

THE CALORIMETER EVENT DATA MODEL IN ATLAS

W. Lampl*, P. Loch, University of Arizona, Tucson, USA,
H. Ma, S. Rajogopalan, Brookhaven National Laboratory, Upton, New York, USA,
S. Laplace, LAPP Annecy, Annecy-Le-Vieux, France, S. Menke, MPI, Munich, Germany,
D. Rousseau, LAL Orsay, Orsay, France

Abstract

The event data model for the ATLAS calorimeters in the reconstruction software is described, starting from the raw data to the analysis domain calorimeter data. The data model includes important features like compression strategies with insignificant loss of signal precision, flexible and configurable data content for high level reconstruction objects, and backward navigation from the analysis data at the highest extraction level to the full event data. The most important underlying strategies will be discussed in this contribution.

INTRODUCTION

Calorimetry in ATLAS

Calorimetry is an essential component of the ATLAS experiment [1] which is currently being assembled at one of the interaction points of the Large Hadron Collider (LHC) at CERN in Geneva, Switzerland. It is planned to start the operation of this 14 TeV proton-proton collider by mid-2007. The physics goals of the ATLAS experiment are the search for the Higgs boson, the precision studies of Standard Model physics, and discovery of new physics beyond this model.

Various technologies are deployed to cope with the different challenges and requirements for electromagnetic and hadronic calorimetry in different pseudo-rapidity (a quantity related to the scattering angle with respect to the beam, noted η) regions. Nonetheless, the measured signals in all ATLAS calorimeters need to be reconstructed and analyzed together in order to archive the required performance, in particular in the hadronic sector.

All calorimeters used in ATLAS are sampling calorimeters. One can distinguish the following sub-systems:

Electromagnetic Calorimeter The electromagnetic (EM) calorimeter is constructed from lead absorbers interleaved with liquid argon as active material. It covers the pseudo-rapidity region $|\eta| < 3.2$.

Hadronic Barrel Calorimeter The hadronic barrel calorimeter has large steel absorber equipped with scintillating tiles for readout (Tile Calorimeter). It has a central barrel part covering the η region up to 1.0 and an extended barrel part on each side that covers up to $|\eta| < 1.7$.

Hadronic End-Cap Calorimeter The Hadronic end-cap also features the liquid argon technology, but with different geometry. The absorbers are flat parallel copper plates. It is placed behind the EM end-cap in the same cryostat. It covers the η range from 1.5 to 3.2 in both detector hemispheres.

Forward Calorimeter The forward calorimeter provides electromagnetic as well as hadronic calorimetry in the very forward region (η between 3.2 and 4.9). It is located in the inner bore of the hadronic calorimeter, around the beam pipe. The first of three forward-calorimeter modules uses copper as absorber, the other two are made of tungsten. To cope with the higher counting rate, the active gaps are much thinner compared to the other liquid argon based calorimeters.

All calorimeters together comprise 187,652 geometrical cells.

Software Framework

The ATLAS reconstruction framework, Athena, is extensively described in [2]. It provides an event data store following a black board model. Algorithms retrieve and record data objects from and to this store, which takes ownership of the objects and destroys them at the end of the event. Additional stores for detector condition and geometry data are also provided, with a different object life cycle.

Athena also handles the data persistency and allows to establish links between persistent data objects, even across different physical storage locations, in addition to support object navigation in the transient event store.

RECONSTRUCTION DATA FLOW

The raw data coming out of the detector is stored as byte-stream on the CERN tape storage facility. The size of one event is about 1.6 MBytes, of which the calorimetry data takes about 0.5 MBytes. These files are the input to the ATLAS reconstruction framework.

The data objects representing the raw data in the transient store are the so-called Raw Data Objects (RDOs). Reconstruction of calorimeter data mainly involves refining the calibration already done online. These signal refinements include cell-by-cell corrections for non-linearities in the electronics chain, and corrections for possible high voltage problems, among others.

The reconstruction software also combines adjacent cells into towers and clusters. The result of the refined calibra-

* walter.lampl@cern.ch

tion are CaloCell objects, while the towers and clusters are represented by the CaloTower and CaloCluster objects, respectively. These composite objects then provide the input for jet-finding, electron/photon/tau reconstruction, and missing transverse energy calculations.

