



ATLAS PUB Note
ATL-SOFT-PUB-2018-002
10th July 2018



The new Fast Calorimeter Simulation in ATLAS

The ATLAS Collaboration

ATLAS relies on very large samples of simulated events for delivering high-quality and competitive physics results, but producing these samples is very CPU intensive when using the full GEANT4 detector simulation. Fast simulation tools are a useful way of reducing CPU requirements when detailed detector simulations are not needed. During Run 1 and 2 of the LHC, a fast calorimeter simulation (FastCaloSim) was successfully used in ATLAS. FastCaloSim provides a simulation of the particle energy response at the calorimeter read-out cell level, taking into account the detailed particle shower shapes and the correlations between the energy depositions in the various calorimeter layers. It is interfaced with the standard ATLAS digitization and reconstruction software, and it can be tuned to data more easily than GEANT4. An improved version of FastCaloSim that incorporates the experience gained with the Run 1 version is currently under development. The new FastCaloSim makes use of machine learning techniques, such as principal component analysis and neural networks, to optimise the amount of information stored in the ATLAS simulation infrastructure. This allows for further performance improvement by reducing the I/O time and the memory usage during the simulation jobs. A prototype is being tested and validated, and it has shown significant improvements in the modelling of cluster-level variables in electromagnetic and hadronic showers.



The simulation of the shower development in the ATLAS calorimeter [1] is very time consuming when using a full detector simulation like Geant4 [2]. About 75% of the simulation time is devoted to modeling shower development in the calorimeters [3]. In the approach of the Fast Calorimeter Simulation (FastCaloSim) [4], the shower development in the calorimeters is replaced by parameterizations obtained from single particle simulations with Geant4 to realistically model the detector response depending on the incoming particle type. An improved version of FastCaloSim, referred to as FastCaloSim V2 (FCS V2), is in development and key elements and first results are presented in this note.

The particles used to derive the parameterizations are photons (γ) and electrons (e^\pm) to represent electromagnetic (EM) showers and charged pions (π^\pm) for hadronic showers. The truth particles are generated on the calorimeter surface following the boundary of the inner detector and the calorimeter in the integrated simulation framework (ISF) [5]. Particles corresponding to discrete values of $\log(\text{energy})$ in the range of 1 GeV to 4 TeV and uniformly distributed in η^1 bins of size 0.05 are currently used for the parametrization. For each sample, up to 10 000 events are generated. The generated events do not take into account the beam spread in the interaction region. The detector response for these true particles is simulated using Geant4 with the standard MC16 Run 2 ATLAS geometry. In the Geant4 simulation step, the detailed spatial position of the energy deposit is saved. In the digitization step of these Geant4 hit outputs, electronic noise, cross talk between neighboring cells and dead cells are turned off.

The energies deposited in the various calorimeter layers as the particle travels through the calorimeter are strongly correlated, which makes it difficult to model the response. The longitudinal energy parametrization is therefore based on a principal component analysis (PCA), which decorrelates these energies. The steps of this PCA-chain are as follows:

- Inputs: Starting from a Geant4 input sample with a fixed energy and a single η bin, the inputs to the PCA are the total energy (summed over all layers) and the fraction of this energy deposited in each layer. Layers are classified as "relevant" when the average energy fraction is above 0.1% per shower in this layer (example in Fig. 1a).
- Transformation: The input distributions are transformed into cumulative distributions (integrating the events over all bins), and then further transformed into Gaussian distributions using the inverse error function and the cumulative distributions (example in Fig. 1b and 1c). The x-axis units of these Gaussian distributions can be interpreted as probability quantiles.
- PCA: The PCA matrix is constructed from these Gaussian distributions. The PCA is a method that converts a set of possibly correlated variables into a set of linearly uncorrelated variables by an orthogonal transformation of the coordinate system. This transformation is computed from the covariance matrix of the inputs and its eigenvectors that are normalized to unity. The basis of the new coordinate system is defined by these unit vectors, and the first vector (ie. the first new coordinate) is chosen such that the data shows the largest variance if projected onto this new axis. The second coordinate is then orthogonal to the first one and the data has the second-highest variance along this new axis, and the third component again is again orthogonal to the first two axes, and so on. The

¹ ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the z -axis along the beam pipe. The x -axis points from the IP to the center of the LHC ring, and the y -axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z -axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

number of PCA components is equal to the number of input variables. The TPrincipal class as implemented in RooT [6] is used to perform this calculation. This step is also referred to as "first PCA".

- After PCA rotation, the leading principal component (PC) then carries the largest variance (example in Fig. 1d). To achieve even better decorrelation, this distribution is used to divide the showers of the input sample into quantiles ("PCA bins"), constructed such that each bin contains the equal number of events, ie. the bins each have the same probability (Fig. 1e).
- Each event in the input sample is sorted into a PCA bin and the PCA chain is performed a second time for events of each bin (also called "second PCA"). This further decorrelates the input energies.

The information that needs to be stored in the parametrization file are the cumulative energy distributions, the PCA matrices from the second PCA transformation, and the mean and rms of the Gaussian distributions after PCA rotation.

To demonstrate the usefulness of the PCA method, the correlations between the energies deposited in a few relevant layers are depicted before and after first PCA transformation, using photons of 65 GeV energy with $0.2 < |\eta| < 0.25$. The correlations are plotted after having been transformed to Gaussian functions, but before the PCA rotation, in Fig. 2. Correlation factors are computed from these 2D histograms, where 1 (or -1) means maximal correlation (or anti-correlation), and 0 means no correlation. Significant correlations are observed. A subset of the output components of the first PCA are then plotted against each other in Fig. 3, and the correlation factors obtained from these 2D histograms are almost 0.

In the fast simulation, the steps of the PCA chain are executed in reverse order. For each simulated particle, a PCA bin is determined from a uniformly distributed random number. Then, starting from uncorrelated Gaussian-distributed random numbers (one for each PCA output component), the numbers are rotated using the inverse PCA matrix. The resulting Gaussians are then transformed into correlated uniform random numbers using the error function, and then from these and the stored cumulative distributions the energy distributions are obtained in each layer. Since the entire parametrization needs to be loaded into memory at the beginning of a simulation job, efficient methods are investigated to reduce the information and therefore memory-consumption of the cumulative distribution, ie. smart rebinning, splines or regression techniques. In the parametrization state presented in this note, smart rebinning is applied.

The longitudinal energy parametrization is presented in Fig. 4 for 1 TeV pions with $0.2 < |\eta| < 0.25$. The parametrization is validated with a toy simulation. This method simulates incoming true pions with 1 TeV energy and then calculates the energy response in the relevant calorimeter layers as well as the total energy. The toy events are not passed through digitization or reconstruction. These toy simulation validations are an efficient method and are typically performed on-the-fly when the parametrization are obtained. The simulated energies are compared to Geant4 and good agreement is observed.

