The new ATLAS Fast Calorimeter Simulation

Flavia Dias[*]

on behalf of the ATLAS Collaboration

University of Edinburgh (GB)

E-mail: flavia.dias@cern.ch

The physics and performance studies of the ATLAS detector at the Large Hadron Collider require a large number of simulated events. A GEANT4 based detailed simulation of the ATLAS calorimeter systems is highly CPU intensive and such resolution is often unnecessary. To reduce the calorimeter simulation time by a few orders of magnitude, fast simulation tools have been developed. The Fast Calorimeter Simulation (FastCaloSim) provides a parameterised simulation of the particle energy response at the calorimeter read-out cell level. In Run 1, about 13 billion events were simulated in ATLAS, out of which 50% were produced using fast simulation. For Run 2, a new parameterisation is being developed to improve the original version: it incorporates developments in geometry and physics lists during the last five years and benefits from the knowledge acquired from the Run 1 data. The algorithm uses machine learning techniques to improve the parameterisations and to optimise the amount of information to be stored in the ATLAS simulation infrastructure. The latest developments of the new ATLAS FastCaloSim are reviewed in this proceeding.

---

[*]Speaker.

1. Introduction

One of the key ingredients for a successful physics program of the ATLAS Experiment [1] at the Large Hadron Collider are the simulated Monte Carlo events. In generating such events it is necessary to simulate both the physical process of interest and the response of the detector to the produced particles. ATLAS uses a detailed detector simulation which makes use of the Geant4 [2] toolkit to model interactions between particles and materials of the detector. The drawback of such a detailed simulation is the CPU time requirement of several minutes per event [4], of which about 78% is spent for the calorimeter systems. With the rapidly increasing LHC luminosity, this CPU time requirement is a challenge to generate Monte Carlo simulated events with sufficiently large statistics. In some situations, such as the simulation of background events, a detailed detector description is not needed, and in such cases fast simulation techniques may be used.

The ATLAS Fast Calorimeter Simulation (FastCaloSim) [3] package provides a parameterised simulation of the particle energy response and of the energy distribution in the ATLAS calorimeter, reducing the calorimeter simulation time to a few seconds per event [4]. This article will introduce the current FastCaloSim as used in Run 1 and early Run 2 by the ATLAS Collaboration (Section 2), a description of the FastCaloSim upgrade (Section 3), and the summary and next steps in Section 4.

2. ATLAS Fast Calorimeter Simulation

The aim of FastCaloSim is to provide a simulation which is fast, but can reproduce the key features of reconstructed particle properties. Simplifications in the simulation model have to be accepted, but should not lead to large and unpredictable differences at the analysis level. The combination of fast simulation in the calorimeter systems with full simulation in the inner detector and muon system is called Atlfast-II. It reduces the overall simulation time by approximately one order of magnitude [4]. FastCaloSim uses the reconstruction geometry which describes calorimeter cells as cuboids (using η, ϕ and r for the barrel and endcap dimension variables and x, y and z for the forward calorimeters). This is a reasonable assumption for the homogeneous regions of the electromagnetic calorimeter, however only an approximation for their edges and the other calorimeter systems. The fast simulation parameterisation reproduces the longitudinal shower properties, including fluctuations and correlations, but only average lateral shower properties and uncorrelated lateral energy fluctuations. The FastCaloSim software is integrated into the Atlfast-II simulation package [4] that is currently used within ATLAS for Monte Carlo bulk production and was used to produce approximately half of the total ATLAS Monte Carlo statistics for physics analysis. FastCaloSim is able to simulate correctly most calorimeter objects. However, in hadronic showers it fails to describe the substructure of the groups of topologically related deposits in the ATLAS calorimeter cells. The goal of the new parametrisation is to incorporate developments in the geometry and physics lists from the past five years, as well as to benefit from knowledge acquired with Run 1 data to address the shortcomings of the current version.

3. ATLAS Fast Calorimeter Simulation Upgrade

The new version of FastCaloSim under development follows the general idea of the original

algorithm, i.e. modelling the calorimeter response to provide the longitudinal energy deposition and the lateral shower shape parameterisations, but using machine learning techniques.

Single particle inputs of photons, electrons and pions are simulated with the latest Geant4 version and latest ATLAS geometry and conditions. The detailed spatial position of each energy deposit is saved. The samples are generated in grids of energy (covering particle energies from 100 MeV up to about 4 TeV) and η bins (bin width of 0.05). The impact angle of the particle on a calorimeter cell surface is also considered.

