Frozen Shower Simulation - Optimization and Compression

Summer student Program 2009, DESY

Olaf Nackenhorst[1]
W. Ehrenfeld[2], A. Glazov[2], M. Karnevsky[2], R. Plačakytė[2]

[1] Universität Göttingen, [2] Deutsches Elektronen Synchrotron

September 2009

Abstract

This report describes the studies performed on the optimization and compression mechanisms for the fast electromagnetic shower simulation of ATLAS calorimeter, i.e. so-called Frozen Shower (FS) library technique. Frozen Shower simulation is typically used for low energy particles where the full GEANT simulation of the ATLAS calorimeter is substituted with a shower template thus reducing the particle simulation time without affecting the accuracy. The studies on a possible optimization of the FS approach with libraries extended to higher energies are described in the first part of this report. In the second part the studies on the reduction of the memory consumption presently required by the FS libraries during GEANT simulation in ATLAS detector are presented. These studies showed that a significant improvement (40%) in memory consumption can be achieved with a simple conversion of library variables and the result of this development has been implemented into the official ATLAS software.

Contents

1 Introduction

1.1 The ATLAS Detector - a very brief Overview

ATLAS (A Toroidal LHC ApparatuS) is one of the two general purpose detectors of the Large Hadron Collider (LHC) at CERN, Geneva, Switzerland. With a design center of mass energy of 14 TeV and design luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ approximately 22 proton-proton collisions per bunch crossing will occur every 25ns in the detector. The structure of ATLAS is almost cylindrical around the interaction point with a length of 42 m, a radius of 11 m and a weight about 7000 tonnes. The detector consist of four major components, the Inner Tracker which measures the momentum of charged particles, the calorimeter which measures the deposited energy by the particles, the Muon spectrometer which identifies muons and measures their momentum and the Magnet system that forces charged particles to move on curved trajectories.

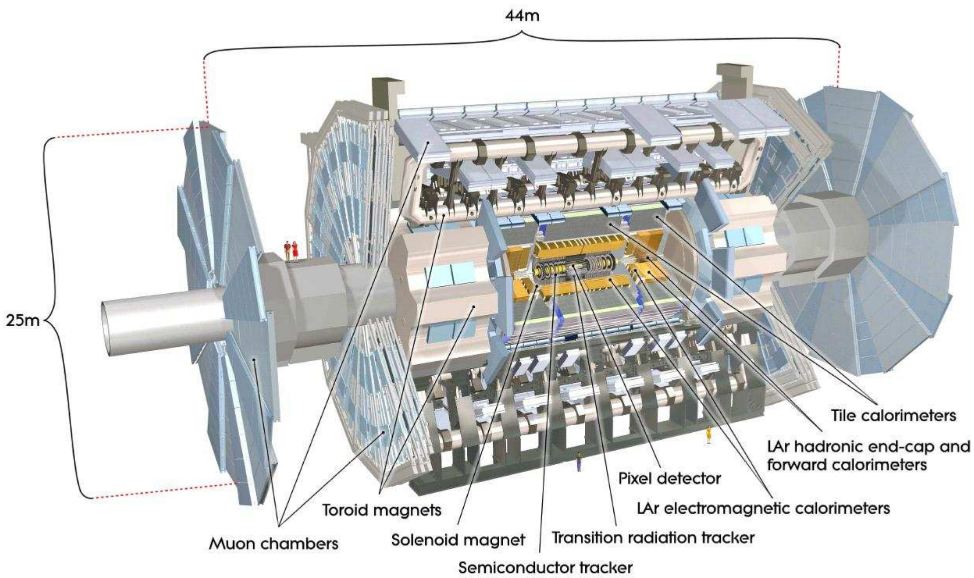

Figure 1: *The ATLAS Detector*

1.1.1 The LAr EM Calorimeter

The Liquid Argon (LAr) electromagnetic (EM) calorimeter is a sampling calorimeter with accordion-shaped lead electrodes in the barrel (EMB) and in the endcaps (EMEC) ensuring continuity in azimuth and a forward calorimeter (FCAL) close to the beam pipe consisting of a copper and tungsten matrix filled with concentric rods and tubes. With the pseudorapidity defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$$

the acceptance for EMB is $|\eta| < 1.5$, EMEC covers the range $1.4 < |\eta| < 3.2$ while FCAL is in the very forward region with $3.1 < |\eta| < 4.9$.

