut.

The resulting invariant mass distribution for $H \to 4e$ is shown in figure 24, where a dramatic increase in the signal efficiency is observed ($\sim 50\%$). A slight deterioration of the mass resolution is observed due to inclusion of softer electrons.

![Graph](image)

Figure 24: $H \to 4e$ invariant mass before (upper plot) and after (lower plot) application of EM object classification. There is a large gain in signal and a deterioration of the resolution coming from events with softer electrons.

The Higgs→4e final results in terms of significance are summarized in tables 3, 4 and should be compared directly with the standard analysis results presented in [20]. In table 3 a fixed normalized impact parameter cut of $10\sigma$ was used. In table 4 the impact parameter cut for the 3+1 sample was tightened to $7\sigma$ reducing the ZZZ background and demonstrating the power of the impact parameter cut optimization in such algorithms.
Table 2: \textit{Higgs\textrightarrow{4\ell} TDR-based Analysis Summary}

<table>
<thead>
<tr>
<th>Higgs Mass (GeV)</th>
<th>130</th>
<th>150</th>
<th>180</th>
<th>200</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross section×BR (fb)</td>
<td>0.903</td>
<td>1.5346</td>
<td>0.7844</td>
<td>2.9812</td>
<td>1.8813</td>
</tr>
<tr>
<td>Signal (fb)</td>
<td>0.0858</td>
<td>0.1999</td>
<td>0.1271</td>
<td>0.5914</td>
<td>0.3424</td>
</tr>
<tr>
<td>pp\rightarrow ZZ\text{} \rightarrow 4\ell (fb)</td>
<td>0.0209</td>
<td>0.0154</td>
<td>0.0789</td>
<td>0.2244</td>
<td>0.0940</td>
</tr>
<tr>
<td>pp\rightarrow tt\rightarrow 4e (fb)</td>
<td>0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>gg\rightarrow Zb\bar{b} \rightarrow 4e (fb)</td>
<td>0.0094</td>
<td>0.0036</td>
<td>0.0014</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4\times30 fb\text{}^{-1} Signal</td>
<td>10.296</td>
<td>23.988</td>
<td>15.252</td>
<td>70.968</td>
<td>41.088</td>
</tr>
<tr>
<td>4\times30 fb\text{}^{-1} Bgdnd</td>
<td>4.236</td>
<td>2.28</td>
<td>9.636</td>
<td>26.928</td>
<td>11.28</td>
</tr>
<tr>
<td>Significance</td>
<td>3.64</td>
<td>8.82</td>
<td>4.04</td>
<td>10.50</td>
<td>8.75</td>
</tr>
</tbody>
</table>

Table 3: \textit{Higgs\textrightarrow{4\ell} Optimized Analysis Summary: a fixed normalized impact parameter cut of 10\sigma was used for all masses.}

<table>
<thead>
<tr>
<th>Higgs Mass (GeV)</th>
<th>130</th>
<th>150</th>
<th>180</th>
<th>200</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross section×BR (fb)</td>
<td>0.903</td>
<td>1.5346</td>
<td>0.7844</td>
<td>2.9812</td>
<td>1.8813</td>
</tr>
<tr>
<td>Signal (fb)</td>
<td>0.1294</td>
<td>0.2813</td>
<td>0.1719</td>
<td>0.8132</td>
<td>0.4872</td>
</tr>
<tr>
<td>pp\rightarrow ZZ\text{} \rightarrow 4\ell (fb)</td>
<td>0.033</td>
<td>0.0268</td>
<td>0.1146</td>
<td>0.2916</td>
<td>0.1355</td>
</tr>
<tr>
<td>pp\rightarrow tt\rightarrow 4e (fb)</td>
<td>0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>gg\rightarrow Zb\bar{b} \rightarrow 4e (fb)</td>
<td>0.0246</td>
<td>0.0137</td>
<td>0.0058</td>
<td>0.0087</td>
<td>0.0051</td>
</tr>
<tr>
<td>4\times30 fb\text{}^{-1} Signal</td>
<td>15.528</td>
<td>33.756</td>
<td>20.628</td>
<td>97.584</td>
<td>58.464</td>
</tr>
<tr>
<td>4\times30 fb\text{}^{-1} Bgdnd</td>
<td>7.512</td>
<td>4.86</td>
<td>14.448</td>
<td>36.036</td>
<td>16.872</td>
</tr>
<tr>
<td>Significance</td>
<td>4.45</td>
<td>9.43</td>
<td>4.5</td>
<td>12.36</td>
<td>10.33</td>
</tr>
</tbody>
</table>

