Upgrading the ATLAS fast calorimeter simulation

Z Hubacek\textsuperscript{1}, on behalf of the ATLAS Collaboration

\textsuperscript{1} Czech Technical University in Prague, FNSPE, Brehova 7, 115 19, Praha 1, Czech Republic
E-mail: zdenek.hubacek@cern.ch

Abstract. Many physics and performance studies with the ATLAS detector at the Large Hadron Collider require very large samples of simulated events, and producing these using the full Geant4 detector simulation is highly CPU intensive. Often, a very detailed detector simulation is not needed, and in these cases fast simulation tools can be used to reduce the calorimeter simulation time. In ATLAS, a fast simulation of the calorimeter systems was developed, called Fast Calorimeter Simulation (FastCaloSim). It provides a parametrized simulation of the particle energy response at the calorimeter read-out cell level. It is interfaced to the standard ATLAS digitization and reconstruction software and can be tuned to data more easily than Geant4. An improved parametrization is being developed, to eventually address shortcomings of the original version. It makes use of statistical techniques such as principal component analysis and a neural network parametrization to optimise the amount of information to store in the ATLAS simulation infrastructure.

1. Introduction

The success of the physics program of the ATLAS Experiment \cite{ref1} at the Large Hadron Collider relies heavily on the large number of Monte Carlo (MC) simulated events. The MC production takes a significant part of the available grid CPU and disk space resources, therefore every possibility for improvement is studied. One of the possibilities is the use of fast simulation tools. The ATLAS Monte Carlo production chain and the Integrated Simulation Framework (ISF) are briefly described in Section 2, followed by the description of the main features of the Fast Calorimeter Simulation (FastCaloSim) in Section 3. The development of the new FastCaloSim version is described in Section 4.

2. ATLAS Monte Carlo Production Chain and ISF Framework

The typical ATLAS MC production chain \cite{ref2} consists of several steps. The event generation simulates the truth-level particle collision with the possibility to choose from common Monte Carlo generators. The event information is stored in the EVNT file format, based on HepMC. The detector simulation computes the sensitive detector energy deposits (HITS) caused by the final particles from the EVNT files. Here either the full detector simulator (Geant4 \cite{ref3}) and/or fast ATLAS specific simulators are used. Next, the digitization converts the HITS into raw data objects. During the digitization, in-time and out-of-time pileup effects can be merged into the signal event. The reconstruction finds and identifies physics objects (photons, leptons, jets, \ldots). With respect to the reconstructed physics objects, the output file contents are equivalent between the simulated and recorded data. In the final step, a conversion to physics analysis format is usually applied to allow fast ROOT \cite{ref4} access to the reconstructed objects used in the physics analyses.
To reduce the CPU utilization, a number of developments are undertaken to speed up the ATLAS detector simulation. The Geant4 based simulation of the detector is taken as the most detailed and the most accurate detector simulation. Precomputed (so-called frozen showers) libraries of electromagnetic showers in the forward calorimeters are used by default also in the full ATLAS simulation. Additional fast simulators available consist of Fatras \cite{5}, a fast ATLAS tracking simulation, and of FastCaloSim \cite{6}, a parametrized ATLAS calorimeter response simulation.

The Integrated Simulation Framework (ISF) \cite{7} was developed to allow combining various simulators (full and fast) for different particles within each event. In this way, one could combine the full and precise simulator for particles of interest and fast simulators for the rest of the event. The ISF is now fully integrated in the ATLAS Athena framework \cite{8}. An example of a speedup of the simulation for different simulation setups is presented in Table 1.

<table>
<thead>
<tr>
<th>ISF Simulation Setup</th>
<th>Speedup</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Geant4</td>
<td>1</td>
<td>best possible</td>
</tr>
<tr>
<td>Geant4 + FastCaloSim</td>
<td>≈ 25</td>
<td>approximated calorimeter</td>
</tr>
<tr>
<td>Fatras + FastCaloSim</td>
<td>≈ 750</td>
<td>all subdetectors approximated</td>
</tr>
<tr>
<td>Fatras + FastCaloSim only simulating particles inside cones around photons</td>
<td>≈ 3000</td>
<td>all subdetectors approximated only partial event simulation</td>
</tr>
</tbody>
</table>

Table 1. Comparison of various ISF simulation setups in the $gg \rightarrow H \rightarrow \gamma \gamma$ events without pileup contributions.

3. FastCaloSim
The simulation of the particle shower evolution takes a large amount of time. In the FastCaloSim simulation, time is saved by considering a simplified detector geometry and replacing the simulation by parametrizations. Only three types of particles are parametrized and used by the FastCaloSim: photons, electrons and charged pions. The charged pions are used for all hadrons (both neutral and charged). In order to reflect the chosen simulation priorities, the parametrization model reproduces the longitudinal shower properties, including fluctuations and correlations, but only the average transverse shower shapes are used. The properties of the shower developments are stored in histogram form for each particle type, each energy and $\eta$ points.