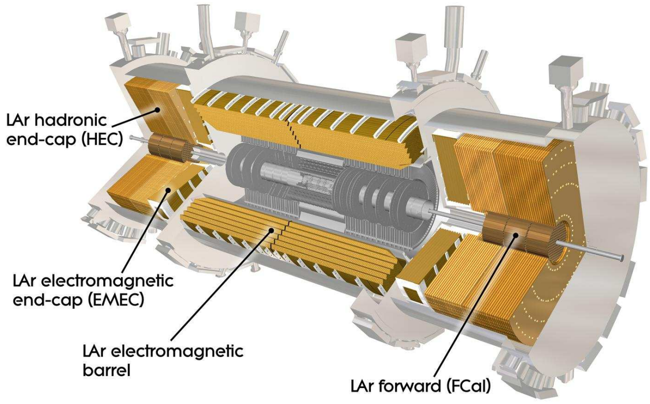

Figure 2: *The EM calorimeter of the ATLAS detector*

1.2 Monte Carlo Production Chain in ATLAS

No analysis in high energy physics can be done without a proper simulation of the experiment to verify the results and to develop analysis methods before the data taking. In the ATLAS experiment this is done by GEANT4 using a detailed microscopic description of the interactions between particles and matter and there is a full chain which consist of the following steps:

- **Generation:** Production of four vectors from specified physics processes which can be done by many different generators, e.g. AlpGen, Herwig, Pythia etc.
- **Simulation:** After generation the events are passed through a GEANT4 simulation of the ATLAS detector to produce hits which determine where the particle crosses the detector and how much energy was deposited by the particle.
- **Digitization:** The simulated GEANT4 hits are subjected to the response of the detector to produce digits, like times and voltages, such as the real detector would do.
- **Reconstruction:** Either simulated and digitized or real data as well has to be reconstructed into tracks and energy deposits that one is able to analyze it.

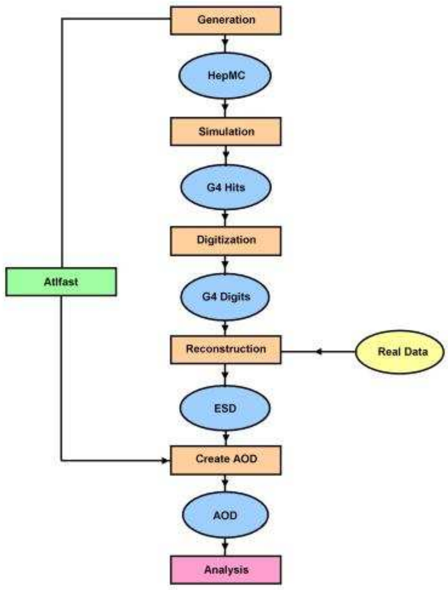

Figure 3: *The Monte Carlo Production Chain in ATLAS*

2 Fast Simulation

For a typical 14 TeV pp collision event modern processors need about 15 minutes of CPU time due to the complexity of ATLAS and the multiplicities of particles. About 70% of simulation time is spent in the calorimeter where electromagnetic particles require the largest part of computing time. There are several fast simulation approaches available in ATLAS, one of them is called *Frozen Showers*.

2.1 Frozen Shower Library

With the Frozen Showers (FS) approach pre-stored EM showers are used instead of low energy (below certain threshold) particle simulation while only the high energy part of the full shower simulation is performed directly by GEANT4. FS are typically used for electrons and photons.Using this method one can reduce the simulation time for electrons by a factor of 20 and for a typical physics event by a factor of 2-3 with a very good description of the shower profiles.

2.1.1 Library Generation

For electrons and photons the Frozen Shower Libraries are organized in typically 10 energy bins and 21 η bins which contain 1000 showers each that the selection is arbitrary enough. These libraries are simulated by GEANT4 with energy values in a logarithmic scale [1, 2, 5, 10, 20, 50, 100, 200, 500, 1000] MeV due to the shower cascade and discrete values for

η [0.1, 0.3, 0,5, 0.81, 0.83, 1.1, 1.3] in the EMB and [1.62, 1.78, 1.82, 1.98, 2.02, 2.08, 2.12, 2.28, 2.32, 2.40, 2.60, 2.78, 2.82, 3.15] in the EMEC due to the detector geometry. In FCAL there are the same energy bins but only 2 position bins either the active material or the passive absorber is hit while η is flat distributed. In order to minimize the additional disk space and memory consumption the Frozen Shower hits are compressed in three steps:

- **Clustering:** A sophisticated algorithm finds the two nearest energy deposits and merges the hits to one at the center and with the sum of the energy if the spatial distance is smaller than a defined maximum value R_{max}. This procedure is repeated until all distances are above R_{max}.
- **Truncation:** To avoid hits with very small energy values all hits are sorted by energy and beginning from the highest energy only those are saved which have a combined energy exceeding 95% of the initial shower.
- **Rescaling:** To preserve the barycenter and the second moment of the true shower the remaining total energy and the radial coordinates are rescaled and only the fraction of the original energy is stored for the hit.