The gain in significance is better shown in figure 25 where it should be noted that in the low masses our method provides a dramatic increase in the signal yield which is expected to aid analyses in which the signal is reconstructed by a fit and not by simple event counting.

The excessive Zb\bar{b} background may be reduced if an impact parameter cut is used throughout the mass region. In particular, decrease of the Zb\bar{b} and tt background may be achieved for ZZ+Jet events when the ZZ candidate is required to be back-to-back with the recoiling jet. In this case radiation from the b lines fills the space between the jet and the ZZ system [20].

5.3 Caveats and Open Questions

Soft electron (p_{\perp} < 20 GeV) isolation is currently being optimized in ATLAS. The isolation criteria used in this section and described in the Appendix, have been optimized for electrons with p_{\perp} > 20 GeV. As a result the evaluation of the “3+1” method in the case where the fourth non-isolated electron is soft, is questionable. It is clear though
Figure 25: Higgs Significance obtained from a $H \rightarrow ZZ^{(*)} \rightarrow e^+e^-e^+e^-$ analysis with full ATLAS simulation and reconstruction. Classification of EM clusters using shower shape information, improves not only the significance shown in the Figure, but also the signal yield by $\simeq 40 - 50\%$ at the low Higgs mass region.
Table 4: Higgs→4e Optimized Analysis Summary: a tighter normalized impact parameter cut of 7σ was used for the 3+1 sample; the standard impact parameter cut of 10σ was used for the 4 isolated electron sample.

<table>
<thead>
<tr>
<th>Higgs Mass (GeV)</th>
<th>130</th>
<th>150</th>
<th>180</th>
<th>200</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross section×BR (fb)</td>
<td>0.903</td>
<td>1.5346</td>
<td>0.7844</td>
<td>2.9812</td>
<td>1.8813</td>
</tr>
<tr>
<td>Signal (fb)</td>
<td>0.1242</td>
<td>0.2723</td>
<td>0.1668</td>
<td>0.7830</td>
<td>0.4718</td>
</tr>
<tr>
<td>pp→ZZ(*)→4e (fb)</td>
<td>0.0312</td>
<td>0.0261</td>
<td>0.1094</td>
<td>0.2772</td>
<td>0.1300</td>
</tr>
<tr>
<td>pp→tT→4e (fb)</td>
<td>0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>gg→Zbb→4e (fb)</td>
<td>0.0203</td>
<td>0.0101</td>
<td>0.0058</td>
<td>0.0065</td>
<td>0.0029</td>
</tr>
<tr>
<td>4 × 30 fb⁻¹ Signal</td>
<td>14.904</td>
<td>32.676</td>
<td>20.016</td>
<td>93.96</td>
<td>56.616</td>
</tr>
<tr>
<td>4 × 30 fb⁻¹ Bgdnd</td>
<td>6.78</td>
<td>4.344</td>
<td>13.824</td>
<td>34.644</td>
<td>16.68</td>
</tr>
<tr>
<td>Significance</td>
<td>4.54</td>
<td>9.59</td>
<td>4.5</td>
<td>12.26</td>
<td>10.10</td>
</tr>
</tbody>
</table>

that in this case the method may still apply and produce gains in the context of a soft-electron selection criteria. This possibility is under study.

