ATLAS-COM-PHYS-2009-424

Report of the Review Panel of the Hadronic Calibration Workshop (Foz do Arelho 2009)

M. Bosman[1], R. Kehoe[2], M. Lefebvre[3], H. Oberlack[4], M. Shapiro[5], K. Tokushuku[6]

[1]Institut de Fisica d'Altes Energies, Barcelona, Spain

[2]Southern Methodist University, Dallas, Texas, USA

[3]University of Victoria, Victoria, British Columbia, Canada

[4]Max Planck Institut fur Physik, Munich, Germany

[5]Lawrence Berkeley National Laboratory, Berkeley, California, USA

[6]High Energy Accelerator Research Organization, KEK, Tsukuba, Japan

August 11, 2009

Abstract

This note presents the results of the review panel for the 2009 Hadronic Calibration Workshop in Foz do Arelho, Portugal. The workshop covered six primary topics: detector performance and Monte Carlo, local and global hadronic calibration, jet energy scale corrections, systematic uncertainties, $\not{E}_T$ calibration and elements necessary to perform and use the calibration during the data-taking period. We summarize key points for each of these topics, and provide subsequent comments and recommendations.

1 Test Beam and Detector Performance

The calorimeter groups have devoted a major effort to the acquisition and analysis of test beam data, both from individual detectors and as part of the combined test beam. The results of this analysis provide an impressive level of precision over a large range of particle types, energies and pseudorapidities. Detailed studies of a large number of distributions that are sensitive to the longitudinal and transverse shower development, as well as the overall energy calibration and resolution have been made. The test beam data is the touchstone for the first-principles understanding of jet energy calibrations and a resource that will be used throughout the time ATLAS is running.

Recommendation:

It is very important that test beam data be stored in a format that will be accessible and analyzable for the lifetime of ATLAS. The most appropriate such format is the ESD. In order to obtain support for this activity, putting the test beam data into ESD format should be identified as an OTP task

2 Monte Carlo Tuning and Detector Description

Detailed comparisons of test beam data with a number of GEANT4 physics lists have been performed. None of the existing lists agree perfectly with the data, but the "best" of these lists agree within a few per cent. In addition, the available lists bracket the data. Therefore running with different lists provides a means of assessing the systematic uncertainties on the jet calibration. A particular difficulty is present in the region 9 to 25 GeV where there is little teat-beam data. For example, the *QGSP_BERT* list which ATLAS uses as a default has unphysical discontinuities intrinsic to the model. This causes a discontinuity in the mean response as a function of beam energy in this region. Fixing this problem will not happen quickly and ATLAS will have to use the existing physics lists at least for the first run.

Recommendation:

Continued collaboration with the GEANT4 team is essential. For the current run, simulation studies with different physics lists are important for understanding the systematic uncertainties on the jet energy scale. For future runs, an improved physics list with more physical performance in the energy range between 9 and 25 GeV is needed.

The lowest energy available in the test beam was 3 GeV. To extend the experimental measurement of the calorimeter response to lower energies, and to verify that the response of hadrons in the final experimental configuration agrees with the test beam, ATLAS will need to study the *in situ* response to isolated hadrons. These studies are complicated by the need to correct the *in situ* data for noise, pileup and physics backgrounds (from nearby particles). Promising work has been done in this area and is included in the ATLAS performance document. Detailed comparisons between the *in situ* response and the test beam data for those energies where both are available will provide an important cross-check on the calorimeter calibrations.

Recommendation:

In situ measurements of the single hadron response are an important component of the strategy for validating the jet energy calibration. These studies are difficult and will require significant effort from the collaboration. The use of the HLT track trigger is necessary to extend the energy range accessible for these studies. The jet performance group must evaluate how much data is necessary *and ensure that these data are available on DPDs.*

