

# 1 Chapter 1

## 2 Monte Carlo

3 The Monte Carlo(MC) method was invented by scientists working on the atomic bomb in the 1940s.  
4 Its core idea is to use random samples of parameters or inputs to explore the behavior of a complex  
5 system or process. Nowadays, MC are essential part of both theoretical and experimental particle  
6 physics research. This chapter gives an overview of ATLAS experiment simulation scheme, simulation  
7 methods and software used. Also, a techniques for fast simulation will be discussed.

### 8 1.1 Monte Carlo simulation at ATLAS experiment

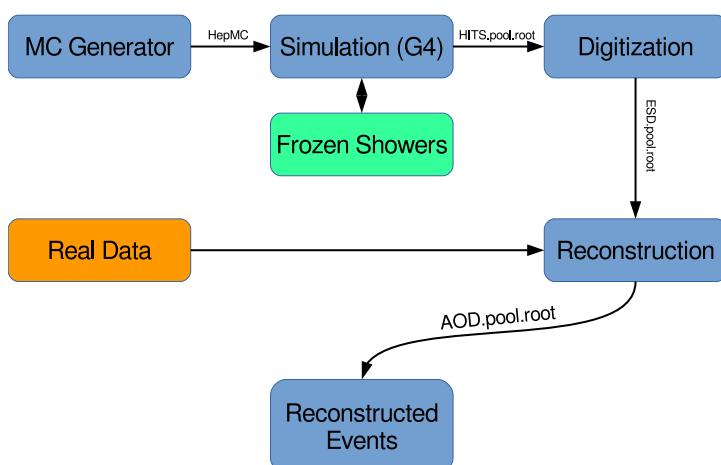


Fig. 1.1: Diagram of the ATLAS MC production chain

9 MC is allowing to make different analysis, such as compare data with predictions, study detector  
10 or selection algorithms performance. All of this applications are requiring MC precision. Simulation  
11 software expects to use precise physics models for sampling and have large enough statistics, to  
12 exclude statistical uncertainties (usually 5-10 times more, than expected in a data). ATLAS simulation  
13 software is integrated into Athena and usually used during large production of events. Simulation  
14 chain is generally divided into 4 main steps (Figure 1.1):

15 **Event generation** Simulation of hard interaction and a resulting high-energy particles parameters.  
16 This step is independent of ATLAS detector geometry.

17 **Simulation** Simulation of energy depositions ("hits") done by final state particles in ATLAS detector.

18    **Digitalization** Simulation of detector response using "hits" information: first, inputs to the read out  
 19       drivers (ROD's), called "digits" are constructed, then, ROD functionality is emulated. Detector  
 20       noise effects are added at this stage.

21    **Reconstruction** Production of the Analysis Object Data (AOD) files, which are containing sufficient  
 22       information for physics analysis. This stage is identical for both data and MC

23    This scheme is allowing to use computing resources efficiently, than with a single-step simulation,  
 24       and simplifies software validation, since it is possible to reuse files from previous stages. In the  
 25       following sections event generation and simulation will be described more detailed

26    **1.1.1 Event generators**

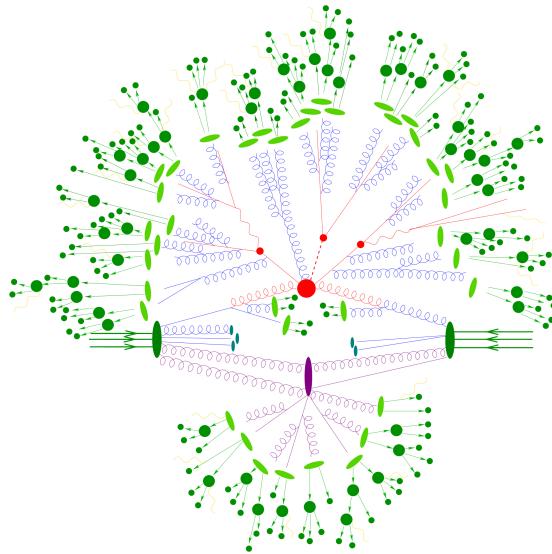


Fig. 1.2: Schematic view of a  $t\bar{t}H$  event produced in a  $pp$ -collision: the hard scattering is shown as a red blob with the solid and dashed lines as the resulting three particles. Independently happening multi-particle interactions are indicated by the violet blob. Parton showers are shown with curly lines. Hadronization yields hadrons as shown in light green, while the final state particle are dark green.

c

27    The outcome of the hard interaction could be simple scattering, their annihilation or a combination  
 28       of two. In any case, the final state consist of a large multiplicity particles. The main goal of event  
 29       generator is to provide a complete picture of this final states: description of the particle types and  
 30       momentia on event-by-event basis. The factorisation theorem allows to make event generation in  
 31       independent stages, which are dominated by different dynamics:

32    **Modelling of hard subprocess** Hard subprocess is happening at the smallest scales of times and  
 33       distance, where all of the colliding partons are considered free and their initial momentia are  
 34       given by selected PDF set. Process of interest is simulated by selecting production channels and  
 35       calculating corresponding matrix elements (ME) in the desired level of accuracy in perturbation  
 36       theory . Most of the generators have leading order or next to leading order ME in  $\alpha_s$ . Scale of  $\alpha_s$

37       is can be determined through renormalisation of factorisation. Also, inclusion of higher-order  
38       correction reduces scale dependency, however it makes calculations technically more difficult.

39       **Parton showering** Quarks and gluons from hard process can QCD radiation, resulting on the  
40       dozens of additional partons associated with the event. This process calculated as step-by-  
41       step evolution of momentum transfer scales from highest (hard subprocess), to the lowest  
42       (around 1 GeV). There is a possibility of double counting between showers and hard subpro-  
43       cess. This can be avoided using matching approach, for which higher order corrections to ME  
44       are integrated with parton showers, or merging strategy, there jet resolution scale is used as  
45       an threshold between matrix elements and parton showers.

46       **Hadronisation** Final stable color-neutral particles, what can be detected in experiment, are formed  
47       during hadronisation. This occurs at larger nonperturbative scales and usually implemented  
48       using different phenomenological models.