The longitudinal shower development is described by the energy parameterisation. It uses a Principal Component Analysis (PCA) [5] to convert the set of correlated energy fractions into a linearly uncorrelated set of principle components. A TMVA [6] neural network (NN) regression analysis is performed to approximate the cumulative histograms used as parameterisation, to reduce the amount of information to be stored. The simulation procedure starts with the generation of uncorrelated random numbers, followed by the simulation of the energy fraction deposited in each calorimeter layer obtained via the PCA weights used as the parameterisation inputs. An example output of the simulation procedure is shown in Fig. 1, which shows good agreement between the Geant4 inputs and the fast simulated output.
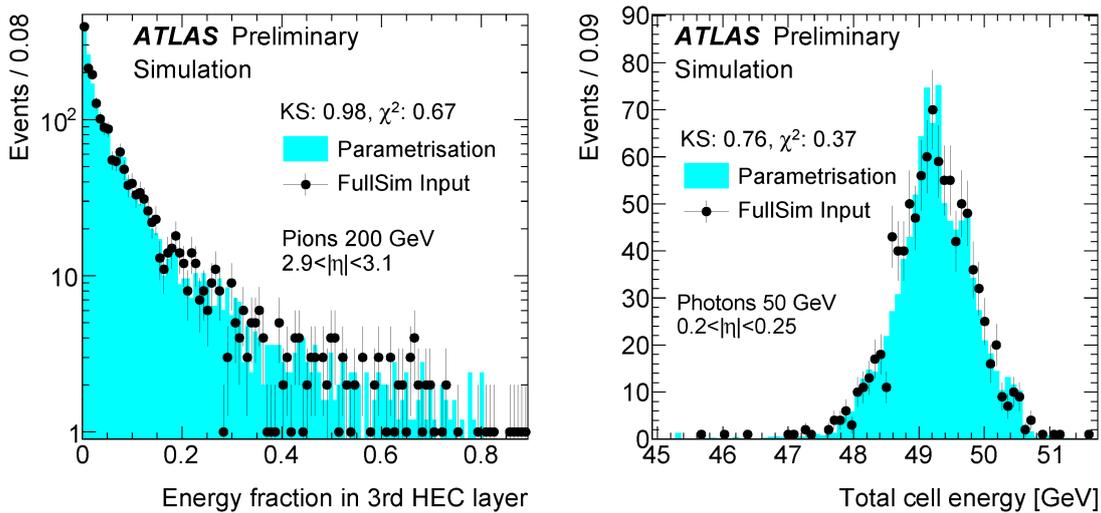

Figure 1: Energy fraction deposited in the third layer of the hadronic endcap calorimeter by charged pions (left) and total cell energy deposited in the calorimeter by photons (right). The black points show the Geant4 inputs, and the result of the longitudinal energy parametrisation is shown in light blue. The results of a Kolmogorov (KS) and χ^2 tests are displayed as well [7].

The lateral shower shape parameterisation describes the average shower shape in each calorimeter layer, for a given particle type, and each bin of the leading principle component variable. A NN is trained to describe the average shower shape instead of saving a full histogram information. The binning is chosen in two variables, r and α, in order to follow the radial symmetry of the barrel calorimeters. The number of bins in each variable is optimised in order to keep roughly the same number of hits in each bin. This condition leads to having a finer binning in the centre of the shower, and coarser bins in the tails. The NN is trained for each type of particle with a particular energy

and pseudorapidity and in bins of the leading PCA component from the energy parameterisation. An illustrative example of the NN training is shown in Fig. 2. The NN fit (right) can reproduce well the overall description of the simulated inputs (left). It does not describe the fluctuations and the details in the distribution, which are features that can be recovered with the lateral shower shape fluctuations.
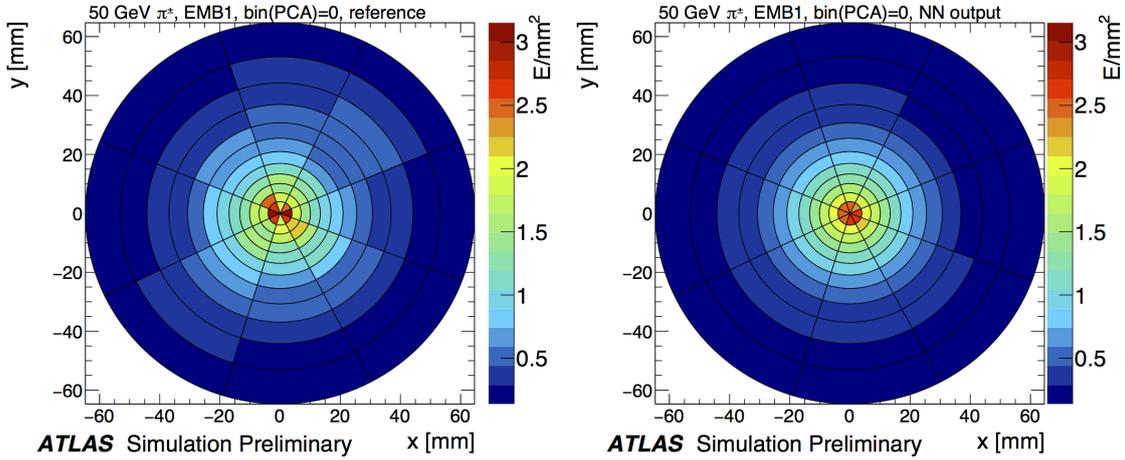