The FastCaloSim was validated and used in the Run 1. The average lateral shower shapes were tuned to data, but some shortcomings of the simplified model were observed, particularly in shower sub-structure (Fig. 1). Because of the limited computing resources, only the average lateral shower shapes could be saved. In the context of the new FastCaloSim development, a random shower shape fluctuations model has also been developed to improve the situation \cite{9}. Overall, a good agreement between the FastCaloSim and the full Geant4 simulation was observed, and the fast simulation reduced the simulation time in the ATLAS calorimeter from several minutes to a few seconds per event.

4. Upgrade of Fast Calorimeter Simulation
A new version of the fast calorimeter simulation is currently being developed. The new parametrization is based on the latest Geant4 version used by the ATLAS experiment. The development is split again into the simulation of the calorimeter response of single particles, the
parametrization of the longitudinal energy deposition, the fitting of the average lateral shower shapes and implementing the shower shape fluctuations.

**Single particle inputs** of photons, electrons and pions are used as in the original FastCaloSim. The detector response to these particles is 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 single particle energies from 100 MeV up to about 4 TeV) and 10 bins (bin width of 0.05). The impact angle of the particle on a calorimeter cell surface is also considered.

**Energy Parametrization** describes the longitudinal shower development. It utilizes the Principal Component Analysis (PCA) to convert the set of correlated energy fractions into a linearly uncorrelated set of principle components. A TMVA [11] neural network (NN) regression is used to approximate cumulative histograms inputs (Fig. 2) in order to reduce the amount of stored information. The simulation procedure will start with the generation of uncorrelated random numbers and will invert the PCA matrix in order to simulate the energy fraction deposited in each calorimeter layer. An example output of the simulation procedure is shown in Fig. 2.

**Lateral Shower Shape** describes the average shower shape for each particle type in each calorimeter layer, and bins of the leading principle component variable. A NN is trained to describe the average shower shape instead of saving a full histogram information. A radial symmetric binning is optimized so that the number of hits remains roughly the same in each bin and a finer binning is chosen in the center of the shower as shown in Fig. 3. The NN is going to be trained for each particle type, each energy and η point and in bins of the leading PCA component from the energy parametrization. An illustrative example of the NN training is shown in Fig. 4.

Typical NN settings for both Energy Parametrization and Lateral Shower Shape steps are shown in Table 2.
Figure 2. Illustration of the iterative regression with an increasing number of neurons for one of the cumulative histograms used in the longitudinal energy parametrization of charged pions. The example here shows the energy fraction deposited in the first layer of the Tile calorimeter (left) and energy fraction deposited in the 3rd layer of the Hadronic Endcap calorimeter by charged pions. The black points show the Geant4 inputs, and the result of the longitudinal energy parametrization is shown in light blue. A good agreement is observed. The results of a Kolmogorov (KS) and $\chi^2$ test are displayed as well (right).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Default value</th>
<th>Value used</th>
</tr>
</thead>
<tbody>
<tr>
<td>HiddenLayers</td>
<td>N</td>
<td>2–10</td>
</tr>
<tr>
<td>TestRate</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>NeuronType</td>
<td>sigmoid</td>
<td>tanh</td>
</tr>
<tr>
<td>TrainingMethod</td>
<td>BP</td>
<td>BFGS</td>
</tr>
<tr>
<td>ConvergenceImprove</td>
<td>1e-30</td>
<td>1e-6</td>
</tr>
<tr>
<td>ConvergenceTests</td>
<td>-1</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 2. TMVA NN parameter settings used in the Energy Parametrization and Lateral Shower Shape steps.

5. Summary
A development of a new ATLAS fast calorimeter simulation is underway. It is aiming at improving the current FastCaloSim speed and precision. The new fast calorimeter simulation is going to use PCA methods for the energy parametrization and NN regression for the fitting of the average shower shapes.

Acknowledgments
The author gratefully acknowledges the financial support from MSMT of the Czech Republic under Grant No. RVO 68407700.

References
[1] The ATLAS Collaboration 2008 The ATLAS Experiment at the Large Hadron Collider *JINST* 3 S08003
Figure 3. Illustration of the definition in terms of $\eta$ and $\phi$ of the hits of the alpha and $r$ variables, which are used to optimise the binning and perform the NN fit. This example is for 50 GeV central ($0.20 < |\eta| < 0.25$) pions in the first layer of the electromagnetic calorimeter (EMB1) and corresponds to events included in the first bin of the PCA energy parametrization (left) and of the number of hits per bin used for the NN fit. The number of hits per bin is kept as constant as possible for the NN fit stability. At the very center, a more detailed description of the shower shape is needed, therefore the bins are sub-divided there, hence the decrease observed in the number of hits per bin (right).

Figure 4. Illustration of the energy normalized per bin area used as input to the NN fit. This example is for 50 GeV central ($0.20 < |\eta| < 0.25$) pions in the EMB1 layer and corresponds to events included in the first bin of the PCA energy parametrization (left) and of the output of the NN parametrization of the input (right).