At two levels of the reconstruction chain, data are made persistent in LCG POOL [3] files. This simplifies the distribution of the acquired data to the various institutes participating in ATLAS and allows to re-do (a part of) the reconstruction without going back to the raw data files. Event Summary Data (ESD) files contain the result of the reconstruction, Analysis Object Data (AOD) contain physics objects suitable for analysis. There are strict size limitations for ESDs and AODs. On the other hand, more content makes physics analysis more flexible. Therefore the specific content of ESDs and AODs is a trade-off between these two requirements. ATLAS aims for an ESD size of 0.5 MBytes/event and an AOD size of 0.1 MBytes/event.

BASIC CALORIMETER DATA OBJECTS

The following section describes the data objects produced in the ATLAS calorimeter reconstruction chain. These objects have a lifetime of one event, e.g. they are constructed when data is read in or as (intermediate) result of the reconstruction and discarded at the end of each event. Some objects are stored at the end of each event in the ESDs and AODs discussed above.

In general, there are many instances of a certain data object (e.g. CaloCells) for each event. They are collected into storable containers, which are the data entity stored in the transient event store of Athena or written into a POOL file. Containers may imply a certain order or sub-structure of the objects they contain in order to speed-up access or to simplify iterations.

Raw Data Objects

The calorimeter raw data contains already a first order calibrated energy (in units of MeV) for each cell, which has been calculated online by a Digital Signal Processor (DSP) from the digitally sampled calorimeter signal.

Since two different readout technologies are used (scintillator readout for the Tile Calorimeter and ionization current readout for the Liquid Argon Calorimeter), the corresponding raw data object differ. For example, each Tile cell has actually two signals, as it is read out by two photomultipliers (PMTs), while only a single ionization signal is obtained from the Liquid Argon Calorimeter cell. In some more detail the two raw data object types, which are called LArRawChannel and TileRawChannel respectively, contain the following information:

Energy For Tile Calorimeter, there are two energy values (one for each PMT) while the LAr Calorimeters have only one energy per cell.

Time For cells with sufficient signal, it is possible to compute the arrival time of the signal.

Quality Gives an estimate of the quality of the ADC samples to energy conversion. It is only available for high-energy channels.

Gain The hardware readout gain

Online Identifier The online identifier encodes the position of the cell in the readout chain e.g. the identifiers of significant elements of the electronic chain and the readout channel number.

The raw data objects are stored in a LArRawChannelContainer or a TileRawChannelContainer, according to their type.

Calorimeter Cell Objects

The data object describing calorimeter cell data content and behavior is the CaloCell. It is commonly used by both calorimeter systems in ATLAS and is the smallest calorimeter signal object available. It contains similar quantities as the raw data object. The main difference is that it has only one energy signal, which is the result of a refined calibration. The scale of this signal is the electromagnetic energy scale in a given calorimeter sub-detector, meaning that these signals are balanced relative to each other across the system with respect to the electron response.

Each CaloCell is uniquely identified by an offline identifier, contrary to the use of an online identifier for the raw data objects. The offline identifier is a 32 bit data word encoding the geometrical position of the cell in a space of pseudo-rapidity (η) and azimuthal (φ) bin indexes, and calorimeter sub-detector and longitudinal sampling indexes, for all uniform calorimeter regions. A dictionary based tool allows to build a hash index, as well as conversion to and from the online identifiers describing the readout electronics, see above. The CaloCell also contains a pointer to a static *Detector Description Element* which contains all relevant geometrical information like position, volume, and lateral extensions in different reference frames.

CaloCells are stored in the transient store in a CaloCellContainer that can be made persistent. This container allows access to each CaloCell object using its unique hash identifier as a continuously running index. The container supports random and sequential access as well as selective access to cells in given calorimeter sub-detectors only.

Calorimeter Cell Compactification

CaloCells are stored in ESD files so that users can re-run higher level calorimeter reconstruction like clustering without going back to the raw data files. The total storage required for all 187, 652 CaloCells, with about 256 bytes of data per cell (plus one pointer), by far exceeds the ESD file size of 500 kBytes per event, even when compressed and without data from the other ATLAS detectors.

A detailed study of several compactification schemes has been conducted [4]. The finally implemented best scheme

with the least loss of signal precision and most gain in data reduction consists of two major steps. First, redundant and invalid information is omitted. The paramount example for redundant data is the cell identifier, which can be retrieved from a geometry database using the position (index) of the CaloCell in the CaloCellContainer.

The best example for the suppression of invalid data is the time information, which is only reliably available for cells with a large enough signal amplitude, by far the minority of all cells in a typical physics event in ATLAS. The time is therefore only stored for CaloCells with signals above the threshold.