49       **Modelling underlying event** Parallel to the main process other collisions of partons can occur,  
50       called underlying event. These additional interaction can produce partons which contribute to  
51       the final state. This is one of the least understood aspect of hadronic collisions.

52       Schematic plan of simulation of ttbar event is shown on a figure ???. The hard scattering itself  
53       is shown as a red blob with the solid and dashed lines indicating the resulting particles, which  
54       themselves decay further. Underlying event is indicated by the violet blob. Parton showers are shown  
55       with curly lines. Hadronization yields hadrons as shown in light green, while the final state particles  
56       are dark green.

57       The current analysis uses samples generated with the following generators:

58       **Powheg** Powheg is generator with NLO ME, that can be interfaced with other generator(such as Pythia  
59       or Herwig) for higher precision of showering.

60       **Pythia** Pythia is a general purpose generator for hadronic, hadron-lepton and leptonic collisions. It  
61       can model initial and final state showers, hadronisation and decays, underlying event (via multi  
62       parton interactions). Pythia contains library with around 240 processes with LO ME. It uses  
63       Lund String model for hadronisation.

64       **Herwig** Herwig is a LO general purpose event generator for simulation lepton-lepton, hadron-lepton  
65       and hadron-hadron collisions. The main difference between Pythia and Herwig is that it is  
66       uses angular ordering in the parton showers and also models the hadronisation step based on  
67       the cluster fragmentation

68       **Sherpa** Sherpa is a generator with tree level of matrix elements, featuring its own implementation  
69       of parton shower and hadronisation models.

70       **Photos** Precision tool for QED radiative corrections in W and Z decays.

71       **Tauola** Generator, used to describe leptonic ans semi-leptonic  $\tau$ -decays.

## 72       1.1.2 Simulation in Geant4

73       After event generation, simulation software obtains hardware response for a final state particles.  
74       The main method used by ATLAS, referred to as *Full Simulation*, makes use of the Geant4 [?]. It is

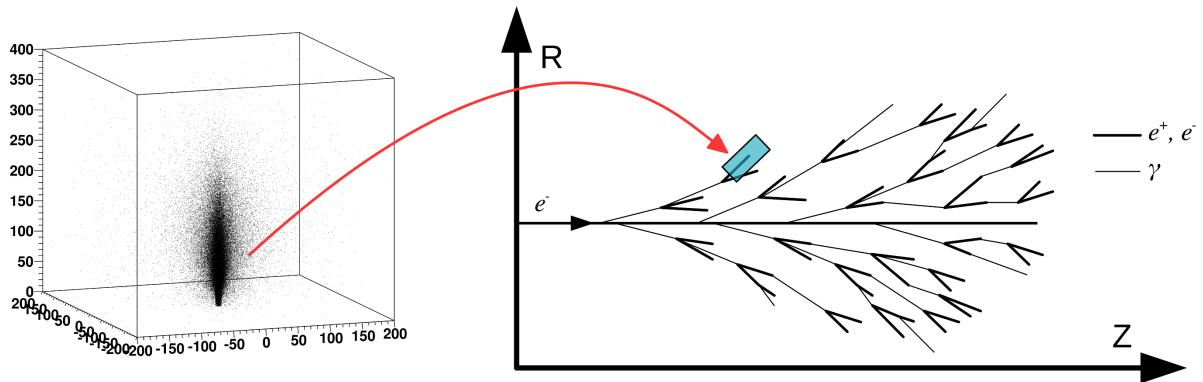


Fig. 1.3: Diagram showing the shower substitution of the low-energy particle, during the high-energy particle simulation.

75 C++ based toolkit for the simulation of the passage of particles through matter. It is used in a wide  
76 range of experiments in a high energy and nuclear physics.

77 Geant4 can simulate complex detector structure with sensitive detector material and correspond-  
78 ing infrastructure. It can also calculate basic properties of materials, like radiation and interaction  
79 length. For detector Geant4 stores "hits" information - snapshots of physical interactions. In Geant4  
80 events and particles are simulated separately and each particle is moved step by step. Size of this  
81 step is chosen to preserve both CPU performance and required precision. Physics processes are  
82 handled as a discrete either at rest, along step or after it. Geant4 package has different models  
83 and approximations of hadronic and electromagnetic processes. Some of them are not so precise,  
84 but computationally fast. It allows to choose set of the models, called physics list, depending on a  
85 case needs. There are several reference physics lists, that are validated for each new release. ATLAS  
86 experiment is using one of this lists.

87 It is necessary to have a mass MC production for each data taking, what is taking most of the  
88 resources. Some of the Run-1 analysis are dominated by available MC statistics. It is possible to  
89 improve in CPU usage by tuning physics list or reworking of B-field access. Also there are long  
90 term developments for multi-threading and vectorisation of the code. Yet, Run-2 has a higher pileup  
91 and luminosity, so even more MC events are needed. This means that fast and accurate simulation  
92 approach is essential. During simulation largest time is spend on calorimeters. This is the motivation  
93 for development of fast calorimetry techniques.

94 There are two main methods used in ATLAS:

- 95 • Parametrisation of calorimeter cells response. Spacial energy response is simulated using lon-  
96 gitudinal and lateral energy profiles.
- 97 • Frozen Showers. This technique will be described more detailed in the following section

### 98 1.1.3 Frozen Showers

99 Main principle of this library is described by its name. It is using pre-simulated "frozen" showers  
100 generated in a full simulation. Particles below minimum energy thresholds are killed and replaced with  
101 with this showers. All of the other particles are using full simulation. This process is schematically  
102 shown on a Figure 1.3.

