

1

Part I

2

Theory introduction

³ Chapter **1**

⁴ **Standart model**

⁵ Chapter **2**

⁶ **Theory of pp collisions**

⁷ Chapter **3**

⁸ **PDF fits and DLGAP formalism**

⁹ Chapter **4**

¹⁰ **Physics of WZ bosons**

¹¹ **4.1 Theory**

¹² **4.2 Cross-section measurement**

13

Part II

14

Experimental setup

15

Chapter **5**

16

LHC

¹⁷

Chapter

6

¹⁸

ATLAS experiment

19

Chapter

7

20

Software

22 Monte Carlo

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

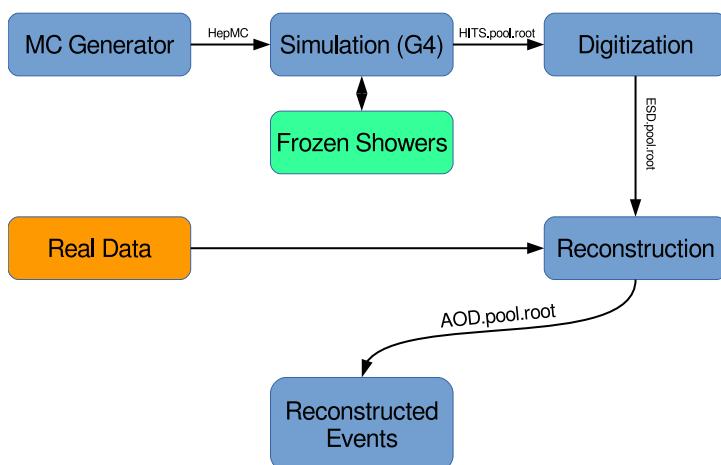
28 **8.1 Monte Carlo simulation at ATLAS experiment**

Fig. 8.1: Diagram of the ATLAS MC production chain

29 MC is allowing to make different analysis, such as compare data with predictions, study detector
 30 or selection algorithms performance. All of this applications are requiring MC precision. Simulation
 31 software expects to use precise physics models for sampling and have large enough statistics, to
 32 exclude statistical uncertainties (usually 5 times more, than expected in a data). ATLAS simulation
 33 software is integrated into Athena and usually used during large production of events. Simulation
 34 chain is generally divided into 4 main steps (Figure 8.1):

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

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

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

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

43 This scheme allows to use computing resources more efficiently, than with a single-step simulation,
 44 and simplifies software validation, since it is possible to reuse files from previous stages. In the
 45 following sections event generation and simulation will be described in more details

46 **8.1.1 Event generators**

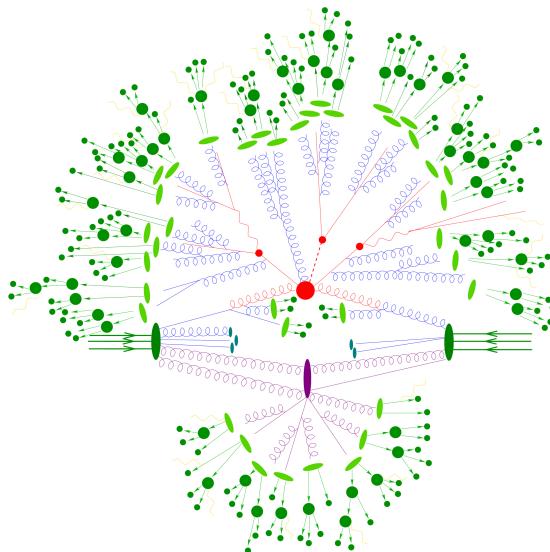


Fig. 8.2: Schematic view of a $t\bar{t}H$ event produced in a pp-collision: the hard scattering is shown as a red blob with the solid and dashed lines as the resulting three particles. Independently happening multi-particle interactions are indicated by the violet blob. Parton showers are shown with curly lines. Hadronization yields hadrons as shown in light green, while the final state particle are dark green.