The second step of compactification is the numerical compression of the calorimeter signal, i.e. the cell energy. The scheme chosen here takes the basic resolution power of calorimetric energy measurement into account, which typically scales as $1/\sqrt{E}$. The resolution limits are set by the electromagnetic calorimeter in ATLAS, as this detector features the best energy resolution in the order of $10\%/\sqrt{E}(\text{GeV})$. This performance is not allowed to be degraded by compactification of the energy data. This is currently achieved by storing the cubic root of the energy for all CaloCells and the logarithm of the time if this measurement is valid. Depending on the hardware readout gain, two energy resolution regimes are used. High resolution covers an energy range from 8 MeV to 50 GeV, while the normal resolution goes from 512 MeV to 3.2 TeV. Figure 1 shows the degradation of the sampling term due to the compactification in high and normal resolution mode.

The total data to be stored for a CaloCell is then one bit tagging the existence of valid time information, one bit for the sign of the energy signal, 12 bits for the cubic root of the absolute energy signal, two bits for the gain indicator, and, if valid, one bit for the sign of the time and 15 bits for the logarithm of the absolute value of the time. The total transient memory is therefore reduced to 0.39 MBytes per event, and the total persistent disk space needed, using POOL storage technology including additional compression, is 0.25 MByte.

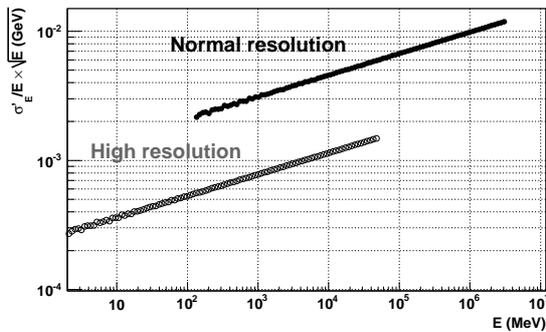


Figure 1: Degradation of the calorimeter sampling term due to the compact storage of the energy.

COMPOSITE CALORIMETER RECONSTRUCTION OBJECTS

The electromagnetic and hadronic showers developing in the calorimeter are generally bigger than one cell. In order to reconstruct the full energy of the incoming particle(s), several cells are combined into calorimeter towers and calorimeter clusters, following the spatial shower development.

Calorimeter Towers

Calorimeter towers are represented in Athena by the CaloTower object. Each tower is uniquely mapped onto a bin in a fixed grid in pseudo-rapidity (η) and azimuth (φ), with typical bin sizes $\Delta\eta \times \Delta\varphi = 5/200 \times 2\pi/256$ for electromagnetic towers in $-2.5 < \eta < 2.5$ and $\Delta\eta \times \Delta\varphi = 10/100 \times 2\pi/64$ for hadronic towers in $-5 < \eta < 5$.

For electromagnetic towers, only cells in the electromagnetic calorimeters are collected into towers according to their (η, φ) position. The tower energy is then the sum of all cell energies within a given tower. These towers are typically input to the sliding window cluster finder discussed below.

Hadronic towers contain cells from the whole ATLAS calorimeter system. They are typically used for jet reconstruction. Both types of towers use the same CaloTower data object, which implements the standard Athena interfaces for access and manipulation of the kinematic variables (four-momentum) and for object navigation.

CaloTowers are not made persistent at any level, but reconstructed on the fly. The only information needed for this is the grid description common to all towers of a certain size or type, and the list of contributing calorimeters. The assignment of CaloCells to CaloTowers is computed once at initialization and stored as a list of hash identifiers for each CaloTower, thus allowing to re-sum the tower energy without any further look-up.

Calorimeter Clusters

Clusters are groups of calorimeter cells with signals correlated by shower development and general energy flow in the collider event. Two kind of clustering algorithms are currently used by ATLAS:

Sliding window clusters The sliding window algorithm searches for a window in the electromagnetic tower grid where the total energy is at maximum. The window can be adjusted to different sizes.

Topological clustering The topological clustering algorithm attempts to aggregate neighboring cells with signals sufficiently above the expected noise. The noise cuts are adjustable. Signal structures (local maxima, for instance) are used to split topologically connected areas, if necessary.

Both algorithms produce the same output data object, the CaloCluster. It is supposed to be made persistent in ESD

and in a slimmed-down version also in the AODs, thus allowing re-building of electron/gamma, tau, and jet objects at the analysis data level, for example. The possibility of configuring CaloCluster for ESD and AOD storage makes it the most sophisticated data object in the calorimeter event data model.