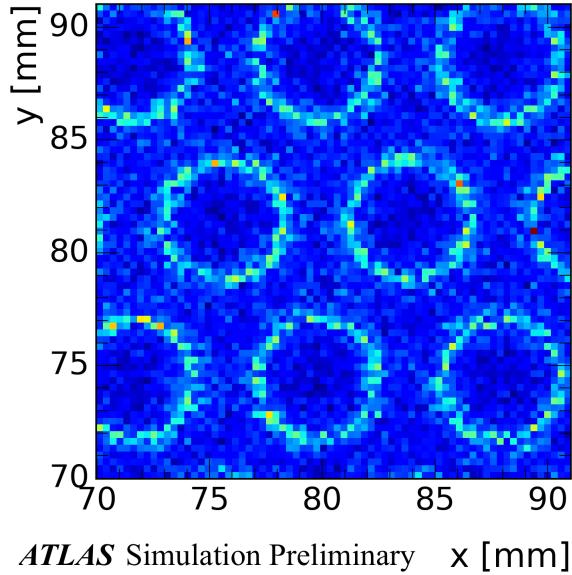


Fig. 1.4: Energy response for electrons in a calorimeter in a x vs y plane and in a distance plane

103     The library itself organized as follows: it contains basic simulation parameters, like Geant4, geom-  
 104     etry and atlas software release version and physics list used. It stores information about position of  
 105     initial particle in a bins, while energy remains unbinned. Each shower stores lateral and transverse  
 106     size and information about all of the hits: four vector (time and position) and energy deposition.

### 107     Production use of Frozen Showers

108     During a simulation, if an energy of particle falls below cut-off energy, particle algorithm examines  
 109     resulting shower containment. It checks that particle is far from the edges of calorimeter, so what  
 110     shower will be by 90% inside calorimeter. This depends also on a energy of particle, because  
 111     shower sizes are growing with energy. When particle is removed and substituted by shower taken  
 112     from corresponding eta and distance bin with the closest energy found. Energies of the hits in shower  
 113     found are scaled to fully correspond to particle energy. Additionally, shower direction is changed to  
 114     the direction of the shower.

Frozen Showers have been used in Atlas Monte-Carlo production since run-1. This method is applicable for all LAr calorimeters in ATLAS, but currently it is enabled for simulation of forward calorimeters (FCAL), since it is showing the smallest differences, compared to the other fast simulation methods (e.g parametrisation). This is because of large of non-uniformly distributed sensitive material, which is giving different responce, than a dead material (Figure 1.5). Resolution of a calorimeter can be written as:

$$\frac{\sigma}{E} \approx \frac{1}{\sqrt{E}} \oplus \frac{1}{E} \oplus const, \quad (1.1)$$

115     where symbol  $\oplus$  indicates a quadratic sum. The first term is 'stochastic term', which is including  
 116     intrinsic shower fluctuations, second takes into account readout noise effects and pile-up fluctua-  
 117     tions. Constant term derives from non-uniformities in a detector, what are causing large fluctuation  
 118     of the energy loss. Resolution of high-energy electrons is mostly dominated by this term. Example  
 119     of this fluctuations is shown on a Figure 1.4. It can be seen, that particles inside sensitive material

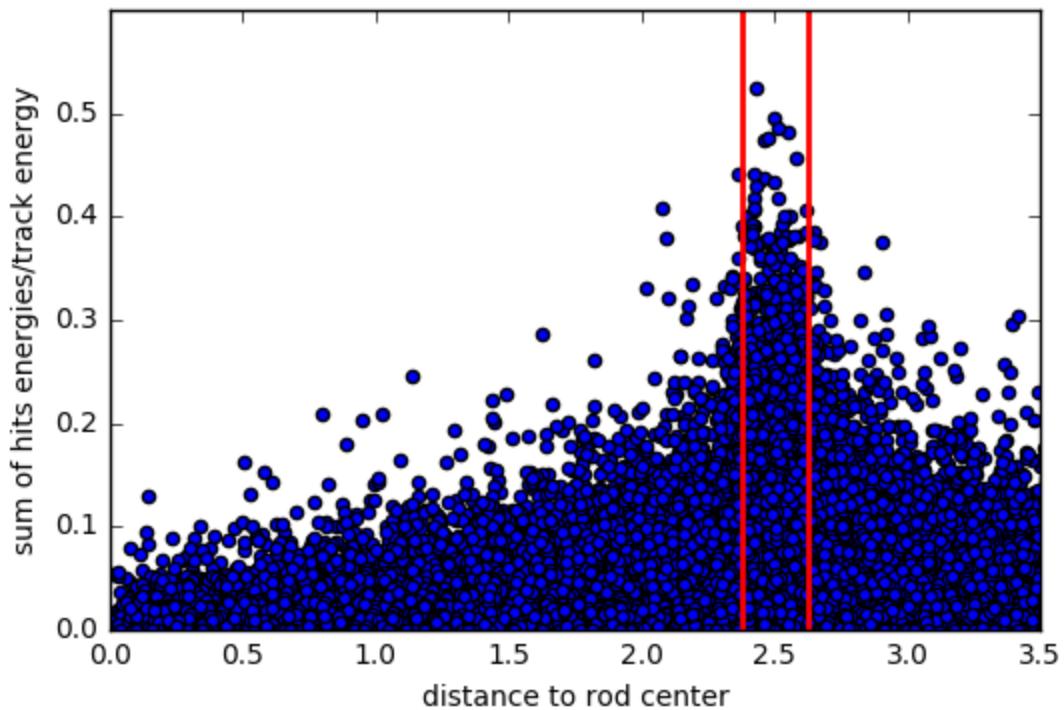


Fig. 1.5: Energy response for electrons in a calorimeter for all electrons in a library