The response of the total energy as a function of the truth energy of the incoming particle is studied for photons, electrons and pions and for a large range of truth energies from 1 GeV to 4 TeV. The results are shown in Fig. 5 as a function of the truth energy, and in Fig. 6 as a function of the truth $|\eta|$. Very good agreement between the simulated responses of the FastCaloSim V2 and Geant4 is found for all tested points. The dips in the energy response for certain η bins is caused by transitions in the detector geometry, changes of material or cell granularity. These transitions are sharper and more distinct for electromagnetic showers as these are narrower compared to those created by pions.

The lateral shower parametrization is derived in each relevant layer for each bin of PCA. The coordinates of the lateral shower are calculated with respect to the extrapolated position of the particle. As the shower is developed in units of millimeter (mm), the coordinates are then transformed to the dimensions in millimeter as shown in Eqn. 1.

$$\begin{aligned}
\Delta\eta &= \eta^{\text{hit}} - \eta^{\text{extr.}}, \\
\Delta\phi &= \phi^{\text{hit}} - \phi^{\text{extr.}}, \\
\Delta\eta^{\text{mm}} &= \Delta\eta \times \eta_{\text{Jacobi, hit}} \times \sqrt{r_{\text{cell}}^2 + z_{\text{cell}}^2}, \\
\Delta\phi^{\text{mm}} &= \Delta\phi \times r_{\text{cell}},
\end{aligned} \tag{1}$$

where, $\eta_{\text{Jacobi}} = |2 \times \exp(-\eta_{\text{cell}}) / (1 + \exp(-2\eta_{\text{cell}}))|$. Here, hit refers to the energy distribution inside a calorimeter cell. Additionally, the symmetry of the shower around the center is exploited by transforming to a new set of coordinates defined in Eqn. 2

$$\begin{aligned}
r^{\text{mm}} &= \sqrt{(\Delta\eta^{\text{mm}})^2 + (\Delta\phi^{\text{mm}})^2}, \\
\alpha &= \arctan2(\Delta\phi^{\text{mm}}, \Delta\eta^{\text{mm}}).
\end{aligned} \tag{2}$$

The shower development is then parametrized in these coordinates and stored in a 2D histogram. Figure 8 shows the polar representation of two such parametrized showers corresponding to EM and hadronic shower in the second layer of the EM barrel and the Tile barrel, respectively. The memory footprint of these 2D histograms is reduced by storing only the $|\phi|$ coordinates ($0 \leq \alpha \leq \pi$). Since the shower is expected to be symmetric in ϕ , this is a good approximation. Figure 9 shows one such parametrization in the polar representation. At the simulation step, hit coordinates are sampled randomly from these histograms and additionally $+\phi$ or $-\phi$ is also assigned randomly. These hits are assigned to calorimeter cells assuming a simplified cuboid geometry. However, such simplification results in energy deposits in wrong cells due to the accordion structure of the EM calorimeter. This effect is clearly seen in Fig. 10a where the energy deposit in each cell is compared to the corresponding Geant4 energy assignment. A function derived considering the accordion structure is used to displace the hits to neighboring cells. The effect of this correction is demonstrated in Fig. 10b. The performance of the parametrization is tested using toy simulation by generating large number of lateral shower distributions from the parametrization and comparing to the Geant4 distribution at the cell level. Figure 11 shows one such validation, where FastCaloSim V2 generated hits are assigned to cells with accordion correction and compared to Geant4 distribution.

The simulation infrastructure for FastCaloSim V2, known as FastCaloSimSvcV2 is implemented inside the ISF and used to simulate the particles. The simulation starts when ISF hands over a particle at the calorimeter boundary from the inner detector. Based on the particle type, energy and η a parametrization is selected to generate detector response to the truth particle. The CPU requirement for FastCaloSim V2 is compared to Geant4, as well as the current ATLAS fast simulation, ATLFASSTII (AF2). In performing this study, single photons in the range $0.2 < |\eta| < 0.25$ and with energies of 8 GeV, 65 GeV or 256 GeV are generated on the calorimeter surface and processed through the different Geant4, AF2 and FCS V2 in the ATLAS software release Athena, 21.0.73. For each energy point, 100 events are used to calculate the average CPU time to simulate an event. Figure 12 shows that FastCaloSim V2 has a performance

compared to the current AF2. The gain in CPU time compared to Geant4 depends on the energy and ranges from about approximately a factor 10 at $E = 8$ GeV to a factor 25 for $E = 256$ GeV. No optimization and profiling has been performed for FastCaloSimSvcV2 yet.

The performance of FastCaloSim V2 is studied by calculating various electromagnetic shower and hadronic cluster variables in simulated single photon and pion events of various energy and η regions and comparing them to the corresponding Geant4 variables. These variables are used in the algorithms for photon and jet identification. Figures 13 and 14 show the comparison for electromagnetic and hadronic showers, respectively, for various particle energies. An energy interpolation is performed for the energies of 44 and 200 GeV. In general, good modeling is observed.

The FastCaloSimSvcV2 can also process events with multiple particles. As a first validation a sample of $H \rightarrow \gamma\gamma$ MC events has been processed. The parametrization used for this study is restricted to $|\eta| < 0.6$. The MC events were generated with POWHEG-Box interfaced with PYTHIA8. One event display that demonstrates the functionality of the service is presented in 15. It shows an event where the Higgs boson decays to two photons which are "back-to-back", one of the photon converts and a high- p_T track is reconstructed.