3 Local and Global Calibration Schemes

Jet finding in ATLAS uses either towers or topoclusters as its input. The topoclusters have an advantage over towers in that they incorporate sophisticated noise-suppression. Three schemes for calibrating jets have been discussed: EM-scale, global cell-weighting (GCW), and Local Hadron (LC) calibration. EM-scale calibrations are the simplest since this scheme does not attempt to correct for the intrinsic e/h of the calorimeters. Because it is simple, it will play an important role during data commissioning and perhaps for early analysis. Both the GCW and LC schemes use the fact that electromagnetic and hadronic interactions result in different energy densities (defined as the energy deposited per unit volume) and longitudinal depth. The GCW scheme performs cell level corrections that depend on the energy density of that cell in the global context of a jet. The LC is derived from single pions and the approach classifies each topocluster as EM-like or hadron-like (based on energy density) and applies a calibration to the cluster. At first glance, the performance of the GCW and LC schemes appears similar; but comprehensive comparisons have just begun.

Because roughly 1/3 of the jet energy is deposited by π^0s, jet energy calibration requires a good EM calibration as a starting point. Providing the calibration for electrons and photons is the responsibility of the e-gamma performance group. However, challenges remain for appropriately using these calibrations for jets. First, photon and electron reconstruction is done using a sliding window cluster algorithm rather than the topoclusters used in the jet group. Out of cluster corrections will be different in the two cases. Second, the e-gamma group makes a number of corrections to the energy scale that are appropriate for isolated photons and electrons but are not applicable to jets. It will be important to properly factor absolute scale, non-uniformity, and dead-material corrections such that they can be implemented at cell level for jets and $\not{E}_T$.

Recommendations:

- *The jet performance group and the e-gamma group should collaborate further to develop a strategy to propagate improvements in the EM calibration, which are not specific to the sliding window algorithm, into the cell level for use by topoclusters, jets and* $\not{E}_T$*.*
- *We believe the establishment of the EM-scale in the forward (*$|\eta| > 2.5$*) region is important for jet and* $\not{E}_T$ *calibration. This will involve the implementation of calibration studies in* Z *events with one electron in this region. This strategy will likely require manpower contributed from the jet group.*

Both the GCW and LC calibration schemes have had an impressive amount of work done to study their performance. The panel strongly supports this work. A major achievement of the past year is the development of debugged and easy-to-use tools to study jet and $\not{E}_T$ performance using these schemes. These include

ParticleID (which allows systematic studies of how the energy deposition for individual particles affects the overall energy distribution), JetPerformance and MetPerformance (which provide an overall framework for evaluating and monitoring performance) and JetTools (which insures that many users can calculate complex quantities in a consistent way and also allows users to easily switch between different jet reconstruction, calibration or selection algorithms). This is a major achievement of the group and should be congratulated.

Recommendation:

The jet group should develop a set of benchmarks to compare the performance of different calibration strategies. Common software and samples must be used for these comparisons.

During the first run, the ATLAS operating conditions will be changing. The number of bunches in the collider, the number and distribution of dead channels, the noise and the amount of pileup will vary significantly. While ATLAS has run static simulations varying each of these factors individually, it is less clear whether the mechanisms to track such changes dynamically and apply appropriate calibration corrections are in place.

Recommendation:

Common analysis tools will be necessary to propogate time-dependent effects into the jet calibration framework. A strategy for handling time dependent effects in the simulation should be developed.

4 Jet Corrections

The Jet/$\not\!\!E_T$ group has an impressive array of studies ongoing which are designed to establish the jet energy scale. Some of these are directed at specific effects, while others attempt to achieve a total correction. In many cases, a decision concerning what methods to use is best evaluated when the first data arrive. As a result, the Jet/$\not\!\!E_T$ group has developed and presented during the Workshop a flexible, factorized approach to implementing the jet energy scale. Such an approach considers the EM scale as a starting point, with the hadronic calibration (LC or GCW) to follow. Then there is allowance for offset (e.g. pileup) and response corrections to ensure the jet reflects the desired reference: the energy in a jet defined by the same algorithm but applied at particle level. Subsequent adjustments to the jet energy based on layer fractions or tracking might be used to further improve jet energy resolution.

The review panel views this development as critical. It allows the swapping of different approaches for different calibration components (e.g. hadronic calibration, jet response) for excellent flexibility for the data-taking period. We have several more specific comments or outstanding questions:

- The JetCalibTools package is very useful to provide software for the delivery of the calibration.