47 The outcome of the hard interaction could be simple scattering of the hadron elementary con-
 48 stituents, their annihilation into new resonances or a combination of two. In any case, the final state
 49 has a large particles multiplicity. The main goal of event generator is to provide a complete picture of
 50 this final states: description of the particle types and momentia on event-by-event basis. The fac-
 51 torisation theorem [?] allows to make event generation in independent stages, which are dominated
 52 by different dynamics:

53 **Modelling of hard subprocess** Hard subprocess is happening at the smallest scales of times and
 54 distance, where all of the colliding partons are considered free. Process of interest is simulated
 55 by selecting production channels and calculating corresponding matrix elements (ME) in the
 56 desired level of accuracy in perturbation theory . Most of the generators have leading order
 57 or next to leading order ME in α_s .

58 **Parton showering** Quarks and gluons from hard process can radiate secondary quarks and gluons,
59 resulting on the dozens of additional partons associated with the event. This process calculated
60 as step-by-step evolution of momentum transfer scales from highest (hard subprocess), to
61 the lowest (around 1 GeV). There is a possibility of double counting between showers and hard
62 subprocess. This can be avoided using matching approach, for which higher order corrections
63 to ME are integrated with parton showers, or merging strategy, where jet resolution scale is
64 used as a threshold between matrix elements and parton showers.

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

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

71 Schematic plan of simulation of ttbar event is shown in Figure ???. The hard scattering itself is shown
72 as a red blob with the solid and dashed lines indicating the resulting particles, which themselves decay
73 further. Underlying event is indicated by the violet blob. Parton showers are shown with curly lines.
74 Hadronization yields hadrons as shown in light green, while the final state particles are dark green.
75 The current analysis uses samples generated with the following generators:

76 Powheg [?] Powheg is generator with NLO ME [?], that can be interfaced with other generator(such
77 as Pythia or Herwig) for higher precision of showering.

78 Pythia [?] Pythia is a general purpose generator for hadronic, hadron-lepton and leptonic collisions.
79 It can model initial and final state showers, hadronisation and decays, underlying event (via
80 multi parton interactions). Pythia contains library with around 240 processes with LO ME. It
81 uses Lund String model [?] for hadronisation.

82 Herwig [?] Herwig is a LO general purpose event generator for simulation lepton-lepton, hadron-
83 lepton and hadron-hadron collisions. The main difference between Pythia and Herwig is that
84 it uses angular ordering in the parton showers and also models the hadronisation step based
85 on the cluster fragmentation

86 Sherpa [?] Sherpa is a generator with tree level of matrix elements, featuring its own implementation
87 of parton shower and hadronisation models.

88 Photos [?] Precision tool for QED radiative corrections in W and Z decays.

89 Tauola [?] Generator, used to describe leptonic and semi-leptonic τ -decays.

90 8.1.2 Simulation in Geant4

91 After event generation, simulation software obtains hardware response for final state particles. The
92 main method used by ATLAS, referred to as *Full Simulation*, makes use of the Geant4 [?]. It is C++
93 based toolkit for the simulation of the passage of particles through matter. It is used in a wide range
94 of experiments in high energy and nuclear physics.

95 Geant4 can simulate complex detector structures with sensitive detector material and correspond-
96 ing infrastructure. It can also calculate basic properties of materials, like radiation and interaction

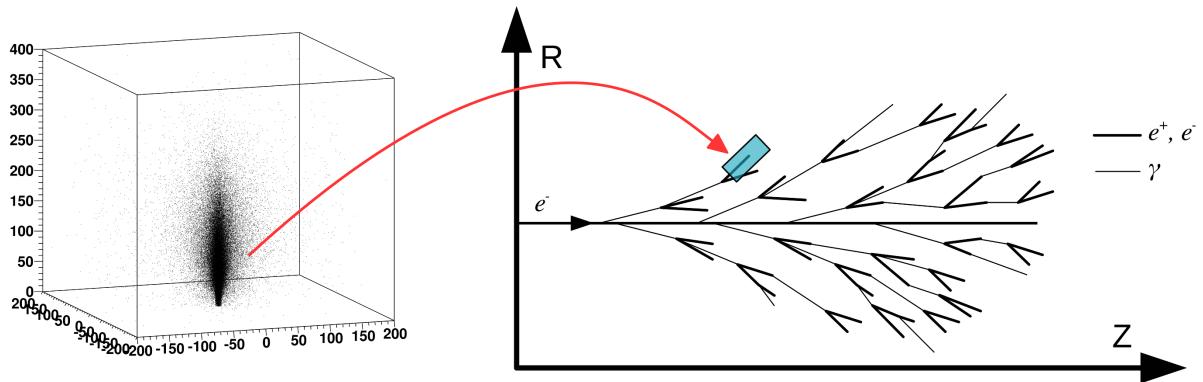


Fig. 8.3: Diagram showing the shower substitution of the low-energy particle, during the high-energy particle simulation.