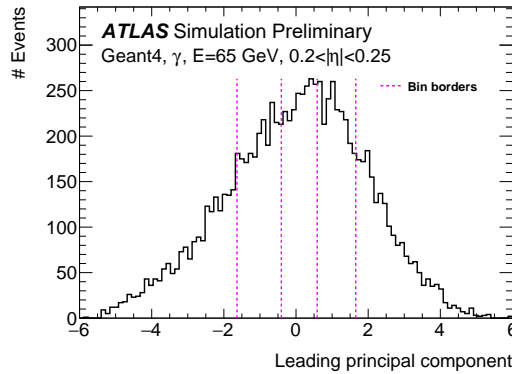
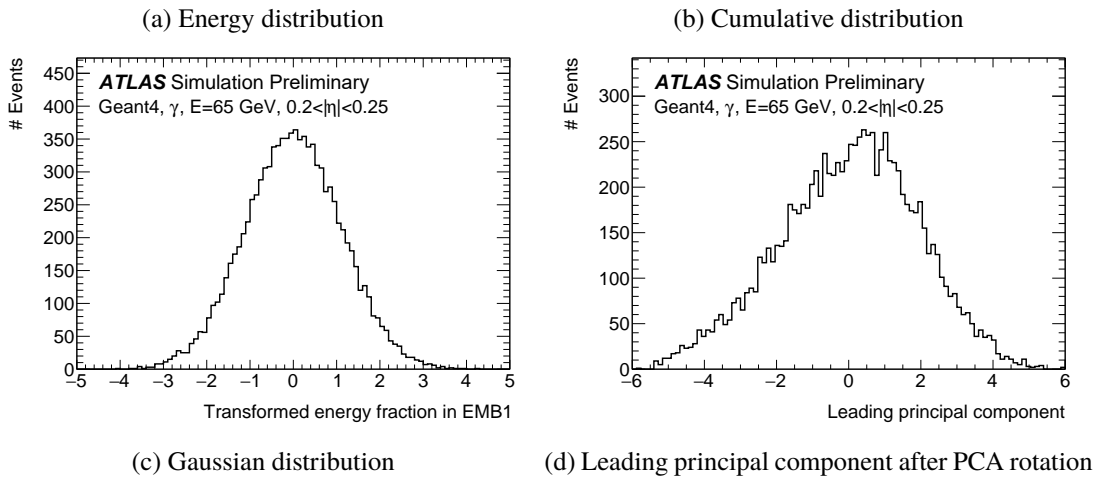
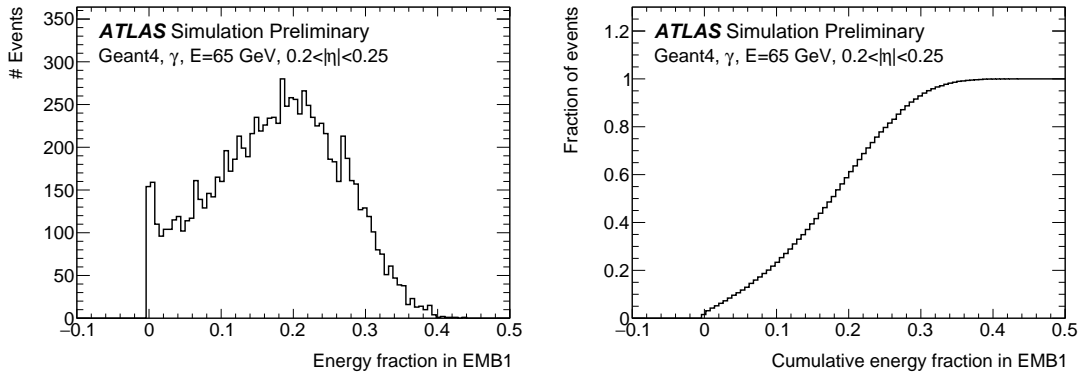


Figure 1: The steps of the PCA chain for one input sample of central photons with 65 GeV energy. (a) the distribution of the fractional energy in EM barrel 1 (termed EMB1), (b) the cumulative distribution, (c) transformed into a Gaussian distribution, and (d) the leading principal component after PCA transformation. This distribution is then used to divide the dataset of the Geant4 showers into several subsets. These "PCA bins" are depicted in (e) in pink, five bins are defined in this example, each with the same (flat) probability.

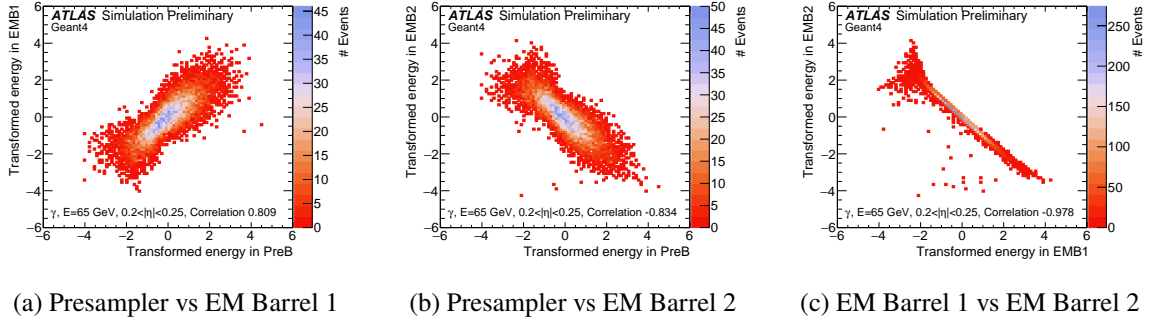


Figure 2: Correlations between the transformed energies deposited in several layers, before PCA rotation, showing (a) pre-sampler barrel vs. EM barrel 1, (b) pre-sampler vs. EM barrel 2 and (c) EM barrel 1 vs. EM barrel 2. The energies were transformed into Gaussian distributions. The correlation factors obtained from these 2D histograms are displayed in the plots.

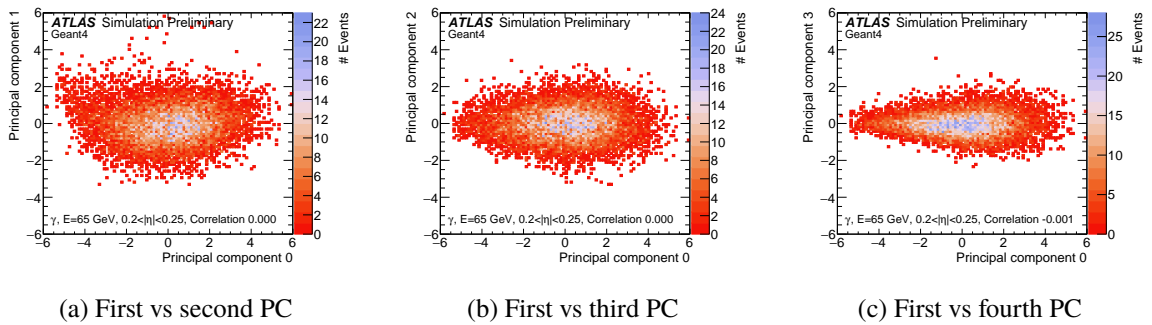
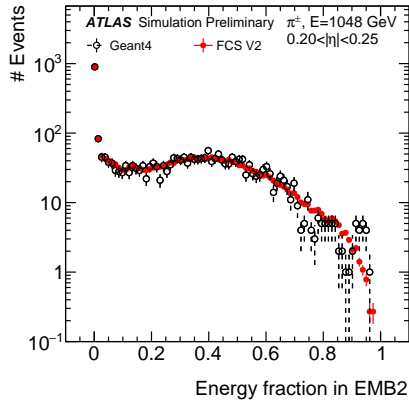
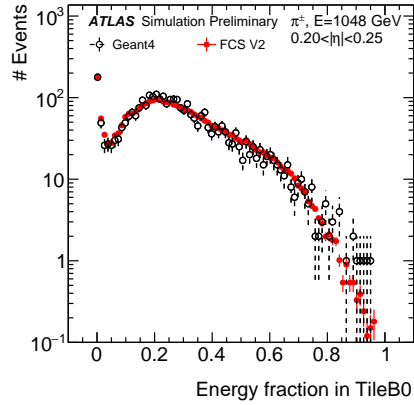