- It was not entirely clear how the order of corrections would be chosen and implemented. For instance, does the performance (e.g. systematic uncertainty) of the calibration suffer from certain choices? Also, what are the implications for applying particular corrections first on the derivation of downstream corrections?

- How does the Jet/$\not\!\!E_T$ group expect to ensure that effects are not double-counted in the correction?
- Establishing systematics for the calibration is more important than improving the resolution.

Recommendations:

- *It is important to keep detector and physics effects separate. There are several effects which alter a parton level jet to a particle level jet. Such effects are specific to the physics sample with which a physics analysis is done. We encourage not to include fragmentation or radiation out-of-cone, or underlying event (i.e. any interactions involving the pp giving the hard interaction). We recommend that these effects be kept out of the jet energy scale as much as possible.*
- *It is important to keep corrections relevant to jets distinct from those relevant for $\not\!\!E_T$. For instance, jets may have offset, response and showering or bending corrections. But the event $\not\!\!E_T$ should only be corrected for the deficit of response from unity, i.e. the signal not measured by the detector.*
- *We suggest to develop more precision concerning key concepts of the correction. There appeared to be several very different definitions of response, out-of-cone, showering, etc. among the discussed analyses.*

5 Noise and Jet Offset Studies

Noise is a concern in the jet reconstruction, particularly after work which revealed a susceptibility of the tower-based cone algorithm to negative energy from noise. This susceptibility essentially arises because of the four-momentum representation used for calorimeter readout channels, which cannot properly handle these energies. In some cases, this has a noticeable impact on the reconstructed jet properties. A proposal was made by the convenors to drop the Tower-based jet collection and use instead one based on TopoTowers. These are towers whose constituents come solely from cells incorporated into topoClusters using $4-2-0$ thresholding. Substantial improvement in the stability of jets was exhibited.

This session also saw the discussion of pileup corrections for jets. In performing a closure test for these studies, however, it was observed that the correction does not quite remove the effect.

Recommendations:

- *The availability of the TopoTower algorithm for analysis may be valuable. Nevertheless, we feel there is also a need for an alternative algorithm for* *validation of TopoClusters-based jets, and this motivates a retention of the Tower-based jets. If necessary to render the algorithm useful, then we support a modification to improve noise suppression or circumvent the problem arising from the four-momentum representation of cell energies.*

- *The pileup closure plot appears to suffer from an inconsistency in the tower usage between jets and the pileup calculation. We encourage this to be resolved soon.*

6 *in situ* Balancing Methods

In order to establish the overall energy scale of jets, either as a calibration or validation after the cell-level hadronic scale, or directly as a calibration from the EM-scale, several studies are being pursued which use p_T comparisons to obtain absolute or relative calibrations for jets. These analyses will use data and MC samples with γ+jet, Z+jet and dijets events. This effort is getting a large emphasis now in the Jet/$\not\!\!E_T$ group, and is employing both $\not\!\!E_T$ and p_T balance approaches. The panel strongly supports this effort.

The primary studies shown used the γ+jets samples. Many of the results indicated satisfactory balance or closure plots after correction. However, there was debate about the best way to do the 'numerical inversion' in these samples, as that appeared to generate a bias at low p_T. Also, a method using the Missing E_T Projection Fraction (MPF) indicated a small undercorrection at high E_T, as well as a small overcorrection of order 2% for GCW and LC calibrated jets.

We have the following comments:

- Most of the methods being pursued employ a direct p_T balance of the jet with the photon. At the physics level, this attempts to equate the jet with the total hadronic final state balancing the photon, and this includes ISR/FSR as well as fragmentation effects. At the detector level, these direct p_T balancing methods also include differences in offset, showering and bending of particles which differentiate photons from jets. These effects deal with energy observed by the calorimeter, and therefore should not be part of a correction propagated into the $\not\!\!E_T$. Thus, such methods have difficulty keeping these two effects separated.