A second open question is the one of the level of Zbb background. There are large theoretical uncertainties in its prediction and the results may change significantly.

A third question is on the treatment of the method for the extraction of the signal significance. Relaxing cuts which allow a better handle of a smooth background may allow for use of “sideband” information which may also change the results presented in this section.
6 Summary and Conclusions

In this note an electron-based calculation of longitudinal weights for the ATLAS Electromagnetic Barrel Calorimeter was performed. The improvement in resolution and linearity for electrons of energies ranging from 10 GeV up to the TeV scale was demonstrated.

A new analysis method was proposed which can be applied in multi-lepton final states where detector level information is exploited to discriminate between signal and background. The method was specifically applied to the \( H \rightarrow 4e \) channel where EM shower shape information was used to discriminate between signal and \( t\bar{t} \), \( Z\nu\nu \) reducible backgrounds. Gains of \( \sim 40 - 50\% \) in signal and \( \sim 10 - 20\% \) in significance were found for low Higgs masses for this particular channel. Further suppression of the \( Z\nu\nu \) background using a more refined track impact parameter determination is expected to lead to larger gains. However, there are some serious open questions concerning the non-optimized soft electron (\( p_\perp < 20 \) GeV) isolation and the level of the \( Z\nu\nu \) background which may affect the results presented here. These issues are under investigation. Application of this general method to other channels with multi-lepton final states is possible but it requires analysis specific algorithms.

Concerning the near future a full calibration of the EMC for \(|\eta| < 2.5\) will be performed within the ATLAS electron-photon group with DC2 samples for offline releases starting from 9.0.0.

7 Acknowledgements

The authors are grateful to Tancredi Carli, Guillaume Unal, Leonardo Carminati, Martina Schaefer, Kyle Kranmer, Monika Wieler, Michael Riveline, Steve Armstrong, Alden Stradling and Karina Loureiro for their invaluable input in this work. Special thanks to Dirk Zerwas for pointing out a problem in the ID-CALO energy combination formalism.
Appendix

A  Electron Isolation

In this analysis we apply the standard LAr Barrel and End-Cap (EMB and EMEC Outer Wheel) calorimeter cuts on shower shapes.

A calorimeter cluster is tagged as an electron/photon like based on the following application of a sequence of (mostly isolation) cuts which are described in this section:

1. Cuts on longitudinal energy leakage of the EM shower in the Hadronic Calorimeter.
2. Cuts on shower shape variables based on the second sampling (a.k.a. middle or S2) of the EMB/EMEC.
3. Cuts on shower shape variables based on the first sampling (a.k.a. strips or S1) of the EMB/EMEC.

The variables used are taken from [21] and described briefly here:

- $\eta_2$: pseudorapidity of cluster extracted from the 2nd sampling.
- $e_{l}^{had}$: $E_{\perp}$ shower leakage into the 1st sampling of the hadronic calorimeter (HAC).
- $e_{237}/e_{277}$: ratio of cluster energy in $3 \times 3$ over $3 \times 7$ cells.
- $w_{1c}$: corrected cluster width in 3 strips in the 1st sampling.
- $w_{2c}$: corrected cluster width in $3 \times 5$ cells in the 2nd sampling.
- $ema_{x2}$: second maximum in 1st sampling (strips).
- $e_{37}$: $E_{\perp}$ of the cluster based on a $3 \times 7$ size.
- $emin$: energy of strip with a minimum between the two maxima.
- $w_{tot}$: total width in 1st sampling in 20 strips.
- $fracm$: energy in strips outside core ($(E(\pm7)-E(\pm3))/E(\pm7)$.