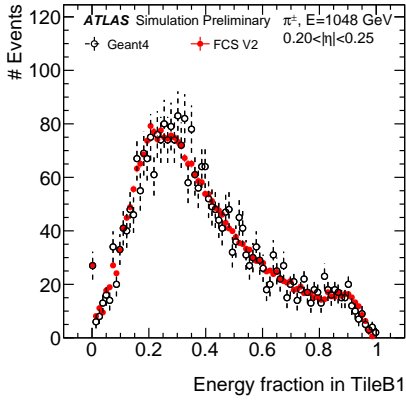
Figure 3: Correlations between the components after PCA rotation. The individual components are approximately Gaussian distributed. After PCA transformation, the correlation factors are consistent with zero.



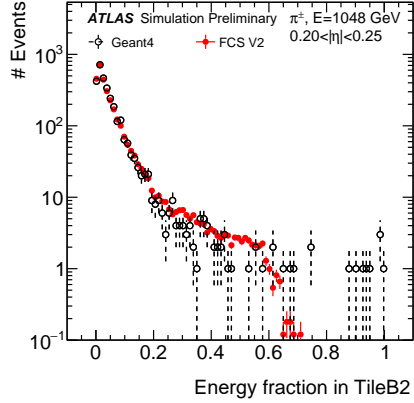
(a) Energy in EM barrel 2



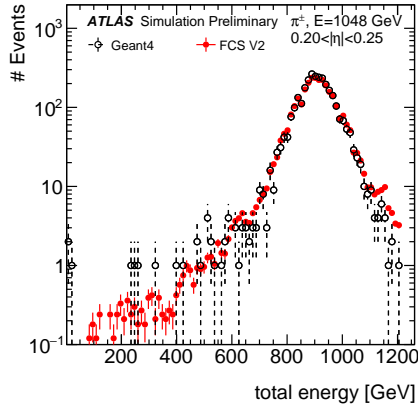
(b) Energy in Tile barrel 0



(c) Energy in Tile barrel 1

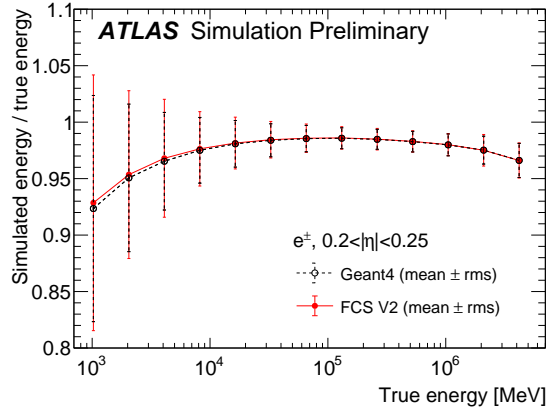


(d) Energy in Tile barrel 2

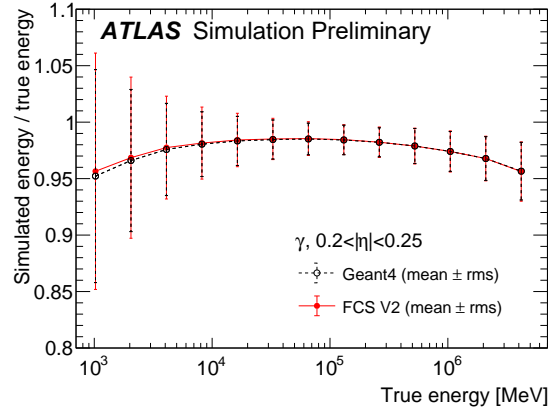


(e) Total energy

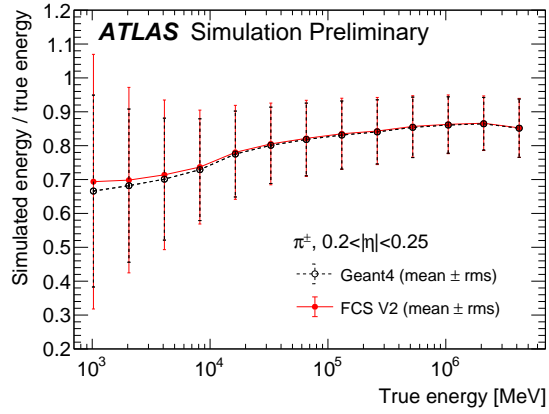
Figure 4: Longitudinal energy parametrisation and validation against Geant4 using a toy simulation for 1 TeV pions with $0.2 < |\eta| < 0.25$. A very good agreement between both simulations is seen in all distributions.



(a) Electrons

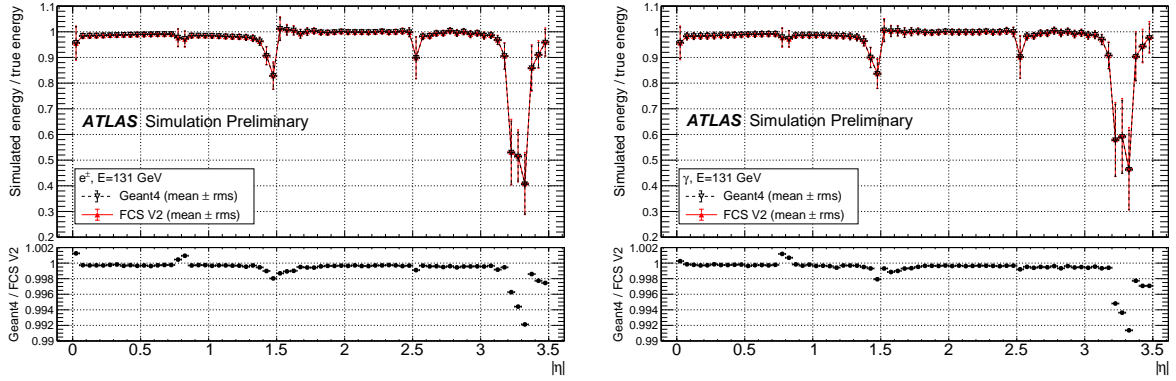


(b) Photons



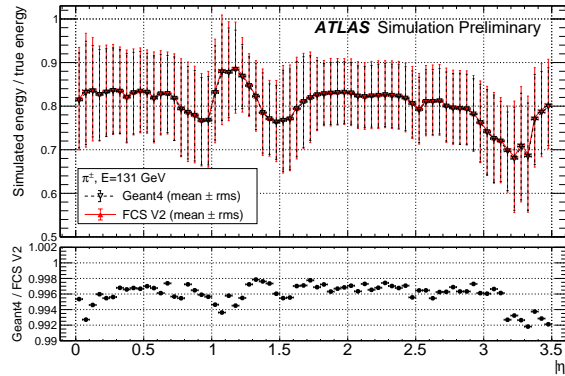
(c) Pions

Figure 5: The total energy response for (a) electrons, (b) photons and (c) pions for truth energies from 1 GeV to 4 TeV for $0.2 < |\eta| < 0.25$. The response is defined as the ratio of measured energy to truth energy. The comparison shows the mean (as points) and the rms (as error bars) of the total energy distributions for the new FastCaloSim (in red) and Geant4 in black.