- Why are GCW and LC calibrated jets being overcorrected in the MPF analysis? Discussions during the workshop suggested a potential double-counting was occurring, at least for the GCW jets. This is an outstanding question we would like to see resolved.

Recommendations:

- *Studies of the sensitivity of the 'numerical inversion' method to fit function assumptions and binning should be performed. A fit-less inversion method, such as using E', should be attempted.*

- *We encourage the exploration of alternative fit functions for jet response vs. jet energy.*

- *We strongly support the development of a basic toy Monte Carlo for these events. Such a simulation should incorporate an adjustable p_T spectrum for the physics sample, resolutions for photons and jets, varying levels of non-compensation in the calorimeter, and eta-dependence of the response. Additional inclusion of simple ISR/FSR models with relevant adjustment of the event kinematics, showering and offset effects will also be useful. Such a simulation can provide a clear validation of the various methods.*

- *In addition to the barrel reference region, we support adding a second reference in the crack-free end-cap region near $|\eta| = 2.2$.*

- *We recommend that Monte Carlo samples be requested of sufficient size to cover the p_T range expected in early data. The most critical samples are dijet, Z+jet and γ+jet samples for these analyses.*

- *We support generation of these samples with different GEANT physics lists, and different generators/fragmentation models. The in situ analyses described do not incur systematic uncertainties from these effects because residual miscalibrations in the data and Monte Carlo would be measured and corrected separately. A mismatch between the reference provided in data and Monte Carlo can occur, for instance from the presence of backgrounds in data, but this motivates a data-driven study of the impact of the background. However, a variety of Monte Carlo samples will still be informative when comparing results of different calibration methods, or understanding the effect of combining different calibration components.*

- *We encourage trigger and background studies for the physics samples used in these methods.*

7 Jet Flavor and Resolution

An analysis was presented in which the ratio of reconstructed p_T to quark p_T was plotted vs. quark p_T. This analysis indicated a several percent difference in this ratio at low p_T (< 25 GeV), and a small difference at high p_T for jets of different flavors. The difference was somewhat smaller at high p_T for the hadronically calibrated jets than for the EM-scale calibrated jets. We identify these open questions or recommendations:

- The comparison of reconstructed jet to quark amounts to a parton-level comparison. In this situation, the semileptonic decay for heavy flavor is included in the flavor differences observed. A correction for this effect can be obtained directly in those cases where the lepton was identified. A comparison to particle-level jets in which the muon and neutrino are also omitted would help to isolate the fragmentation differences from the semileptonic decay differences. We would be interested in a clear itemization of these two effects separately.

- The size of the difference is comparable to the 5-10% systematic that ATLAS hopes to achieve in early data. The details of the flavor dependence of the energy scale are complex and tied somewhat to the physics sample being studied. It may take some time to identify the detector-only differences between jets of various flavors. It will also necessitate some cross-checks of the Monte Carlo with data. In early data, the emphasis should be on larger effects to establish a valid energy scale at the 5-10% calibration – a prerequisite before addressing the question of flavor-dependent corrections. A systematic uncertainty can be defined and determined from early studies. This will apparently have a significant impact at low E_T. Once the effects are understood, jets identified to be of specific flavors, such as $b-$jets in a top quark analysis, should have appropriate corrections.

Taking the approach of factorization one step further, an analysis has been pursued which measures the functional dependence of jet response on various layer energies. The motivations for the layers chosen is actually taken from the GCW approach, but a simpler analysis is performed. Using jets at the EM scale, the layer dependence of response is fitted and corrected for. Such a correction could be performed after all other jet-level corrections meant to attain the particle-jet energy scale for the jet on average. This particular calibration would therefore be dedicated to improving the jet resolution. This simple method does appear to achieve a similar jet energy resolution to the GCW approach. Dijet events are used, meaning that this correction can be independently obtained *in situ* for the ATLAS detector. We support continuing to explore this analysis.