length. For detector Geant4 stores "hits" information - snapshots of physical interactions. In Geant4 events and particles are simulated separately and each particle is moved step by step. Size of this step is chosen to preserve both CPU performance and required precision. Physics is treated as a set of discrete processes. They could be handled either at rest, along step or after it. Geant4 package has different models and approximations for hadronic and electromagnetic processes. Some of them are not approximate, but computationally fast. It allows to choose set of the models, called physics list, depending on particular requirements. There are several reference physics lists, that are validated for each new release of Geant4 software. ATLAS experiment is using one of this lists.

It is necessary to have mass MC production for each data taking, what is taking most of the resources. Uncertainties of some of Run-I analyses are dominated by available MC statistics.

It is possible to improve in CPU usage by tuning physics list or working on a magnetic field parametrisation. Also there are long term developments for multi-threading and vectorisation of the code. Yet, Run-2 has a higher pileup and luminosity, so even more MC events are needed. This means that fast and accurate simulation approach is essential. During simulation largest time is spent on calorimeters. This is the motivation for development of fast calorimetry techniques.

There are two main methods used in ATLAS:

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

8.1.3 Frozen Showers

Main principle of this method is described by its name. It is using pre-simulated "frozen" showers generated in full simulation and stored in a library. Particles below minimum energy thresholds are killed and replaced with these showers. All of the other particles are simulated using full simulation. This process is schematically shown in a Figure 8.3.

The library itself organized as follows: the header contains basic simulation parameters, like Geant4, geometry and ATLAS software release version and physics list used. Showers are stored in a bins of positional variables (see sec. 8.1.3), while energy remain unbinned. Each shower stores lateral and transverse size and information about energy, time and positions of the hits.

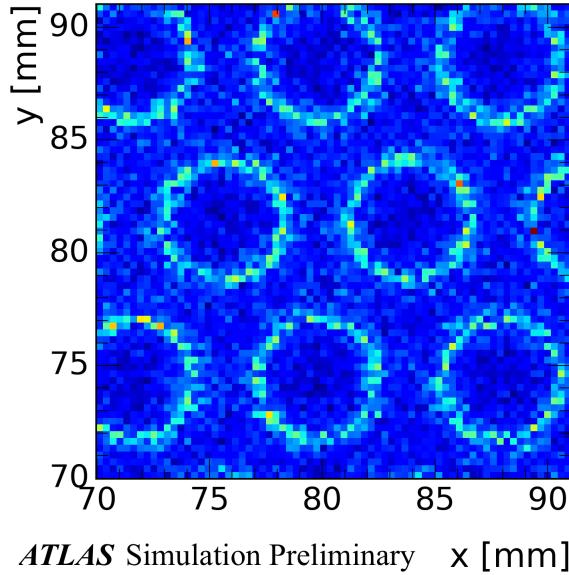


Fig. 8.4: Average energy response in a calorimeter vs x,y for electrons with energy less than 1 GeV

125 Production use of Frozen Showers

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

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 8.5). Resolution of a calorimeter can be written as:

$$\frac{\sigma}{E} \approx \frac{1}{\sqrt{E}} \oplus \frac{1}{E} \oplus const, \quad (8.1)$$

where symbol \oplus indicates a quadratic sum. The first term is 'stochastic term', which includes intrinsic shower fluctuations, second takes into account readout noise effects and pile-up fluctuations. Constant term derives from non-uniformities in a detector, what are causing large fluctuation of the energy loss. Resolution of high-energy electrons is mostly dominated by this term. Example of these fluctuations is shown on a Figure 8.4. Circles are corresponding to a LAr gaps inside FCAL. It can be seen, that particles inside sensitive material are having more energetic showers, than particles in a dead material. It is possible to capture this structure by introducing distance to a closest rod center:

$$D = \min(\sqrt{dx^2 + dy^2}, \sqrt{(step_x - dx)^2 + dy^2}, \sqrt{(step_x/2 - dx)^2 - (step_y - dy)^2}), \quad (8.2)$$