3 Optimization of FS Libraries

Although the technique of Frozen Showers is already quite successful there are a few ways to speed up the simulation process even more.

3.1 Pion Libraries

After electrons and photons are substituted with FS libraries, pions become the simulation dominating particles and are natural candidates for the FS approach. A significant percent of simulation time is spent in the FCAL. The $Z \rightarrow e^+e^-$ decay is an appropriate process to determine a good energy binning of the pion library:[150, 500, 1000, 5000, 10000, 15000] MeV. Like for electrons and photons η is uniform distributed within the acceptance of the FCAL and usually 2000 events are generated for each bin due to the wide range and binning. The use of pion libraries can reduce the simulation time further significantly, e.g. for $Z \rightarrow e^+e^-$ by 14% and the shower profiles of the fast simulation show a good agreement with the full simulation.

3.2 Neutron Cuts

Having the pion libraries applied to simulation, next thing to investigate is the remaining particles which dominates the simulation time. It turns out that roughly 50% of the left over particles in the EM are neutrons. Most of these neutrons have energy below 50 MeV. Such thermal neutrons have almost no influence on simulation of most of physics processes and therefore can be safely removed from the simulation, i.e. are *killed*. This common approach merely dumps the energy of the neutron locally in one spot. This easy implementation can impressively improve the simulation time by roughly 40% and has no visible effect on the distributions which show the shower profiles.

3.3 Maximum Energy of Libraries

Another open question is whether one can reduce the computing time by extending the libraries to higher energies. As a test, the libraries for electrons and photons were generated which contain showers up to 2 GeV and 5 GeV. The 2 GeV libraries were tested in simulation with pion libraries included and neutron cuts applied. Showers were pulled out of the libraries if the electrons or photons reached the threshold of 2 GeV. Although the performance for single electron simulation could be improved by 17% in timing, an improvement in a physics event like $Z \rightarrow e^+e^-$ could not be observed. In the following table the most important results are summarized:

Event	$Z \rightarrow e^+e^-$	single e
Full Simulation	552.84 ± 7.96	49 ± 1.20
Frozen Shower / 1 GeV	166.53 ± 2.79	0.47 ± 0.014
Frozen Shower / 2 GeV	162.53 ± 2.73	0.39 ± 0.013
FS + ncut / 1 GeV	103.79 ± 1.85	
FS + ncut / 2 GeV	105.60 ± 1.91	
FS + ncut + pions / 1 GeV	100.70 ± 1.81	
FS + ncut + pions / 2 GeV	97.94 ± 1.78	

Table 1: Simulation time in seconds

The conclusion is that with all optimizations applied the simulation time is 94.94 ± 1.78 s, while the simulation time is 100.07 ± 1.81 s if one sets the energy threshold to default which is 1 GeV for electrons and 10 MeV for photons for the Frozen Shower simulation. There is a small tendency to lower simulation time but within the deviations no visible improvement. Since the memory consumption increases considerably when the libraries up to 2 GeV are used the conclusion of these studies is that the libraries are already at there optimal energy threshold. Further studies with energies up to 5 GeV were therefore neglected.

4 Compression of Frozen Shower Libraries

Due to the historical development of the Frozen Shower Simulation the memory consumption constantly increased and end up at approximately 1200Mb. In the beginning only electrons were replaced by shower templates in libraries, then photon and later pion libraries were added. This results in a high FS memory and a large contribution to the full simulation time of an ATLAS event. The hits of the showers are already partly compressed as mentioned before but they are read in as doubles during the Frozen Shower simulation. By converting the hits from double to float or even further to short integer one should be able to save a lot of memory. While the conversion from double to float is quite trivial, the conversion to short integer needs some studies.

4.1 Class Hit

The LArG4ShowerLib package in the ATLAS software framework *ATHENA* contains the data objects used for Geant 4 based shower parametrization and one of the shower library class is called *Hit*. This class stores the single hit information and contains all the Get-functions for the hits. The constructor reads in the energy E, the x-, y- and z-coordinate of the hit as doubles and recalculates the coordinates to cylindrical coordinates, the angle ϕ, the radius r and the axis z which are private values. The energy E is stored as a fraction of the total energy of the shower and therefore defined in the interval [0, 1], ϕ is in [$-\pi, \pi$], r in [0, r_{max}] and z in [z_{min}, z_{max}]. Formally an empty destructor is defined but not used in the simulation because the showers are reused every time. The class contains functions for the input and output stream and several Get-functions which not only return the private values in original coordinates but also can add energy and do a random rotation in ϕ coordinate.