Figure 2: Left: An example case of the energy normalised per bin area used as input to the NN fit. The distribution is for pions with 50 GeV, located in the central pseudorapidity region (0.20<$|\eta|$<0.25) of the EMB1 layer and corresponds to events included in the first bin of the PCA energy parameterisation. The coordinates are defined with respect of the centre of the incident particle. Right: Distribution of the output of the NN parametrisation of the input on the left [7].

Lateral shape fluctuations are able to address the deficiency of the original FastCaloSim to describe the substructure of the groups of topologically related deposits in the ATLAS calorimeter cells. Those deposits, called topoclusters, are the base of ATLAS jet reconstruction, and hence FastCaloSim cannot be used for analyses that need a realistic simulation of these clusters, such as jet substructure studies. The shower fluctuations consist of depositing the energy in fewer cells than the original algorithm, while still following the average distribution given by the shape parameterisation, resulting in less homogeneous and more structured showers. In addition to the shower fluctuations, to address the region where the distance between the original pion and a cluster is large ($\Delta R(\pi,\text{cls}) > 0.15$), a complementary splitting method can be used, which defines a Geant4 based parameterisation of the energy deposition as a function of the distance between the pion's incidence position and its shower path. The parameterisation provides a probability with which a pion can be treated as two split pions, for a given energy and angular distance. These two approaches produce more clusters after the full simulation and reconstruction chain, giving a better agreement with the full Geant4 simulation than seen when using the standard FastCaloSim approach, as can be seen in Fig. 3 [8].

Hit to cell assignment is the last ingredient towards a new FastCaloSim prototype. The ATLAS liquid argon electromagnetic calorimeter has an accordion shape, which is not described by the simplified geometry used by the fast simulation. When assigning a hit produced by FastCaloSim to a
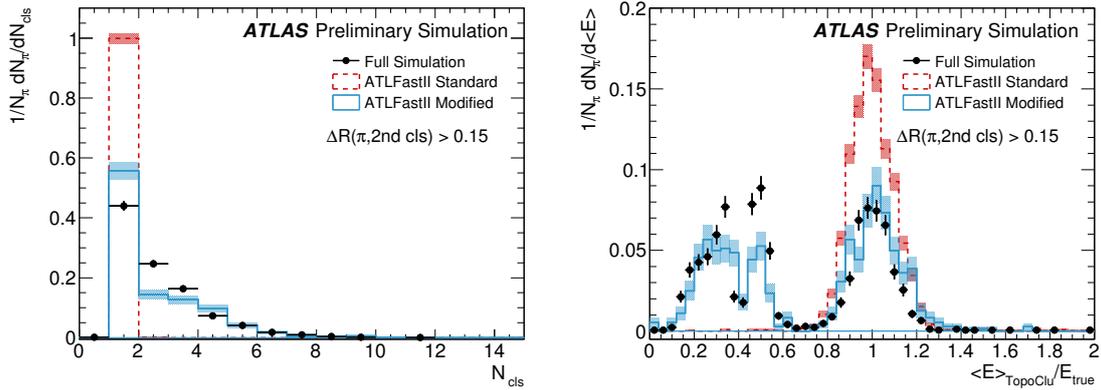
Figure 3: The number of topoclusters (left) and mean energy of a topocluster as a fraction of the true energy (right) for $\Delta R(\pi, \text{2nd cls}) > 0.15$) in the full Geant4 simulation (black points), standard FastCaloSim (red dashed histogram) and modified FastCaloSim (blue histogram) applying both random fluctuation and splitting approaches [8].

given calorimeter cell, the difference in geometry can introduce a bias of cell misassignment. This can be seen in the left plot of Fig. 4, where the energy profile ratio between the fast simulation and the reconstructed cell energy used as reference is above one in the central region of the ΔR (π, cell) and below one in the outer regions in ϕ direction. This bias can be greatly reduced when the hit displacement method ("wiggle") is used, as it can be seen in the right plot of Fig. 4. The "wiggle" method consist of changing the hit position by a small variation, $d\phi$, before calling the geometry tool to assign the hit to a cell. If the hit is in the middle of the cell, the small displacement will not change which cell the hit will be assigned to. If the hit is close to the edge of the cell, the displacement method might change the cell assignment. The variation $d\phi$ depends on the calorimeter layer and is tuned in order to give the least biased ratio of the FastCaloSim and reconstructed cell energy profile.