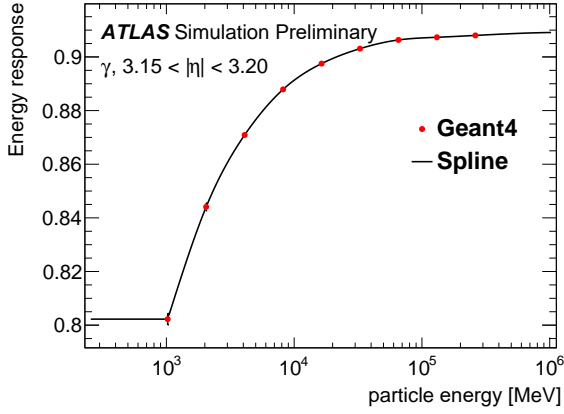
(a) Electrons

(b) Photons

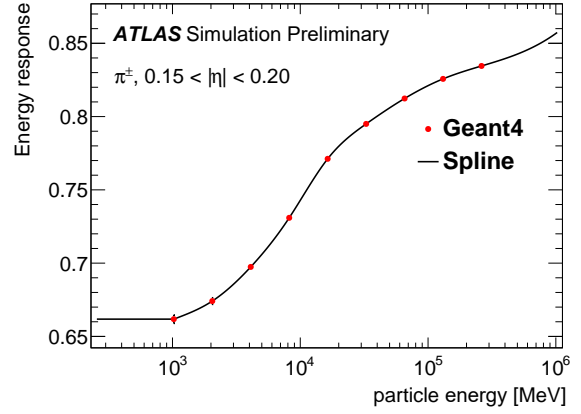


(c) Pions

Figure 6: The total energy response for (a) electrons, (b) photons and (c) pions with energy of 131 GeV for all eta bins covering values between 0 and 3.5. The FCAL is not included here. The response is defined as the ratio of measured energy to truth energy. The comparison shows the mean and the rms of the total energy distributions for the new FastCaloSim (red) and Geant4 (black). The lower panels show the ratio of the response from Geant4 to that of the new FastCaloSim.

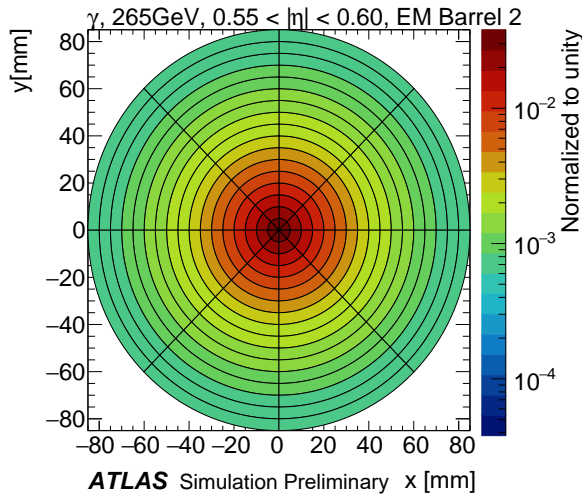


(a) photon

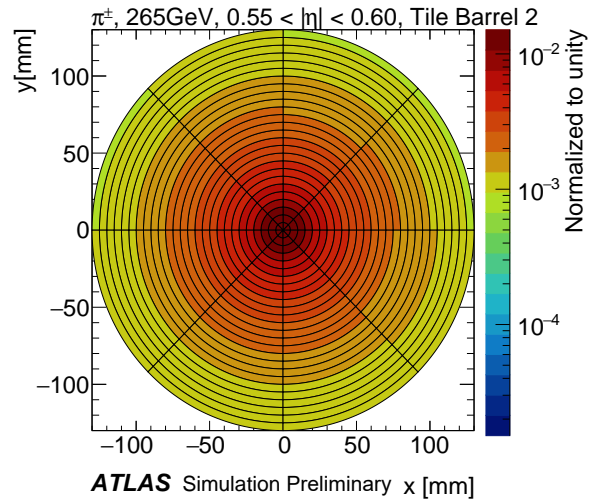


(b) pion

Figure 7: Energy response, defined as the ratio of the reconstructed energy in the calorimeter cells to the energy of the particle, for (a) photons in $3.15 < |\eta| < 3.20$ and (b) pions in $0.15 < |\eta| < 0.20$. The red dotted point represents the response derived at discrete energies, using Geant4 simulated single particles. The black line is a spline fit used to interpolate between discrete energy points.



(a) photon



(b) pion

Figure 8: The lateral shower development of (a) photons and (b) pions of energy 265 GeV in the range $0.55 < |\eta| < 0.60$ parametrized in the second layer of EM barrel and Tile barrel respectively.