4.2 Implementation

With the knowledge described in the section before it is quite easy to convert the private values of the Hit class to short integer

$$E(int) = E(double) \cdot \text{Max(int)} \tag{1}$$

$$\phi(int) = \phi(double) \cdot \frac{\text{Max(int)}}{2\pi} \tag{2}$$

$$r(int) = r \cdot 10 \tag{3}$$

$$z(int) = z \cdot 10 \tag{4}$$

with the maximal unsigned integer value Max (int)=$2^{16} - 1$=65535. Only E and r are defined as unsigned integer because ϕ and z can become negative, r and z are multiplied with a factor of 10 to take microns into account for higher precision. In the Get-functions the integer values are recalculated back to double and the Add- and Rotation-functions have also been taken into account.

4.3 Results

The new classes with float and short integer values were tested for single particles (e, γ, π) and $Z \to e^+e^-$ event simulation including pion libraries and compared with the full simulation and the fast simulation with double values. It turns out that one can save an additional amount of memory by commenting out the destructor which is not used in the simulation anyway.

4.3.1 Memory

It is quite impressive how much memory usage one can reduce with this simple method. With the conversion from double to float one can save 23% of the memory, the conversion

	double	**float**	**short int**	**full**
Memory [Mb]	1225	942	817	592
without destructor	1163	1006	752	-
virtual Memory	1325	1070	944	718

Table 2: Memory Consumption

to short integer can reduce the memory usage even by 33%. In addition one can gain 60 Mb simply by deleting the destructor of the Hit class. Overall a reduction by roughly 500 Mb which corresponds to 40% of reduced memory consumption and is a good result.

4.3.2 Distributions

After the memory usage by FS has been significantly reduced, one has to make sure that the shower profiles of the simulation with float and integer values agree with the simulation using same values as doubles which has been shown to agree with the full simulation. As a representative example a simulation of single 10 GeV electron simulated within the full detector acceptance ($-4.9 < \eta < 4.9$. The simulation using float as well as short integer show a good agreement and are presented in figure [?].

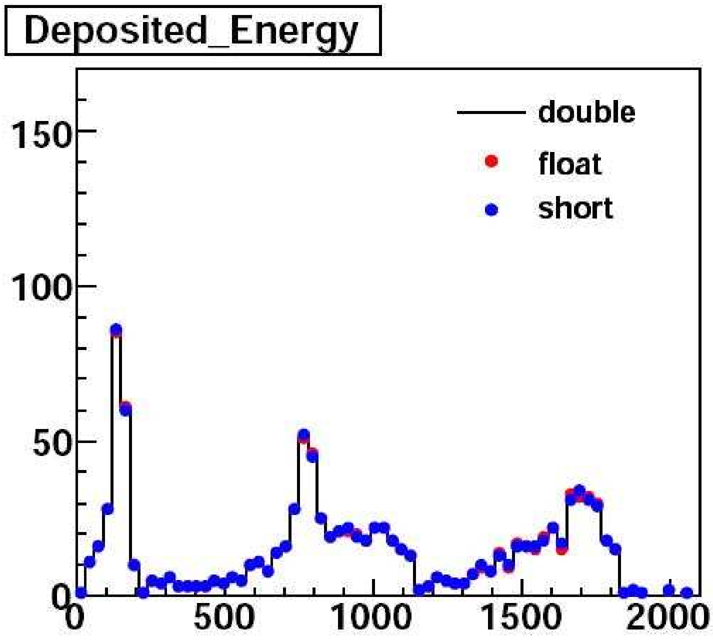

Figure 4: *The deposited energy of a 10 GeV single electron in the ATLAS detector for short integer, float and double variables*

An interesting observation can be seen in CPU time distribution of the simulation (figure [?]) where is visible that simulation using library variables as float is faster with respect to other simulations. This features is not yet understood. The distribution of the short integer simulation gives a close match to the double simulation - one would expect a slightly longer CPU time due to all the additional calculations. All other distribution for electrons and other single particles like γ and π can be found in the Appendix.