4. Summary and Outlook

An update of the ATLAS Fast Calorimeter Simulation software is ongoing. It aims to improve the current FastCaloSim in terms of speed, precision, and memory management. The updated algorithm uses the latest ATLAS geometry, physics list, and knowledge from Run 1 and early Run 2 to improve the parameterisation of the energy deposits and lateral shower shape. Additionally, principal component analysis and neural network regression are used to provide the best description of the objects being parameterised while keeping the stored information slim and quick to access. The updated FastCaloSim also aims to solve some of the shortcomings form the original version, in particular the description of the shower sub-clusters, a fundamental feature to allow fast simulation to be used in cases where jet substructure is relevant. This is achieved by the addition of lateral shape fluctuations and the muon splitting algorithm, which were shown in preliminary studies to improve the performance of sub-clusters variables with respect to the current FastCaloSim.

(a) Simplified geometry without corrections   (b) Simplified geometry with hit displacement
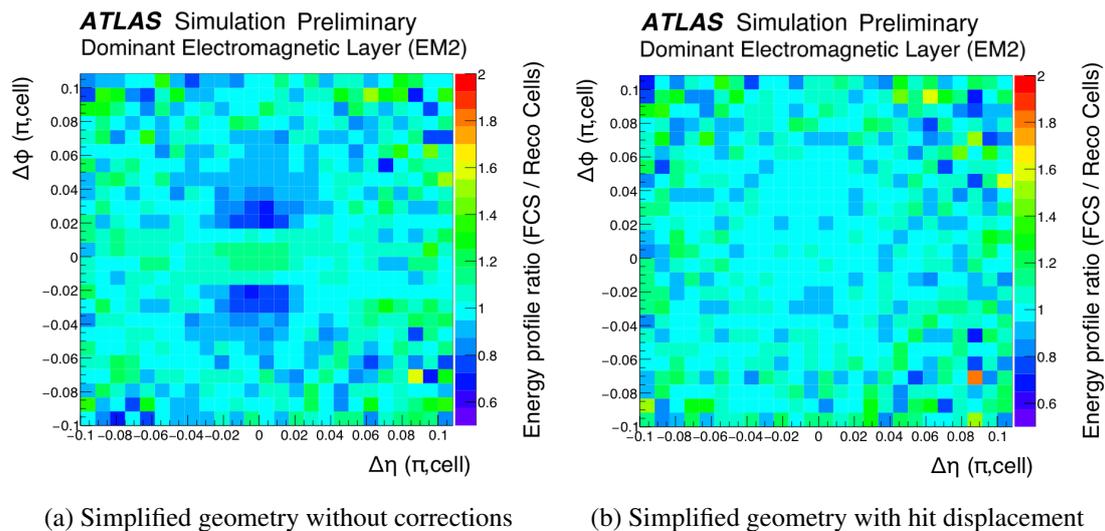
Figure 4: The ratio of the FastCaloSim energy profile and the reconstructed cell energy profile, as a function of the distance of the centre of the cell and the pion calorimeter entrance position $\Delta\eta(\pi,\text{cell})$, $\Delta\phi(\pi,\text{cell})$, for the original hit-cell assignment with the simplified geometry (left) and the modified hit-cell assignment using the wiggle hit displacement method (right). The bias in ϕ due to the wrong description of the accordion shape of the calorimeter in the simplified geometry is greatly reduced when using the hit displacement method. The dominant electromagnetic layer (EM2) is the ECAL layer where most of the particle's energy is deposited [7].

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the Large Hadron Collider*, *JINST* **3** S08003

[2] S. Agostinelli et. al., *Geant4 - a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** no. 3

[3] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, *ATL-PHYS-PUB-2010-013*

[4] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** 823-874

[5] Pearson, K., *On Lines and Planes of Closest Fit to Systems of Points in Space*, *Philosophical Magazine* **2** (11)

[6] Hoecker A et. al., *TMVA: Toolkit for Multivariate Data Analysis*, *Proc. of Science PoS ACAT 2007* **040**

[7] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ComputingandSoftwarePublicResults, retrieved October 2016

[8] ATLAS Collaboration, *Performance of the Fast ATLAS Tracking Simulation (FATRAS) and the ATLAS Fast Calorimeter Simulation (FastCaloSim) with single particles*, *ATLAS-SOFT-PUB-2014-01*