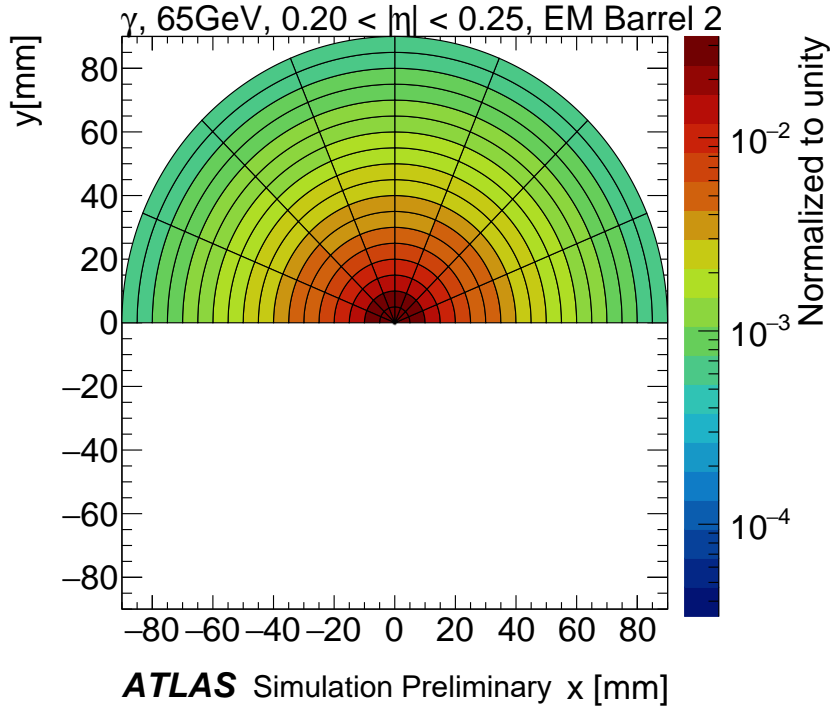


Figure 9: The symmetry of shower in $\pm\phi$ -direction is exploited and only $|\phi|$ coordinate of the shower is used in Eqn. 2 leading to $0 \leq \alpha < \pi$. Thus only the upper half of the parametrization is necessary to store.

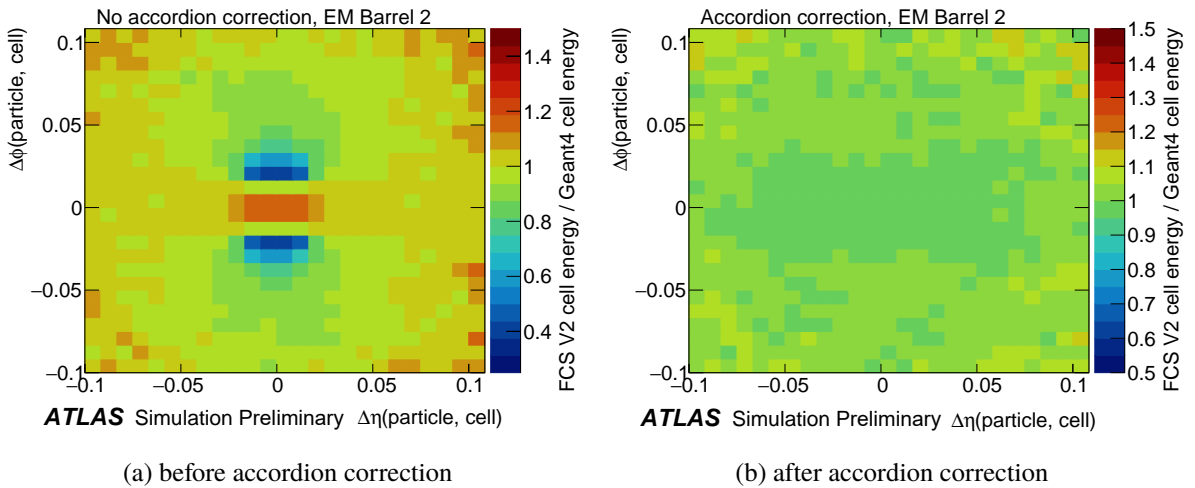


Figure 10: The difference in the energy assigned to each cell of the second electromagnetic (EM) barrel layer using fast simulation (FCSV2) for a photon of 65 GeV in the range $0.20 < |\eta| < 0.25$ compared to Geant4 (a) considering a simplified cuboid geometry (b) after applying a correction for accordion geometry.

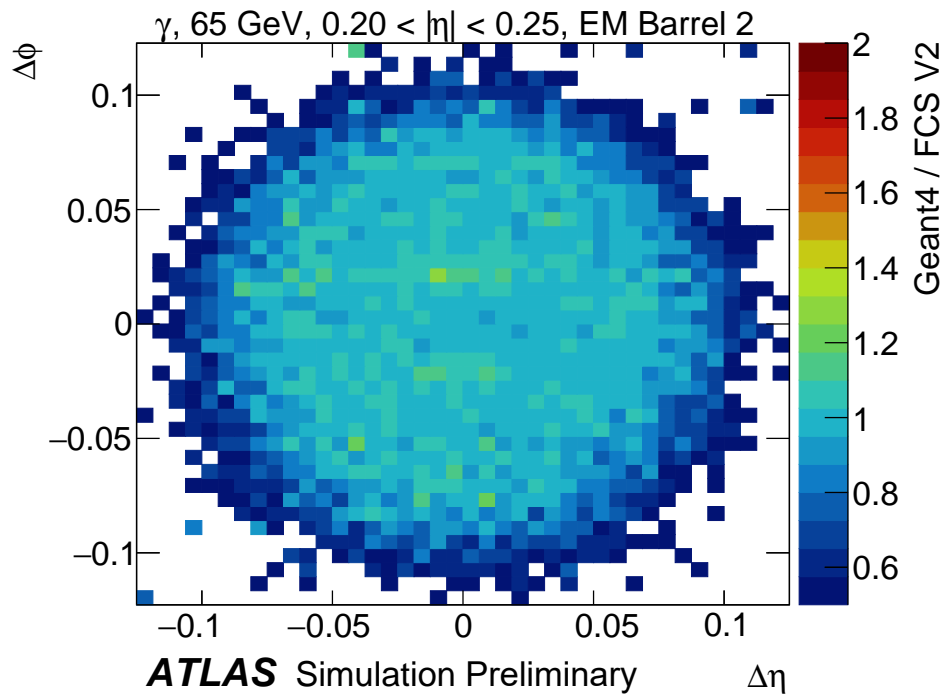


Figure 11: Comparison of the lateral shower parametrization derived for a photon of 65 GeV in the range $0.20 < |\eta| < 0.25$. Presented is the ratio of the shower shape derived from Geant4 to that simulated with FCSV2. The coordinates of the lateral shower ($\Delta\eta$, $\Delta\phi$) are calculated with respect to the position of the incident particle on the calorimeter surface along the layer.