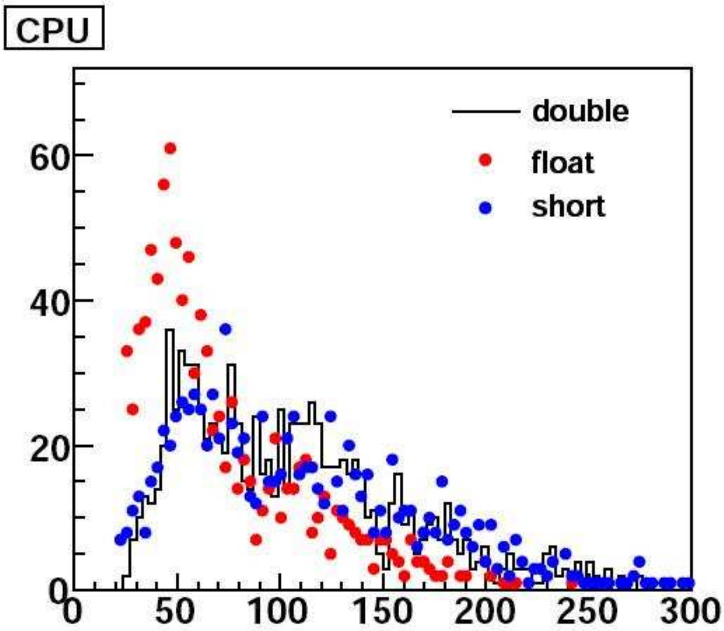

Figure 5: *The simulation time of a single electron in the EM calorimeter for short int, float and double variables*

There was another test performed with FS using a real physics event like $Z \rightarrow e^+e^-$. The event was simulated under the same condition like for single particles and reconstructed after some simple kinematic cuts. It was asked for 2 electrons with $P_t > 15$ and $\eta < 2.37$ and a invariant mass window $70 < M_e e < 110$ GeV of the e^+e^--system. Because one would naturally expect a good agreement for double and float we compare only the distribution of double and short int methods while the the blue line shows short and the black dots with line shows double.

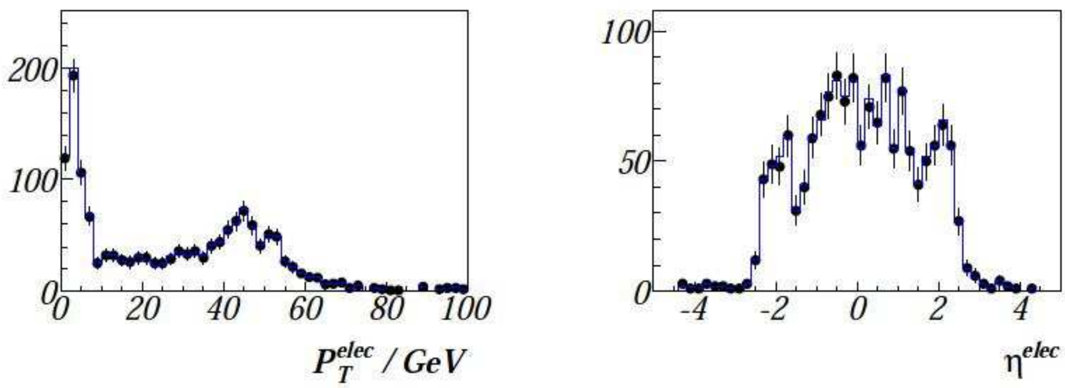

Figure 6: *The distribution of transverse momentum (left) and pseudorapidity (right) of simulated $Z \rightarrow e^+e^-$ with 10 GeV energy in the EM calorimeter*

5 Conclusion

The FS approach can significantly reduce the event simulation time in the ATLAS detector. Pion libraries in addition to e/γ FS libraries brings additional 14% improvement in simulation time. Due to high numbers of low energy neutrons in calorimeter, the *killing* of neutrons brings a remarkable improvement. The optimum energy threshold for FS libraries were teted. The memory consumption by FS libraries can be reduced to 40% with the method described in this report without loosing shower information.

6 Acknowledgements

First I would like to thank Ringaile Plačakytė for her incredible patience answering all my questions and showing me all needful technical details and tools. W. Ehrenfeld, A. Glazov and M. Karnevsky helped and instructed me whenever it was necessary or whenever Ringaile was not available. Joachim Meyer made it possible to me to take an active part in a fantastic organized and structured summer student program. Thank you all for a wonderful and very interesting time at DESY in Hamburg, it was a marvelous opportunity to deepen my knowledge in high energy physics and in addition to meet a lot of great young students from all over the world with similar interests.