Recommendations:

- *It should be investigated whether response and resolution decouple. Applying multiple corrections on top of each other, especially if they are correlated, can create potential biases. How is it proposed to ensure this does not happen?*

- *It should be demonstrated that the trigger biases in this sample are understood. The triggers can bias the shower properties meaning that in the data, it could be difficult to extract an unbiased correction.*

8 Systematic Uncertainties

Attaining the desired particle jet energy scale is important, but a necessary corollary is that there must be well-understood systematic uncertainties. In the period leading to the Workshop, the panel has requested some roadmap for the development of such uncertainties for the various methods chosen. We believe it is important that studies strive to compare methods with systematics. A partial list was presented and discussed at the Workshop. Although the panel does not anticipate recommending a complete set of systematic studies for the jet/$\not\!\!E_T$ group, we identify the following open questions:

- For those data sets with only one GEANT4 version available, how will detector simulation systematics be assessed?
- For almost all calibration methods, what is the impact of pileup on the calibration?
- For hadronic calibration methods (LC, GCW), what are the dead-material and coherent noise uncertainties? What is the out-of-cone uncertainty for GCW?
- For approaches extracting E/p from data and using Monte Carlo, there are two categories of question.
  - On the Monte Carlo side, what is the size of the effect for fragmentation model, calorimeter response model, shower profiles and ISR/FSR model on the jet energy scale?
  - For the E/p measurement, how is this affected by track efficiency uncertainty, resolutions, and backgrounds from noise and physics?
- For the γ/Z+jet approaches, what are the impacts of ISR, event topology, photon backgrounds, and mis-vertexing?

Recommendation:

- *We believe this category of issue is critical to the calibration effort, and strongly encourage the Jet/$\not\!\!E_T$ group to develop a clear road map to assess systematics.*

9 Cosmic Cleanup, Cell Masking, Data Quality and Monitoring

Cosmic ray shower muons interact in the ATLAS detector, and their overlap with proton-proton collision events is a potential source of background for searches of physics beyond the Standard Model through large $\not\!\!E_T$ signatures. It is therefore important to develop strategies and tools to identify, and possibly reject, events with large $\not\!\!E_T$ potentially caused by high energy cosmic muons. Relevant observables include calorimeter cluster time, calorimeter shower shape variables (including cluster EM fraction), jet EM fraction, number of tracks associated to a jet, and the fraction of transverse energy in a jet associated to tracks. They appear to be well described by Monte Carlo. Clearly, the use of such observables for event rejection is physics analysis dependent.

Recommendation:

A coordinated effort is required to develop a software tool to identify potential cosmic muon contaminations in events with large $\not\!\!E_T$. The observables and *criteria used should be clearly defined, so that the use of the software tool can be understood in the context of various physics analyses.*

If calorimeter cell data (or other objects not normally on AOD) are required to identify cosmic muons, then data samples should be prepared as DPD with the relevant information. If this is not possible, a mechanism should be found to make the outcomes of a cosmic muon finder tool available on AOD.

Calorimeter cells identified as misbehaving are contained in 'bad channels' data bases. They are typically labelled as "dead", "noisy", or "affected". For LAr cells, each of these categories is further divided in various sub-categories. Software tools have been developed to allow the masking of known bad channels. Some "noisy" LAr channels are now masked for data reprocessing, and the effects of this masking is being monitored.

Recommendations:

- *The cell masking strategy for HLT needs to be defined.*
- *The cell masking strategy for LVL1, HLT and offline reconstruction should be made consistent.*

The data quality monitoring (DQM) of jets and $\not{E}_T$ is well developed. Initially, the corresponding DQM flag (green, yellow, red) is set to the detector flag. Clearly, close contact with the LAr and Tile detector groups and with the e/gamma performance group is critical. The relation between the size of the dead region (coverage) and the DQM flag is still an issue.

For both cell masking and DQM development, it is important to gain experience with data taking; the cosmic data taking period starting during this workshop is a good opportunity.