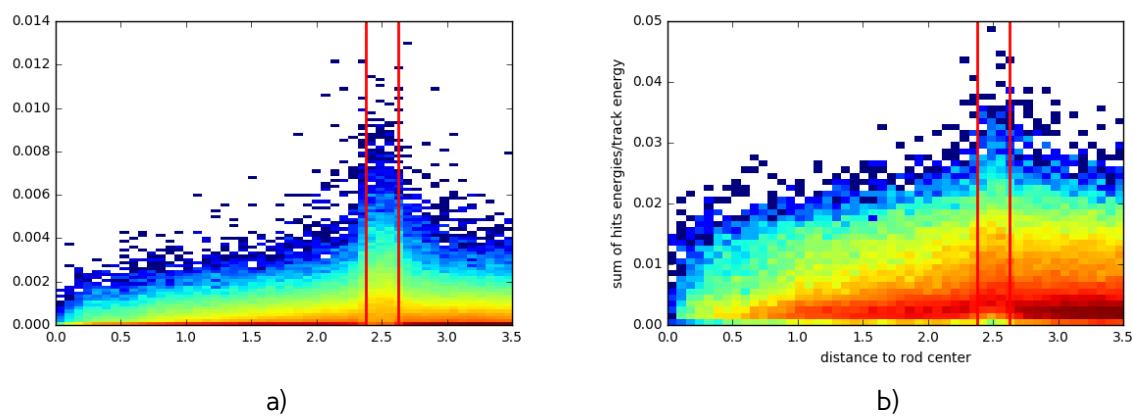


Fig. 1.6: Energy response for electrons in a calorimeter for  
 a) electrons with energy less than 100 MeV  
 b) electrons with energy bigger, than 100 MeV

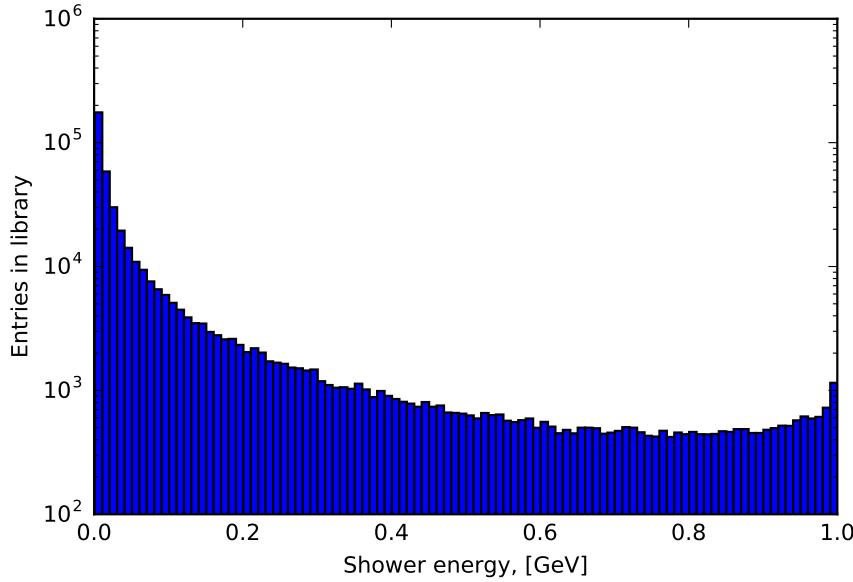


Fig. 1.7: Distribution of shower energy used in production of 1000 GeV electrons.

120 are having more energetic showers, than particles in a dead material. Introducing additional spacial  
 121 binning, that is defined as distance to the closest rod center is allowing to capture this structure.  
 122 Dependency of summed energy of shower on the distance defined above is shown on a Figure 1.5.  
 123 Gap is marked by the red lines. Size of this differences between sensitive material and dead material  
 124 is depending on a initial particle momentia (Figure 1.6 a and b). For electrons with energy greater  
 125 than 500 GeV they are almost negligible. This is why upper limit for frozen showers was set to 1000  
 126 MeV.

127 Performance of frozen showers is also depending on a lower limit of a method. Distribution of  
 128 shower energies, used for production of high-energetic electron (1000 GeV in that case), is shown on  
 129 a Figure . More than 50% of them are having energy less than 20 MeV. Studies have showed, that  
 130 Frozen showers are slower, than a standard Geant4 simulation for showers with energy 3 MeV. This  
 131 is happening due to library non-binned structure for energy. This makes search of closest energy  
 132 shower in a library slower, than simulation of shower with zero or one hit in a sensitive detector.

### 133 Generation of Frozen Showers library

134 In a Frozen Shower method there are separate libraries for each particle and subdetector used.  
 135 Showers should cover fully energy and pseudorapidity region and be able to describe data, that is  
 136 needed during simulation. This is why 2 stages simulation approach have been used. The first stage  
 137 takes initial particle parameters from a physical processes (ttbar or a single electron).

138 The first stage is to take initial particle parameters, that later will be used in a library from a  
 139 physical process. This is done using simulation of some process (e.g. ttbar or single electron). Every  
 140 time, when particle becomes eligible for Frozen Showers, it parameters are saved in a HepMC format.  
 141 Particles inside calorimeter tend to cluster tightly around initial track, so random truncation of initial  
 142 particles is used to obtain better detector coverage. On the second stage, this primary particles are  
 143 propagated through the calorimeter using standart Atlas simulation infrastructure. Resulted shower

The general frozen showers parameters	
Detectors used	FCAL1, FCAL2, FCAL3
Type of the particle	photons, electrons, neutrons
Energy range	$E_\gamma < 10 \text{ MeV}$ , $E_e < 1000 \text{ MeV}$ , $T_n < 100 \text{ MeV}$
Containment requirement	$\Delta E_{\text{shower}} > 98\%$
The library post-processing parameters	
Generation clustering cutoff	$(\Delta R_{\text{cluster}})^2 < 25 \text{ mm}$
Generation truncation cutoff	$R_{\text{hit}}^2 < 50000 \text{ mm}$ , $\Delta E_{\text{shower}} < 1\%$

Table 1.1: Main parameters used for the frozen shower libraries in FCAL

parameters are saved in a library. This procedure allows to take into account sampling fluctuations and charge-collection effects on a hit information automatically. Additionally, in order to save disc space as well as a memory consumption, hit information is compressed. This compression is done in a two steps, hit merging and truncation:

- if the distance between any two hits is smaller, than a given parameter  $R_{\min}$ , then hits are merged into one deposit at the energy weighted center of them. This process is done iteratively.
- hits whose energies are below the fraction  $f$  of the total energy sum of all hits, are truncated. The energy of remaining hits is rescaled back to preserve the total deposited energy.