Usually CaloClusters have a wealth of information associated with them. Contrary to CaloTowers, which have the same cells assigned to the same towers for a given grid, CaloClusters are more dynamic and therefore need a list of actual constituents, realized as list of links to CaloCell objects. Similar to CaloTowers, CaloCluster have kinematic information like a four-momentum and support object navigation, both implementing the standard Athena interfaces for these features.

In addition CaloCluster can have associated data, like signals and directions in individual calorimeter samplings included in the cluster, or shape variables (moments) describing the cluster for hadronic calibration purposes, for example. The AOD size restriction do not allow to store CaloClusters with all possible data cached in the object, not even with the potentially large number of links to CaloCells. At the level of the ESD it is still possible, though, to store all relevant data for a CaloCluster.

The solution presently implemented is that all links to CaloCells are actually stored in an external data object CaloCellLink. The CaloCluster then stores a link to this object, which resides only in the transient store or the ESD, thus reducing the number of cached links to CaloCells in CaloCluster to one, down from potentially several hundred. This concept can be considered as a relayed cell link store.

The already mentioned associated variables can also be stored in an external data object CaloShower, which again introduces only one link stored in CaloCluster, rather than a whole list of cached variables. All links are fully navigable, meaning that they can be followed within the transient store, within a given file, or even across different physical files, like from AOD to ESD. Assuming all variables are stored externally, a retrieval by a client at AOD level then triggers back navigation to the ESD to access the requested variable, or CaloCell, for example. Of course, this comes at a performance penalty when working with AODs.

Some cluster variables are very important for physics analysis. To allow access to these variables without the expensive back navigation across file systems, clients can store these (or all, for that matter) directly into the cluster. This concept can be viewed as a split store, where part of the data is locally cached while another part is stored in an external data object. The client controls the storage location when setting a variable by a simple logical flag in the interface. On retrieval, the CaloCluster always checks first if a requested variable is locally cached. If this is not the case, the client can again control the behavior by setting a logical flag in the interface requesting navigation to the external data object or not. In the latter case the variable return value is indicated as "invalid".

The moment store in CaloCluster is also organized following the split store design described above. It can be accessed randomly with the same client control features. In addition an iterator is supplied which allows to only access the cached store in CaloCluster, or both stores in a transparent way. Again, the client can define the behavior concerning the inclusion of the external store by a simple logical flag in the iterator interface.

CONCLUSION

The calorimeter Event Data Model (EDM) in ATLAS, as implemented in the standard reconstruction framework Athena, provides the three basic reconstruction objects for all calorimeters in a consistent way. All relevant physics reconstruction use cases have been taken into account in the design, including the ability to run downstream algorithms like jet finding on CaloCells, CaloTowers, and CaloClusters without code changes through implementing the standard Athena interface for data objects providing a four-momentum and supporting object navigation.

The data content of the most complex composite calorimeter data object CaloCluster is configurable by design, so that this object can be used at ESD and AOD level, thus allowing common tools to be used for both kind of input data. This also avoids the introduction of yet another data object representing the same reconstructed signals for the AODs, for example.

The ATLAS calorimeter EDM is successfully used in large scale simulation studies as well as test beam data analysis and detector commissioning. Small refinements are envisioned once collision physics data becomes available.

ACKNOWLEDGEMENTS

The ATLAS calorimeter software has been developed in a collaborative effort by many people throughout the last few years. We like to thank all contributors for their input at this time. We also like to recognize the important contribution from the ATLAS core software team. Last but not least we like to thank our respective funding agencies for their continuous support of our efforts.

REFERENCES

- [1] ATLAS Coll., *ATLAS Detector and Physics Technical Design Report*, CERN/LHCC/99-14/15 (1999).
- [2] The ATLAS Collaboration, *ATLAS Computing Technical Design Report*, CERN/LHCC/, May 2005
- [3] for example, see D. Duellmann, *The LCG POOL Project: General Overview and Project Status*, in Proc. 2003 Conf. for Computing in High Energy Physics (CHEP03), San Diego, La Jolla, California, March 24-28, 2003, eConf C0303241:MOKT007 (2003)
- [4] S. Menke, G. Usai, Proposal for Compactification of Calo-Cell Objects, ATLAS internal note, ATL-SOFT-2004-005, 2004