A Plots for Compression Validation

A.1 single electron

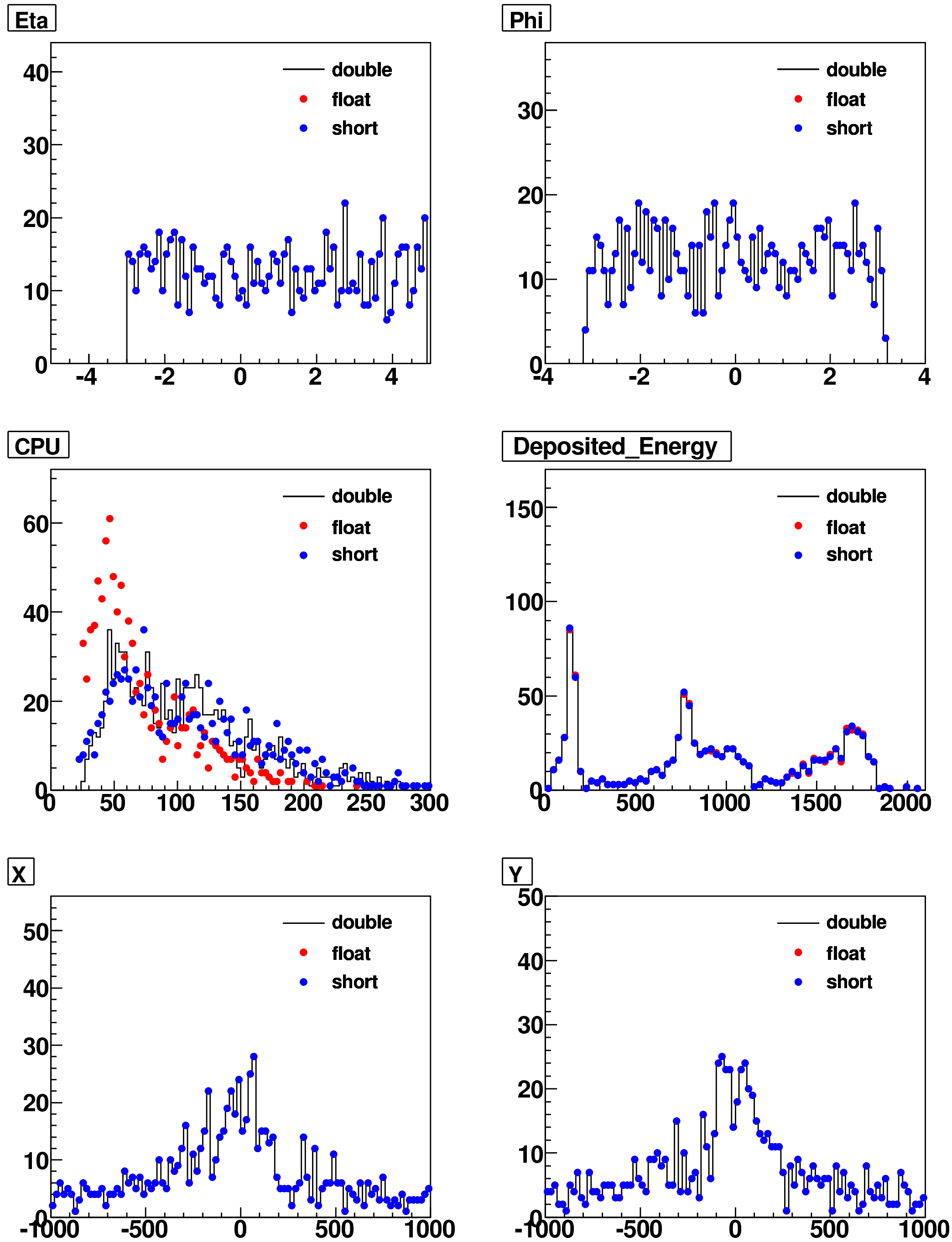

Figure 7: *shower profile for single e*

A.2 single gamma