Unfortunately, for a Frozen Showers, generated for Run-1 monte-carlo, additional tuning of electron libraries was needed. This was done using reconstructed energy of electrons. Frozen Showers tend to underestimate fluctuations of energy loss, that is leading to a smaller electron resolution for a high energies. Correction is done by enlarging bin, corresponding to a gap position. Also, correction of the mean shift is done by scaling energy response of all showers. After this frozen showers are showing good agreement with full simulation. This procedure needs to be done every time, when something is changing in software. Because tuning is done manually, lots of a manpower is needed for each Monte Carlo mass production campaign.

### Distance binning problem

As it was mentioned before, process of library generation can be complicated and take a lot of the time because of the needed tuning. In this subchapter possible ways to improve frozen showers performance have been studied.

As it was mentioned before, that there are two type of material used in a FCAL. Showers within them are giving different response, what is affecting overall reconstructed electron energy resolution. At the first generations distance bin have been corresponding to LAr gap or dead material positions. During tuning bin with LAr was enlarged to gain a better agreement with full simulation. So, one of the basic ideas to improve frozen showers performance is to change a size of LAr gap in a library generation.

It was decided to treat showers, that have been born near LAr gap and crossed it on a radiation length, in a same way with showers in sensitive material gap, and call them sensitive material showers. Oppositely, showers, that haven't crossed LAr gap, are called dead showers. This model leads to a bigger gap width by a definition. One of the possible ways to find this bin position automatically is to use machine learning tools.

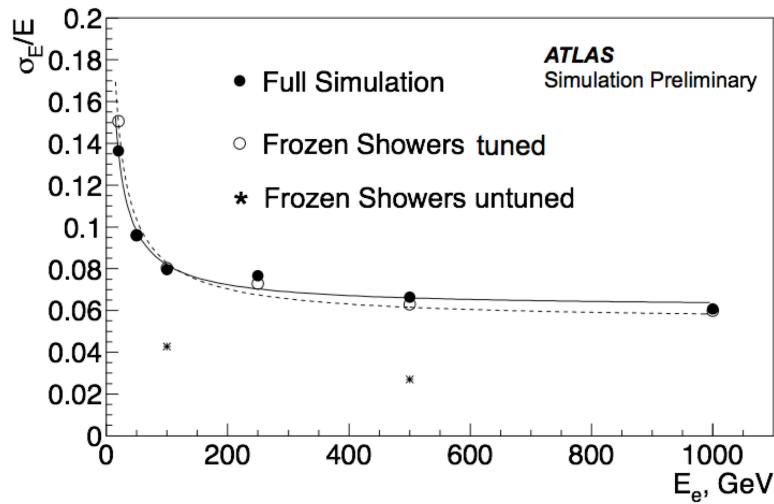


Fig. 1.8: Electron resolution for full simulation, tuned and untuned frozen showers

176 Machine Learning is a set of algorithms, what allows computers to learn and give a predictions  
 177 without being specifically programmed. This is a modern field of computer science, that is wildly  
 178 used in a different fields like computer vision, natural language processing, data science etc. There  
 179 are two main types of machine learning algorithms: supervised, where example of desired output  
 180 is given by the "teacher" and the goal is to learn a general rule, that maps inputs to outputs and  
 181 unsupervised learning, then there are no labels given to algorithm, and algorithms is discovering  
 182 hidden patterns in data. Initial data parameters of interest, that are used in algorithm to learn are  
 183 called features. It is important to have right proper set of features and good training sample.

184 From a geometrical point of view, one of the main parameter is a direction of the shower. Eta  
 185 momentum distribution is showed on a Figure 1.9 . Most of the showers are collinear to an electron  
 186 direction. Because of this it was decided to use as a training sample simulation results for electrons  
 187 with energies less than 1 GeV and momentum uniformly distributed between eta 3.0 and 4.0. This  
 188 allowing to study equally low and high energy showers equally.

189 From our definition of 2 classes of showers, it is simple to construct a pre-labelled training sample.  
 190 This is done by reducing initial sample and taking showers near rod center and inside liquid argon  
 191 gap. Output of this classifier, that was trained on with sample with shower features, such as energy  
 192 response and number of hits, than can be used to expand our labels to a full distance range. Then  
 193 it can be used as an input to a second classifier, which will separate two types of showers using  
 194 particle parameters, such as energy and distance to a rod center. For a first step decision trees have  
 195 showed good classification efficiency (around 97%). For a second classifier support vector machines  
 196 have been used. This method is trying to reconstruct a hyperplane, that is dividing two classes.  
 197 Outputs of both of this classifiers are shown on a figure . New gap position is determined using  
 198 borders of hyperplane. This procedure is giving expected from the initial model results. Gap is wider,  
 199 than and original one. It is also getting bigger with bigger energy, because of the radiation length  
 200 growth. Validation results for two different eta bins are shown on figure a) and b). In a bin this new  
 201 binning is performing better, than original one without any additional tuning. Unfortunately this is not  
 202 true for all of the bins, as we can see on a figure b). This eta bin have showed worst performance  
 203 for a new binning, but it is performing still better, than original binning without tuning.

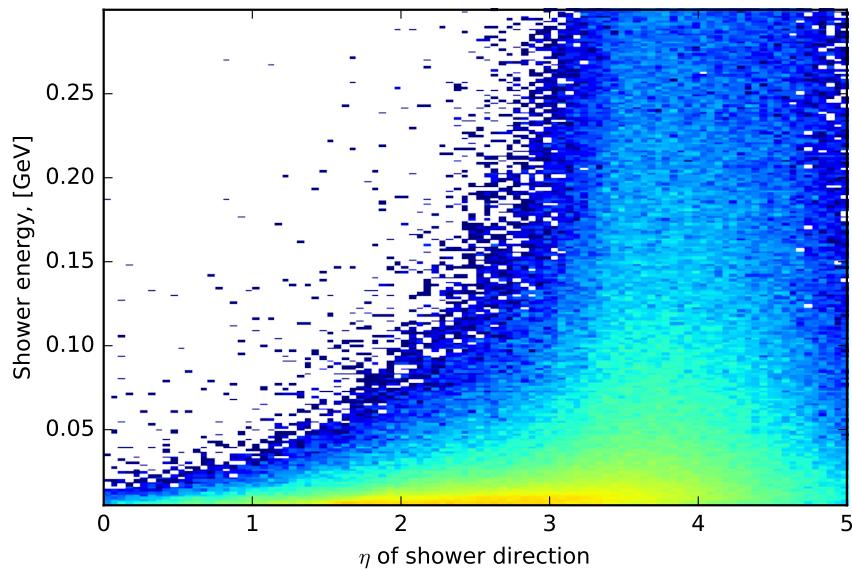


Fig. 1.9: Distribution of showers used in production of 1000 GeV electrons on shower energy vs  $\eta_{momentum}$  plane.