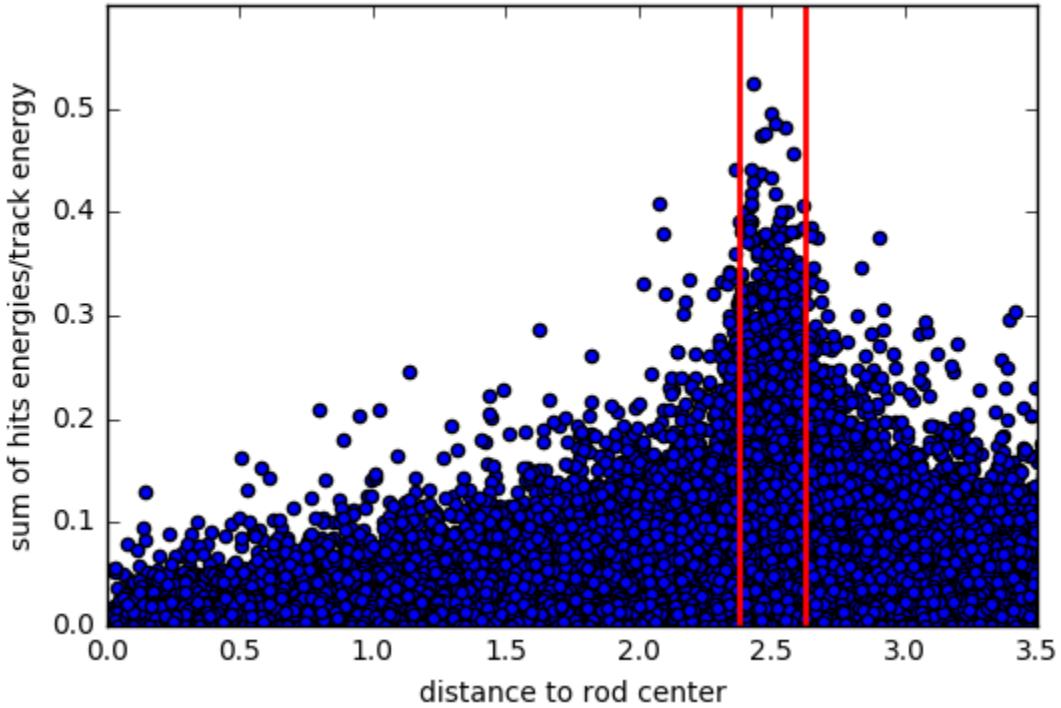


Fig. 8.5: Energy response for electrons in a calorimeter for all electrons in a library

where dx and dy are the distances to a rod center in a x and y plane respectively. They are calculated as:

$$dy = y - n[] \quad (8.3)$$

133 Dependency of summed energy of shower on the distance defined above is shown on a Figure 8.5.
 134 Gap is marked by the red lines. Size of this differences between sensitive material and dead material
 135 depends on a initial particle momentia (Figure 8.6 a and b). For electrons with energy greater than
 136 500 GeV they are almost negligible. Additionally, at higher energies gain in a CPU time is moderate,
 137 while library becomes bigger. This is reason for an upper limit to be set at 1000 MeV

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

144 **Generation of Frozen Showers library**

145 In a Frozen Shower method there are separate libraries for each particle and subdetector used.
 146 Showers should cover fully energy and pseudorapidity region and be able to describe data, that is