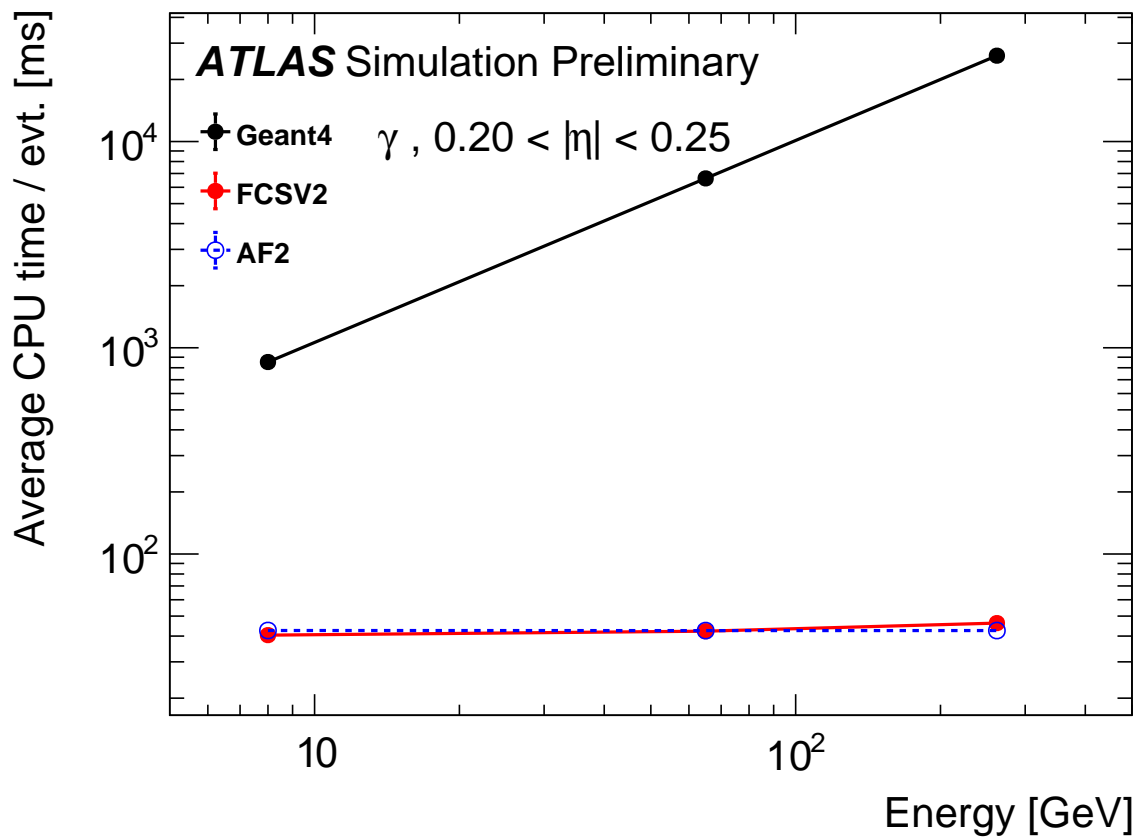
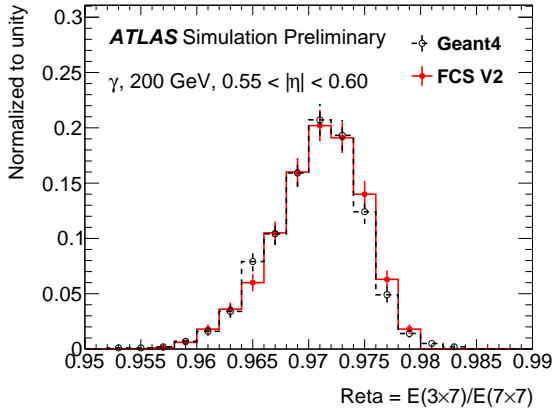
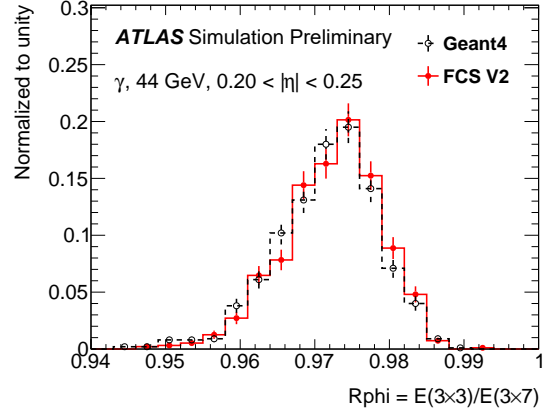


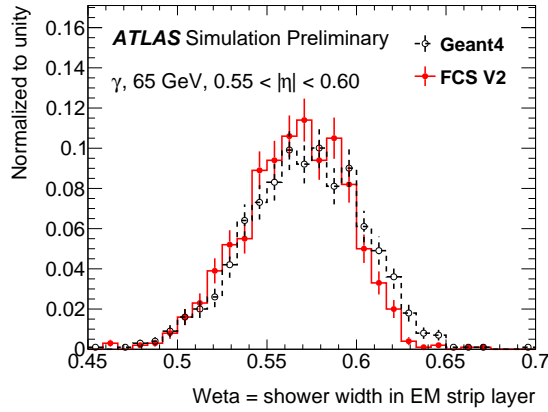
Figure 12: CPU time to simulate photons of 8 GeV, 65 GeV and 256 GeV in the range $0.20 < |\eta| < 0.25$ using Geant4 (black), FCSV2 (red) and AF2 (blue open circle). The average time is calculated by generating 100 events for each simulation type using the ATLAS software release Athena, 21.0.73.