Recommendations:

- *For cell masking and for jets and $\not{E}_T$ DQM, the performance from the current cosmic run should be compared to the performance from last year's cosmic run. They should be reviewed prior to collision data taking.*
- *In addition to the jet/$\not{E}_T$ monitoring, and the LAr/Tile specific monitoring, we support the inclusion of the CaloMonitoring outputs for Jet/$\not{E}_T$ shifters and data quality assessment. This includes cells, clusters and towers which are the direct constituents of jets and $\not{E}_T$.*

10 *in situ* $\not{E}_T$ Calibration

Details of the current $\not{E}_T$ calculations and calibration strategy were presented. A roadmap for $\not{E}_T$ calibration for the first collisions up to 100 pb^{-1} was developed including the use of minimum bias events, di-jet events, W/Z samples, and eventually $t\bar{t}$ events. Furthermore, the METPerformance package offers a common set of tools for the study of $\not\!\!E_T$. All this should be considered as great achievements indeed. We note that more thorough documentation is required in order to address $\not\!\!E_T$ concerns such as double counting of energy, and consistent handling of jets, electrons and muons.

11 DPDs and Triggering

The many requirements on DPDs were discussed. One of the requirements is that the data volume of DPDs should be at the level of 1% to 2% of the ESD data volume. We note that one of the most critical issues in designing the DPD content is the knowledge of the number of events needed to reach a given desired level of calibration accuracy. When this number is known, one can then argue between a larger prescale (and likely changes in data taking conditions) or a smaller (or no) prescale (and data taking in one day with fixed beam and detector conditions). It was noted that low prescale values for special runs could be formally requested to the trigger group and to the run coordinator.

The effect of trigger bias on the DPD samples used for calibration purposes should be checked carefully. Some ideas emerged from the workshop, including

- no requirements on jets for γ+jet studies;
- no calorimeter trigger requirements for E/p studies, for example minimum bias trigger and offline (or HLT) tracking requirements;
- no calorimeter trigger requirements for low E_T jets, for example minimum bias trigger and offline reconstructed jets (HLT?) requirements.

Different event samples are required for the purpose of hadronic calibration, including various jet samples, γ+jet, high p_T tracks, W/Z, and perhaps high p_T photons, and forward jet samples. It is not clear whether full ESD content is required for all samples. We note that it may be more efficient to identify one person responsible for each sample.

Recommendation:

A complete list of the samples required for calibration should be finalized, along with their data content. The related request for trigger and DPD production should be made as soon as possible.

12 Towards One Reference Calibration

The complexity of the jet and $\not\!\!E_T$ calibration requires a roadmap which attempts to balance the need for a calibration with known, but perhaps large, systematic uncertainties in early data-taking with the desire for optimum performance with later data. As a result, the convenors have presented a set of guidelines and milestones which seek to achieve a single reference calibration during the run.

This roadmap allows for the possibility of other calibrations to be developed in parallel, and ultimately to be used in specialized circumstances.

We have the following comments to the roadmap presented:

- A timeline was proposed in which the goal of achieved systematic uncertainty was staged from 10% in very early data to 1% with 1 fb^{-1}. We view this timeline as critical because it will allow the Jet/$\not{E}_T$ group to address large questions first before working on smaller details later. The luminosity milestones appear to be optimistic, but this is a good goal.

- Specific, somewhat hypothetical milestones were described by the convenors to lead the discussion on the process to a reference calibration. We support this kind of thinking ahead, and have several comments about the first milestone. This milestone suggests that if there are problems in early data taking in the Monte Carlo effort, then the γ+jet methods should be used to bring jets from the EM-scale to the particle-jet scale.

  We see two open questions relevant to this proposal. Does such a calibration cause an excessive systematic uncertainty? Will the necessary systematic studies be completed in time? Beyond these, we came to three specific conclusions. A decision on whether to take this route should await the first data-taking. There should be a pre-agreed set of parameters to make a decision, which includes systematic uncertainties for jets as well as established performance metrics to gauge methods on a common footing. Provided these conditions are met, it is possible that the γ+jet sample could provide the calibration from the EM scale in early data.

Recommendations:

- *There was a schematic showing which calibration algorithms whould be provided for which reconstruction algorithms. By default, LC would only be available for small cone size jets. We recommend that this calibration also be included for the 'large' cone size.*

- *We recommend that the γ+jet can provide a potential residual correction if necessary after the GCW or LC calibrations.*