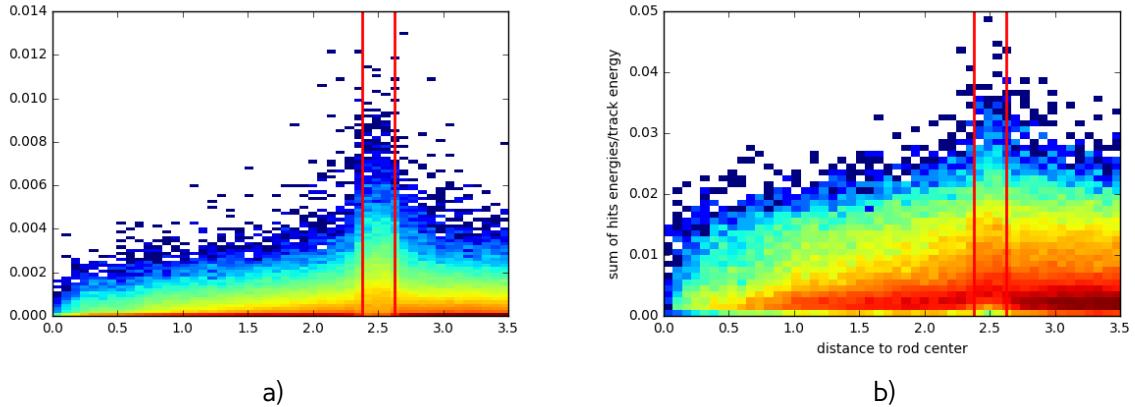


Fig. 8.6: Energy response for electrons in a calorimeter for a) electrons with energy less than 100 MeV b) electrons with energy bigger, than 100 MeV

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 8.1: Main parameters used for the frozen shower libraries in FCAL

needed during simulation. This is why 2 stages simulation approach have been used. The first stage takes initial particle parameters from a physical processes (ttbar or a single electron).

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

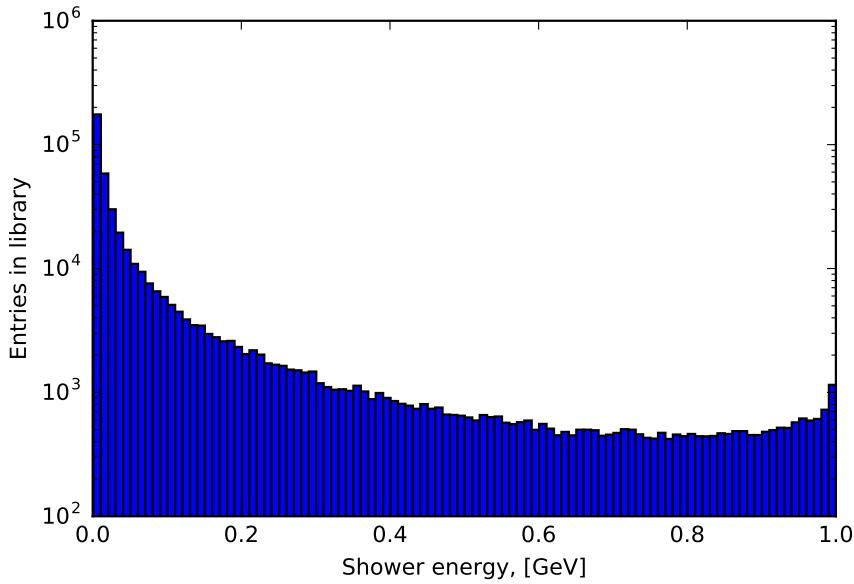


Fig. 8.7: Distribution of shower energy used in production of 1000 GeV electrons.

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

172 Distance binning problem

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

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

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

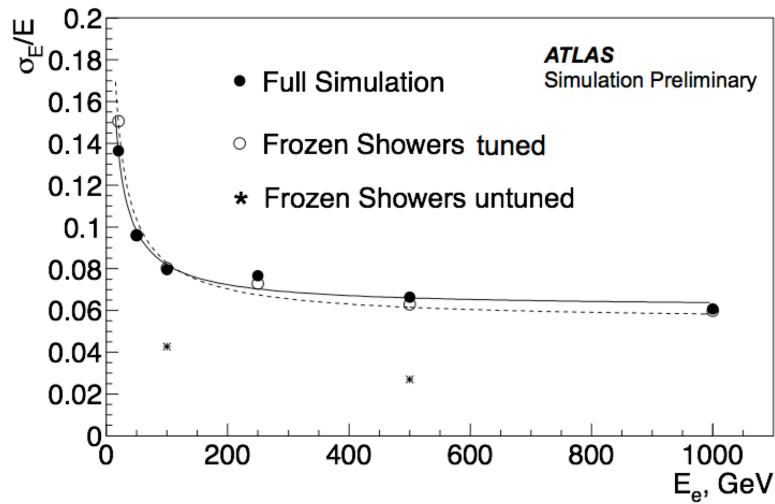


Fig. 8.8: Electron resolution for full simulation, tuned and untuned frozen showers

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

195 From a geometrical point of view, one of the main parameter is a direction of the shower. Eta
 196 momentum distribution is showed on a Figure 8.9 . Most of the showers are collinear to an electron
 197 direction. Because of this it was decided to use as a training sample simulation results for electrons
 198 with energies less than 1 GeV and momentum uniformly distributed between eta 3.0 and 4.0. This
 199 allowing to study equally low and high energy showers equally.