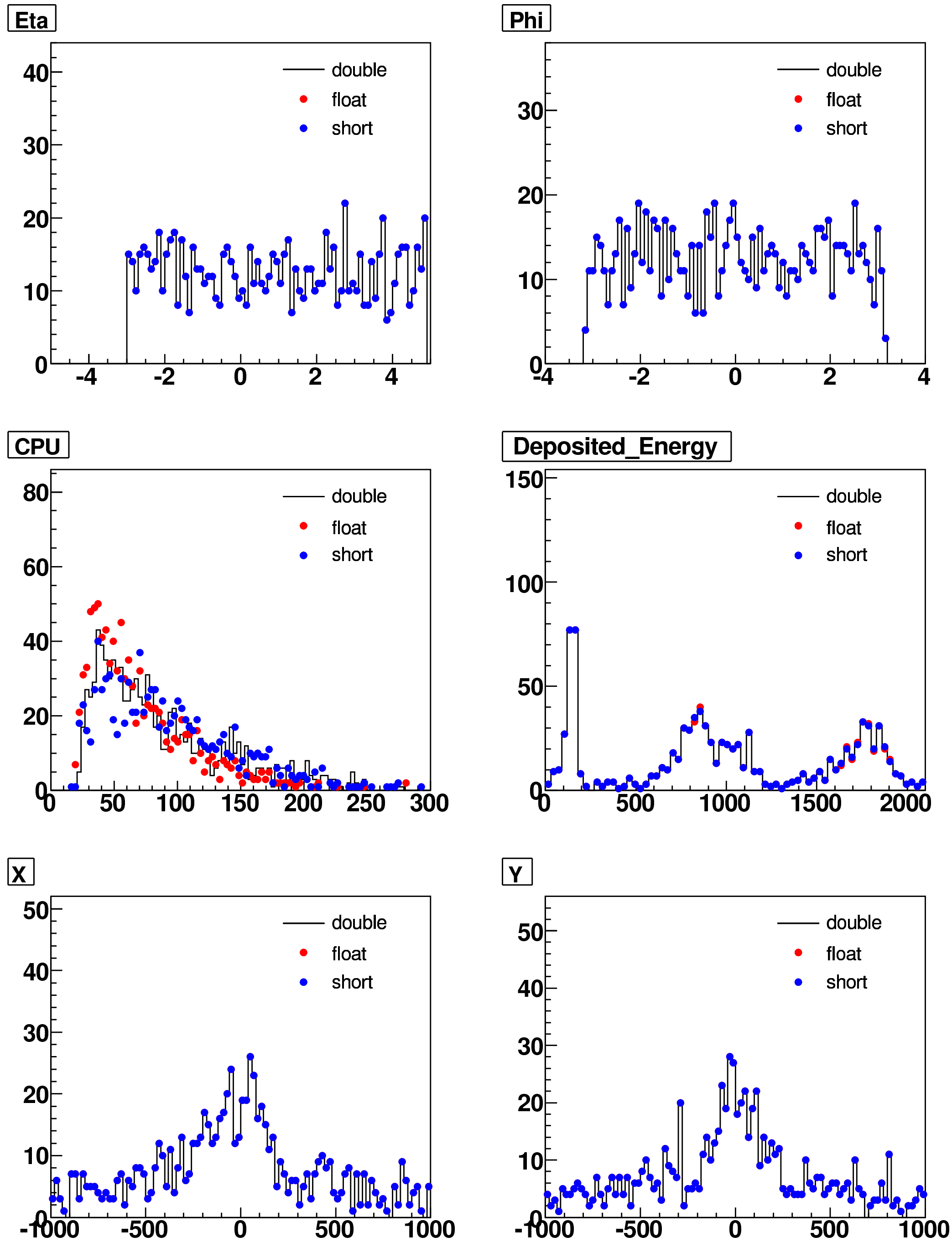

Figure 8: *shower profile for single γ*

A.3 single pions

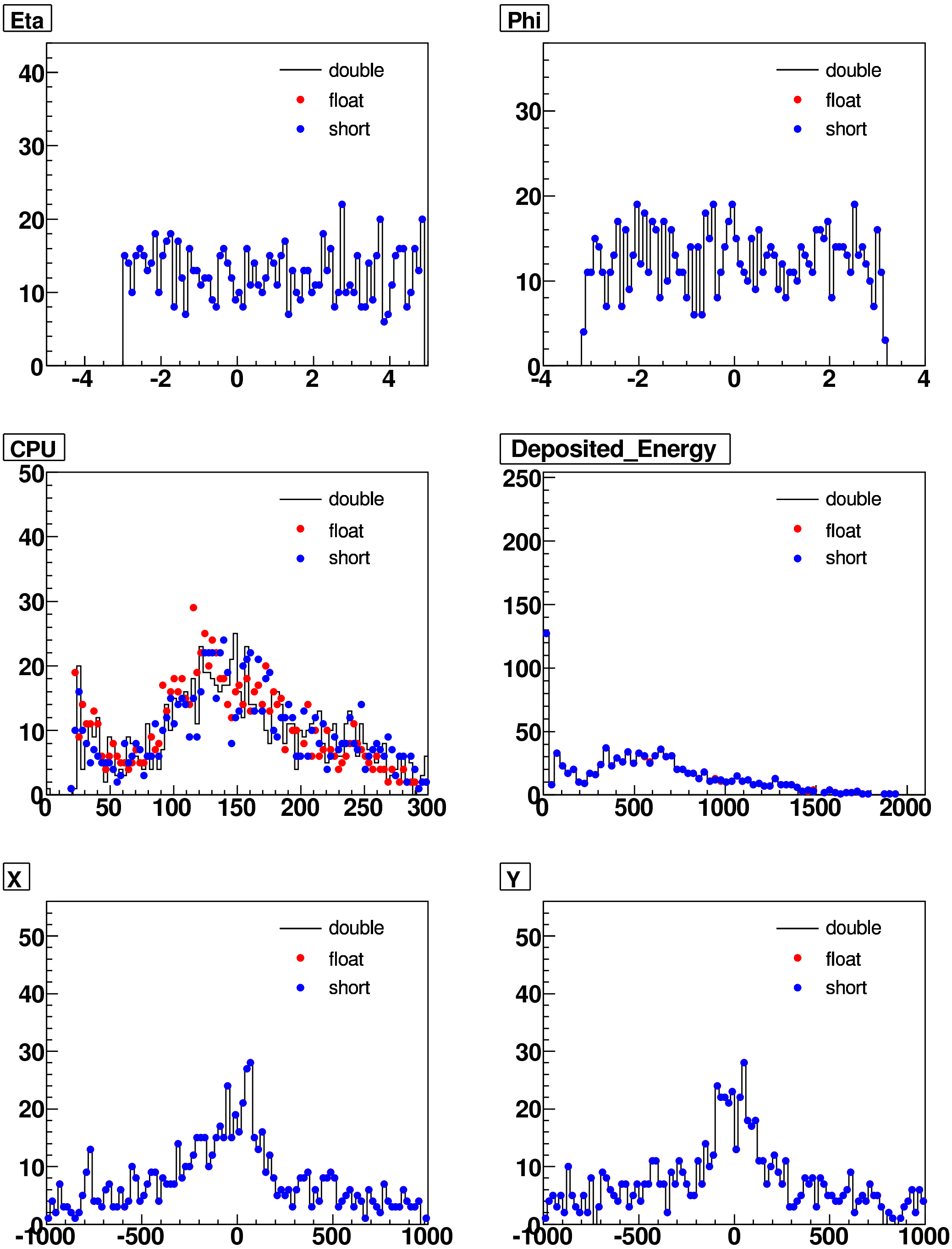