204     This binning was used in a production of new libraries for Monte Carlo in a Run-2. It is planned to  
205     use more precise training sample for a future iterations of this procedure for improving performance  
206     of outlying eta bins.

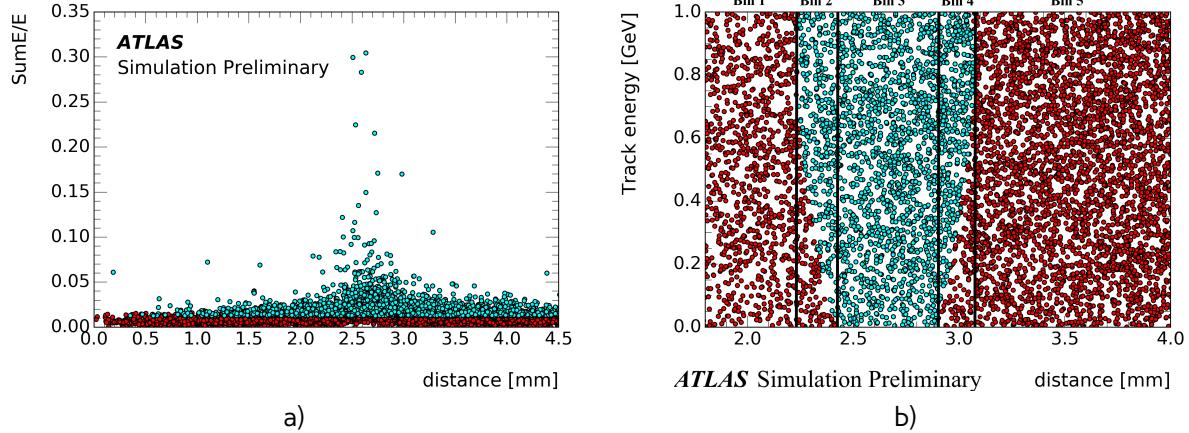


Fig. 1.10: Results of machine learning for a) first classifier b) second classifier. Cyan dots are corresponding to sensitive material showers, red - dead material showers

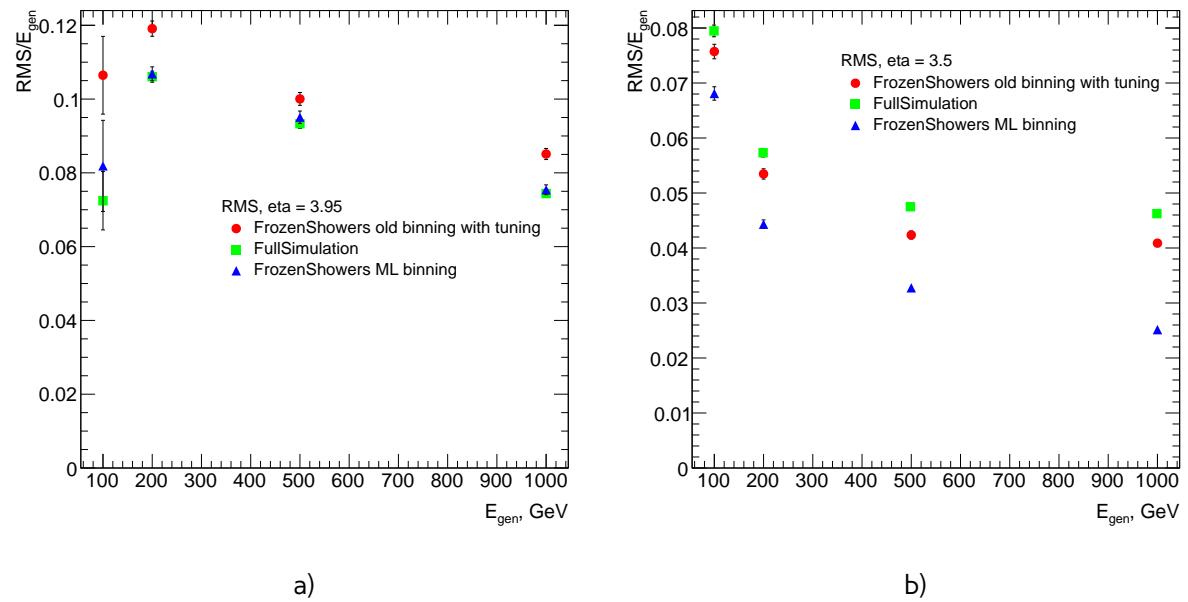


Fig. 1.11: Resolution of reconstructed electrons for full simulation, new libraries with ML binning and old tuned libraries with original binning for a) eta =3.95 b)eta=3.5



207

# Chapter 2

## 208 Selection

209 Selection criteria is the set of requirements, that is applied both on data and MC. Analysis is de-  
 210 pending on a selection, that can separate process of interest (signal) from other processes. For  
 211  $pp \rightarrow W \rightarrow ev/\mu\nu$  and  $pp \rightarrow Z/\gamma^* \rightarrow ee/\mu\mu$  selection criteria can be divided into 3 groups: data  
 212 quality, lepton and boson cuts. In this chapter all of them will be discussed and a cut flow presented  
 213 In this chapter selection criteria for  $pp \rightarrow W \rightarrow ev/\mu\nu$  and  $pp \rightarrow Z/\gamma^* \rightarrow ee/\mu\mu$  are presented.