For reference the low luminosity part of the actual $gejflag$ method in ATHENA is given below:


//
// cuts on hadronic energy
//
if (fabs(eta2) < 0.8) {
    if (ethad/1_et_37>3.5*(0.01046-0.0121*fabs(eta2)+0.002*eta2*eta2))
        iflag = 2;
} else if (fabs(eta2) >= 0.8 && fabs(eta2) < 1.5) {
    if (ethad/1_et_37>0.008) iflag = 2;
} else if (fabs(eta2) >= 1.5 && fabs(eta2) < 1.8) {
    if (ethad/1_et_37>0.03) iflag = 2;
} else if (fabs(eta2) >= 1.8 && fabs(eta2) < 2.0) {
    if (ethad/1_et_37>0.02) iflag = 2;
} else if (fabs(eta2) >= 2.0 && fabs(eta2) < 2.47) {
    if (ethad/1_et_37>0.015) iflag = 2;
}
if (iflag == 2) return iflag;

// cuts on 2nd sampling
if (e277<0.) {
    iflag=3 ;
    return iflag;
}
float wgt=1.0;
if (fabs(eta2) < 0.8) {
    if (e237/e277<=0.915*wgt) iflag = 3;
    if (weta2c>0.012) iflag = 3;
} else if (fabs(eta2) >= 0.8 && fabs(eta2) < 1.5) {
    if (e237/e277<=0.910*wgt) iflag = 3;
    if (weta2c>0.012) iflag = 3;
} else if (fabs(eta2) >= 1.5 && fabs(eta2) < 1.8) {
    if (e237/e277<=0.89*wgt) iflag = 3;
    if (weta2c>0.012) iflag = 3;
} else if (fabs(eta2) >= 1.8 && fabs(eta2) < 2.0) {
    if (e237/e277<=0.92*wgt) iflag = 3;
    if (weta2c>0.0115) iflag = 3;
} else if (fabs(eta2) >= 2.0 && fabs(eta2) < 2.47) {
    if (e237/e277<=0.91*wgt) iflag = 3;
    if (weta2c>0.0125) iflag = 3;
}
if (iflag == 3) return iflag;

// cuts on 1st sampling

32
if (f1<0.005 || fabs(eta2)>=2.37) return iflag;
if (fabs(eta2) < 0.8) {
    if (emax2>0 && (emax2/(1.+0.009*1_et_37)>=0.25 ||
            (emx2-emin)>=0.15)) iflag = 4;
    if (wtot>=2.7 || fracm >= 0.35 || weta1c >= .75) iflag = 4;
} else if (fabs(eta2) >= 0.8 && fabs(eta2) < 1.37) {
    if (emax2>0 && (emax2/(1.+0.009*1_et_37)>=0.5 ||
            (emx2-emin)>=0.15)) iflag = 4;
    if (wtot>=3.5 || fracm >= 0.6 ||
        weta1c >= .75) iflag = 4;
} else if (fabs(eta2) >= 1.52 && fabs(eta2) < 1.8) {
    if (emax2>0 && (emax2/(1.+0.009*1_et_37)>=1.1 ||
            (emx2-emin)>=0.35)) iflag = 4;
    if (wtot>=3.5 || fracm >= 0.68 ||
        weta1c >= .8) iflag = 4;
} else if (fabs(eta2) >= 1.8 && fabs(eta2) < 2.0) {
    if (emax2>0 && (emax2/(1.+0.009*1_et_37)>=0.40 ||
            (emx2-emin)>=0.2)) iflag = 4;
    if (wtot>=2.0 || fracm >= 0.30 || weta1c >= .7) iflag = 4;
} else if (fabs(eta2) >= 2.0 && fabs(eta2) < 2.35) {
    if (emax2>0 && (emax2/(1.+0.009*1_et_37)>=0.3 ||
            (emx2-emin)>=0.15)) iflag = 4;
    if (wtot>=1.4 || fracm >= 0.20 || weta1c >= .6) iflag = 4;
}
if (iflag == 4) return iflag;
References


[8] New parametrization first proposed and studied by Tancredi Carli (CERN) in 2003 (private communication).


[13] The toy Geant4 MC and the sample generation was performed by Tancredi Carli (CERN) in 2003 (private communication).

[15] Samples provided by Leonardo Carminati (Torino) in Feb/2004 (private communication).