Figure 9: *shower profile for single* π

A.4 $Z \to e^+e^-$

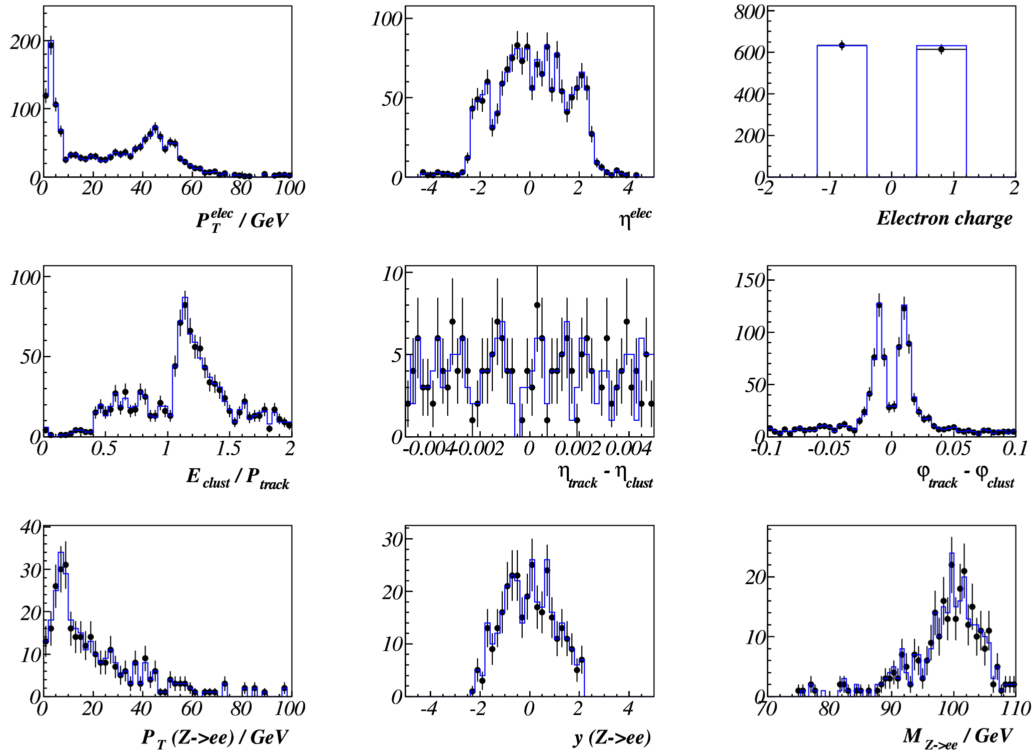

Figure 10: *reconstructed values for* $Z \to e^+e^-$