### 214 2.1 Data quality cuts

Table 2.1: Analysis selection

Event selection	
Single lepton trigger	
Good Run List	
Reject events with LAr errors	
Number of tracks at primary vertex $\geq 3$	
Electron Selection	Muon Selection
$P_T > 20\text{GeV}$	$P_T > 20\text{GeV}$
$ \eta  < 2.47$	$ \eta  < 2.5$
excluding $1.37 <  \eta  < 1.52$	
Medium electron identification	Medium muon identification
$\text{PtCone20} < 0.1$	$\text{PtCone20} < 0.1$
W boson selection	Z boson selection
$\text{EtMiss} > 25 \text{ GeV}$	$\text{EtMiss} < 20 \text{ GeV}$
$M_T > 45 \text{ GeV}$	$66 < M_{ee} < 116 \text{ GeV}$

215 For a measurement we must use the data with a proper quality. Unfortunately not all of the events  
 216 satisfy this criteria. One of the possible source of the problems could be that LHC was not in a  
 217 stable beam mode, or parts of the detector have been switched off, or event had too many noisy  
 218 cells. The information about luminosity blocks, that need to be excluded is stored in a "Good Run  
 219 List". Events, where LAr calorimeter was malfunctioning are excluded by LAr quality criteria. Events  
 220 are furthermore required to have at least one primary vertex from a hard scattering with at least 2  
 221 associated tracks reconstructed.

## 222 **2.2 Lepton quality cuts**

223 Online selection of events is based on a single lepton trigger, depending on a flavor of analysis.  
224 For electron analysis it is required to have EF\_e15\_loose1 trigger, which records electrons with  $E_T >$   
225  $7\text{GeV}$ . This trigger is also using additional "loose" isolation requirements to exclude jets, that are  
226 misidentified as electrons. In muon channel lowest single lepton trigger available used (EF\_mu10). It  
227 records events with muons  $E_T > 10\text{GeV}$ . Moreover, matching between trigger and lepton is required.

228 All of the analysis are using similar selection criteria, applied on a leptons. All of the leptons must  
229 satisfy requirement  $P_T > 20\text{GeV}$  Electron candidates are required to be within pseudorapidity range  
230  $|\eta| < 2.47$ . Candidates within the transition region between the barrel and endcap electromagnetic  
231 calorimeters,  $1.37 < |\eta| < 1.52$ , are removed. Additionally, for better multijet background rejection  
232 medium identification and PtCone20 < 0.1 criterias are applied.

233 Muons are satisfying following criteria: they should be reconstructed by a staco algorithm in a  
234 muon spectrometer and ... within range  $|\eta| < 2.5$  . Set of medium requirements is applied. They  
235 must also satisfy PtCone20 < 0.1 isolation criteria

## 236 **2.3 Boson selection**

237 Events, contained W boson are required to have exactly one selected lepton. Events, where there are  
238 additional "good" leptons are rejected. Missing transverse energy is required to be  $E_{T\text{Miss}} > 25\text{GeV}$ .  
239 W boson, formed out of etMiss and lepton should have transverse mass  $M_T > 45\text{GeV}$ . After the  
240 full selection total number of events in electron channel is .. (.. and .. for  $e^+$  and  $e^-$  respectively).

241 The reconstructed lepton pair in case of Z boson analysis is required to invariant mass between  
242 66 and 116 GeV. Both upper and bottom limits allows to exclude regions with high background  
243 contamination and low statistics.

244 Full set of cuts is summarized in a table 2.1.

## 245 **2.4 Cut flow**

246 The results of this set of cuts, applied both on data and Monte Carlo are summarized in a table ??.

247

# Chapter 3

248

## Monte Carlo corrections

249 Monte Carlo plays important role in cross-section measurement. It is constantly undergoing correction  
 250 to data, in order to obtain a required precision. Part of this corrections have been described in  
 251 a chapter 1. Unfortunately, not everything can be taken into account during simulation itself. This  
 252 leads to a differences between data and monte carlo, that needs to be accounted for. There are two  
 253 possible methods to correct monte carlo without regenerating it. First on is to apply event weight, so  
 254 what each mc event can contribute to non 1 entries in a histogram. This is called reweighting. Second  
 255 one is to smear MC. It is using random number to alter reconstructed 4-vectors. This chapter de-  
 256 scribes all additional corrections, what have been applied on MC in this analysis. All of this correction  
 257 are introducing additional systematic error, that will be discussed in the chapter ??

258 **3.1 Number of vertexes reweighting**

259 The effect of multiple particle interactions per bunch crossing is accounted in MC. But there is  
 260 sill remaining difference between simulation and collected data. This mismodelling can affect many  
 261 distributions. One of the ways to correct this is to apply weight on number of vertexes. This variables  
 262 have a connection between each other, because each pp interaction produces separate vertex. Fig  
 263 ?? shows a comparison of number of vertexes before and after reweighting.

264 **3.2 Lepton efficiency corrections**

265 Lepton detection efficiency at ATLAS detector can be divided into three components:

- 266     • The reconstruction efficiency  $\epsilon_{rec}$  is associated to the reconstruction algorithm. It is a proba-  
 267       bility to reconstruct lepton as a lepton of this flavor .
- 268     • The identification efficiency  $\epsilon_{id|rec}$  is the probability that a reconstructed lepton survives iden-  
 269       tification requirements.
- 270     • The trigger efficiency  $\epsilon_{trig|midrec,id}$  is the probability, that lepton satisfies trigger requirement.