(a) Reta

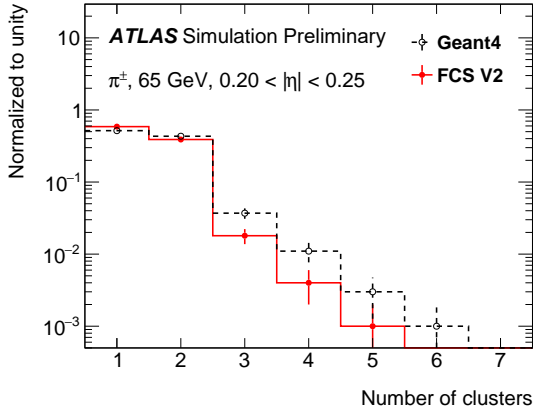


(b) Rphi

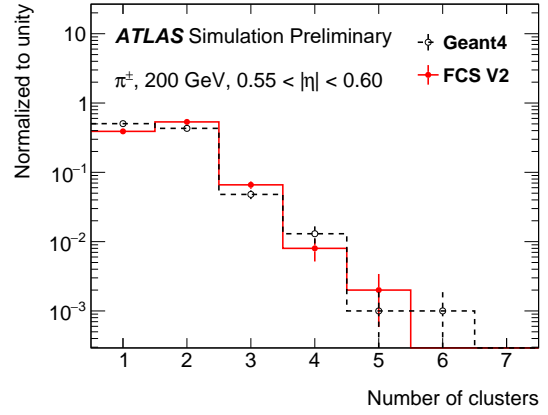


(c) Weta

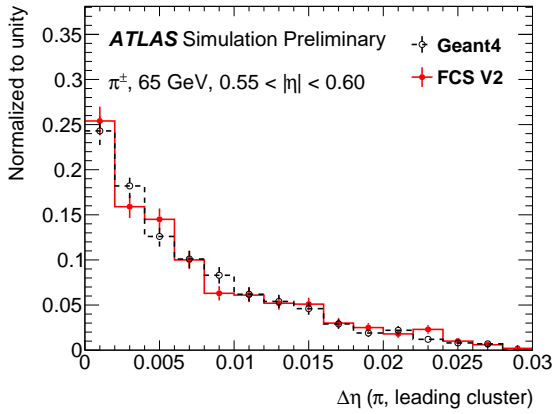
Figure 13: Lateral shower shape variables calculated using FCSV2 (red solid line) and compared to Geant4 (black dashed line): (a) fractional energy deposit in the η direction for the second EM barrel layer for a photon of 200 GeV in the range $0.55 < |\eta| < 0.60$, (b) fractional energy deposit in the ϕ direction for the second EM barrel layer for a photon of 44 GeV in the range $0.20 < |\eta| < 0.25$ and (c) fractional energy around the maximum energy deposit in the strip EM barrel layer for a photon of 65 GeV in the range $0.55 < |\eta| < 0.60$.



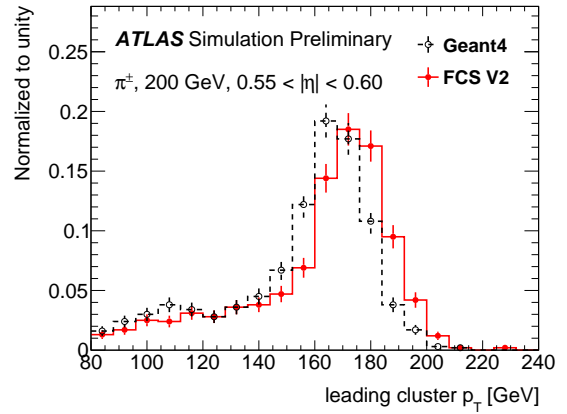
(a) No. of clusters



(b) No. of clusters



(c) $\Delta\eta$



(d) p_T

Figure 14: Cluster variables related to single charged pion shower: (a) and (b) show the number of clusters for energies 65 GeV in $0.20 < |\eta| < 0.25$ and 200 GeV in $0.55 < |\eta| < 0.60$ respectively. (c) distance in η -direction between the pion and the leading constituent cluster. (d) transverse momentum calculated from the leading constituent cluster.

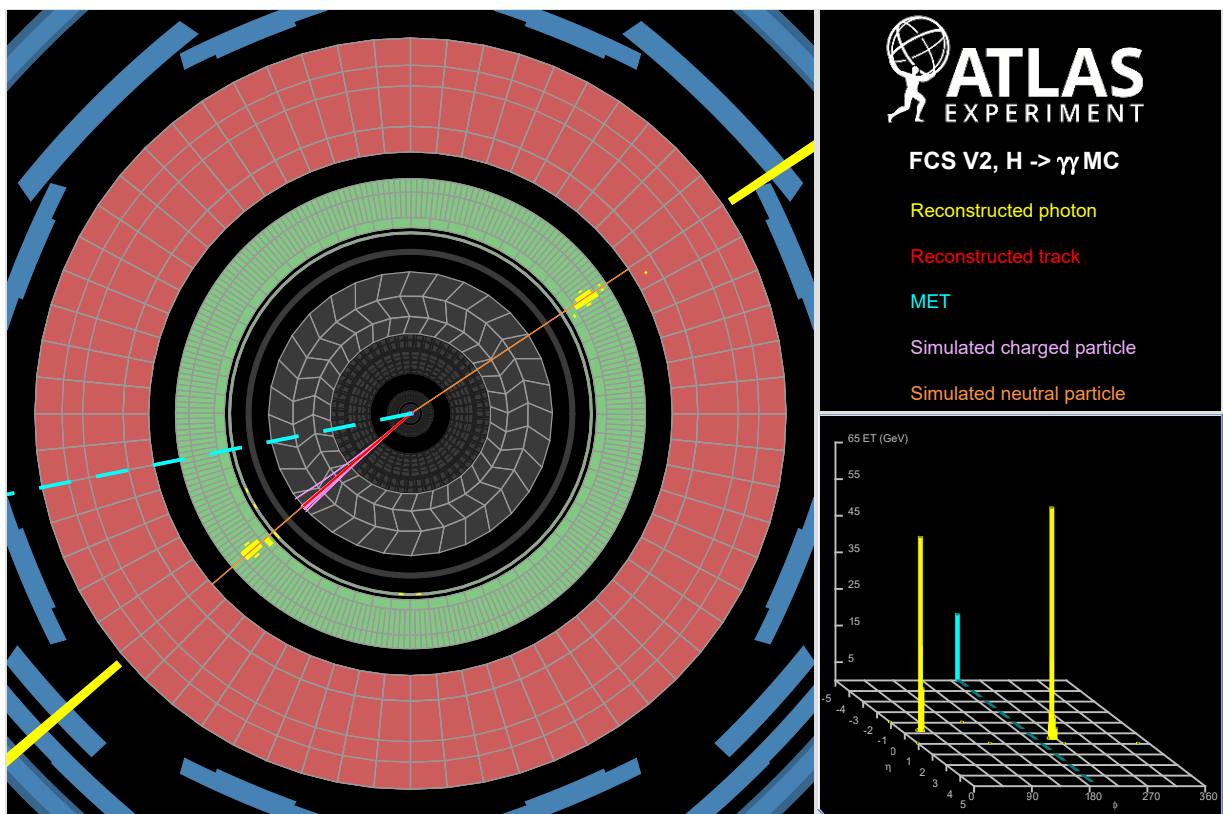


Figure 15: Event display of a $H \rightarrow \gamma\gamma$ MC event that was simulated with FastCaloSim V2. Tracks are displayed if their p_T exceeds 3 GeV. The left photon is converted, while the right one is unconverted.

References

- [1] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.
- [2] Agostinelli, S. et al., *GEANT4: a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.
- [3] ATLAS Collaboration, *The ATLAS simulation infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: [1005.4568](https://arxiv.org/abs/1005.4568) [[physics.ins-det](https://arxiv.org/abs/1005.4568)].
- [4] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, (2010), URL: <https://cds.cern.ch/record/1300517>.
- [5] Ritsch, E. on behalf of the ATLAS collaboration, *The ATLAS Integrated Simulation Framework*, (2013), URL: <https://cds.cern.ch/record/1532476>.
- [6] Brun, R. and Rademakers, F., *ROOT - An Object Oriented Data Analysis Framework*, Proceedings AIHENP'96 Workshop, Lausanne, Nucl. Inst. Meth. in Phys. Res. A **389** (1997) 81.