The full efficiency for a single lepton can be written as:

$$\epsilon_{total} = \epsilon_{rec} \times \epsilon_{id|rec} \times \epsilon_{trig|rec,id} \quad (3.1)$$

All of this efficiencies are measured using Tag and Probe method in  $Z \rightarrow ll$  decays. This is allowing to insure, that all of the reconstructed lepton candidates are coming from an actual leptons. One of the leptons from Z boson, called "probe", is initially selected with all of the cuts, minus one under

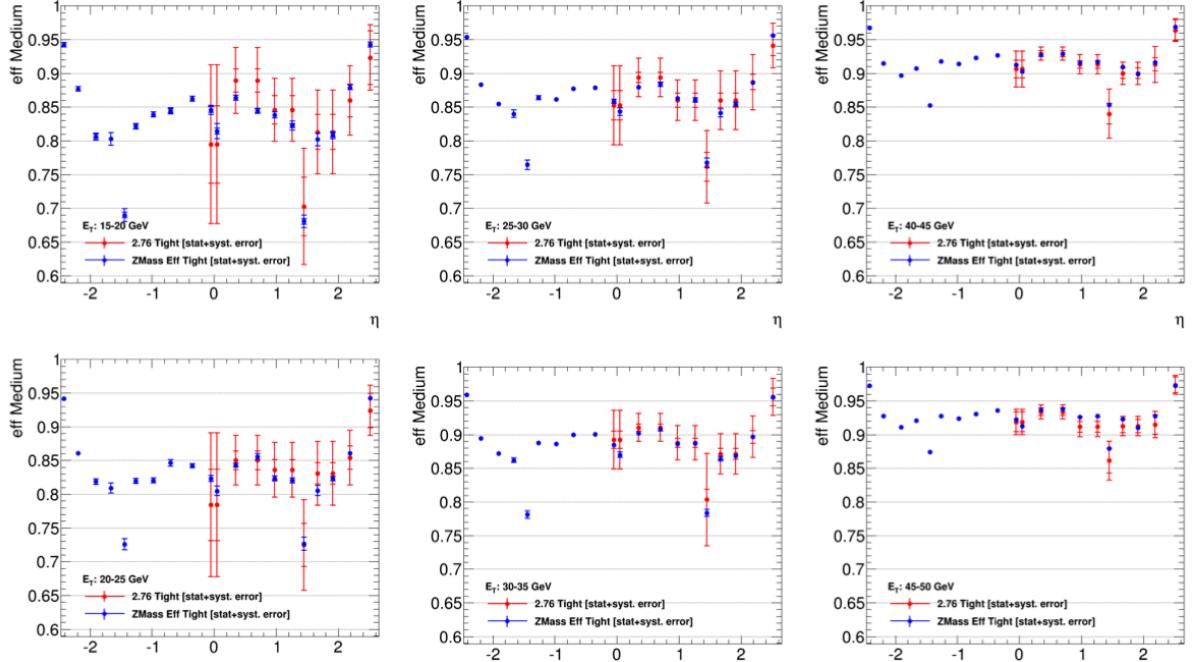


Fig. 3.1: Comparison of electron efficiencies as calculated for 8TeV (blue points) and 2.76TeV (red points) for MC simulation. Efficiencies are shown as a function of pseudorapidity ( $\eta$ ) for different electron  $E_T$  bins. Both statistical and systematic uncertainties are shown.

study. Second one, called "probe" satisfies more tighter selection with additional cut, such as, for example, trigger matching. Simulation samples are corrected to data by a scale-factor :

$$SF_{reco,id,trig} = \frac{\epsilon_{reco,id,trig}^{data}}{\epsilon_{reco,id,trig}^{MC}} \quad (3.2)$$

271 Each of the scale factors calculated in a  $p_t$  and  $\eta$  bins and has an associated statistical and system-  
 272 atical uncertainty component. Statistical component is connected to a size of  $Z \rightarrow ll$ , which is in  
 273 our case is around 500 event per each lepton flavor. This means that precise calculation of scaling  
 274 factors based on this data is difficult. It is possible to use scale factors for 8 TeV 2012 data. The  
 275 main difference between this data samples are center of mass energy and a pile-up conditions (10  
 276 in 2012 and less than 1 in 2013). This effects have been studied on a  $Z \rightarrow ee$  sample. Fig. 3.1 shows  
 277 that all of the differences in a scale factors are negligible and fully covered by the statistical error.  
 278 This justifies the usage of 8 TeV scaling factors with increased uncertainty in this analysis.

279 Unfortunately, single muon trigger haven't been presented in a 2012 data, so muon trigger scale  
 280 factor is calculated as a single number in order to minimize systematics contribution.

### 281 3.3 Electron energy scale and resolution

282 Electrons clusters tend to shift in a reconstructed energy compared to a truth energy of initial  
 283 electron. Correction of this shift is done on a both data and MC as a 3 step process:

- 284 • Electronic calibration, that transfers a raw signal from a readout to a cluster energy deposit.

- 285 • MC based calibration. It corrects effects of energy loss in the material in front of calorimeter  
 286 and leakage into the hadronic calorimeter. This calibration is applied on both data and MC.
- 287 • Correction of calorimeter cell response in data. This is allowing to get right response in non-  
 288 optimal HV-regions and exclude biases in a calorimeter electronics reconstruction.

Energy shift is parameterised, as:

$$E^{data} = E^{MC}(1 + \alpha_i), \quad (3.3)$$

where  $E^{data}$  and  $E^{MC}$  are the energies in data and simulation, respectively and  $\alpha_i$  is a mean shift in a given bin  $i$  in  $\eta$ . Electron resolution is described by a formula 1.1. It is assumed, that sampling and noise terms are modeled well by MC and the main difference is coming from a constant term. The electron resolution correction then can be written as:

$$\frac{\sigma_E}{E_i}^{Data} = \frac{\sigma_E}{E_i}^{MC} \oplus c_i \quad (3.4)$$

289 where  $c_i$  is  $\eta$  dependent relative resolution correction. Parameters  $\alpha_i$  and  $c_i$  are obtained from  $Z \rightarrow ee$   
 290 invariant mass distribution with pseudorapidity configuration  $(\eta_i, \eta_j)$  using  $\chi^2$  fit. Resulting energy  
 291 scale is applied on a data, while resolution is corrected for MC.

## 292 3.4 Muon momentum correction

$$\frac{\sigma_{ID}(p_T)}{p_T} = a_{ID}(\eta) \oplus b_{ID}(\eta) \cdot p_T \quad (3.5)$$

$$\frac{\sigma_{MS}(p_T)}{p_T} = a_{MS}(\eta, \phi) \oplus b_{MS}(\eta, \phi) \cdot p_T \oplus \frac{c(\eta, \phi)}{p_T} \quad (3.6)$$

## 293 3.5 Hadron recoil correction