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

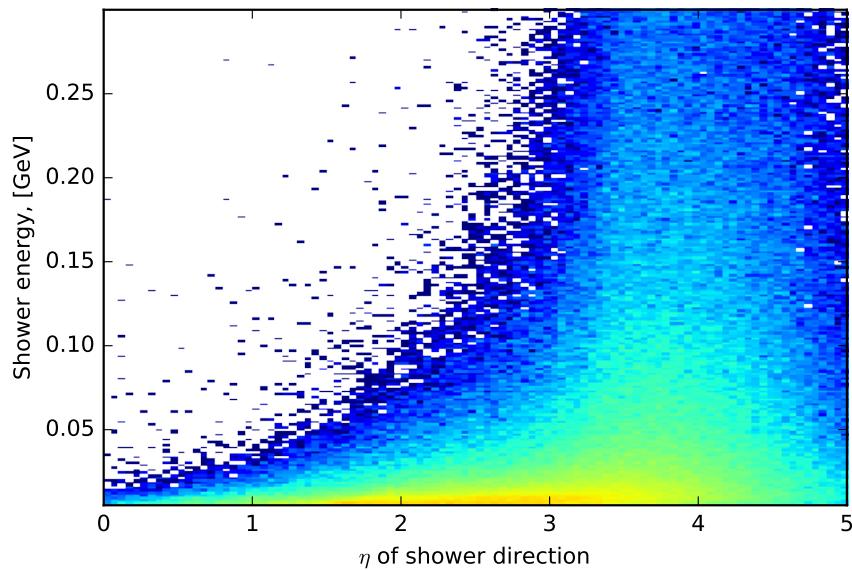


Fig. 8.9: Distribution of showers used in production of 1000 GeV electrons on shower energy vs $\eta_{momentum}$ plane.

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

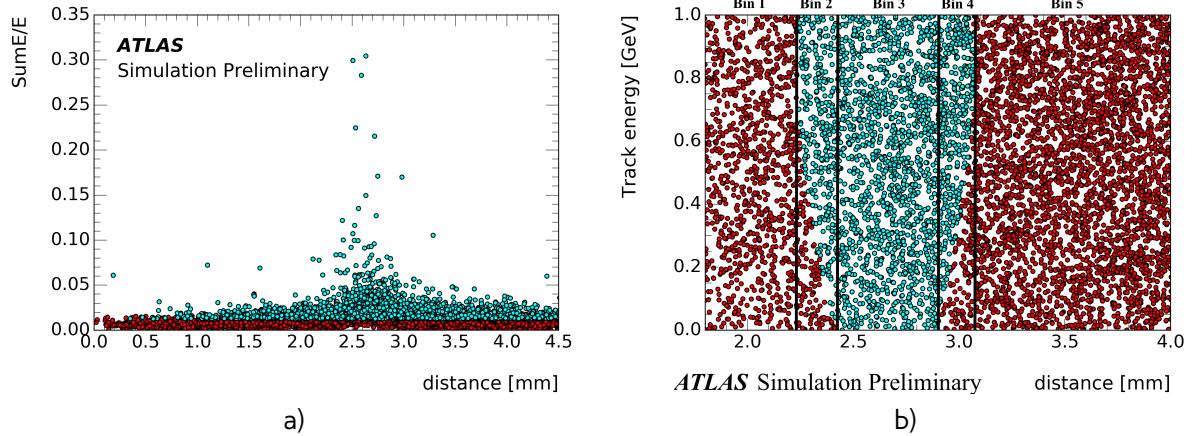


Fig. 8.10: Results of machine learning for a) first classifier b) second classifier. Cyan dots are corresponding to sensitive material showers, red - dead material showers

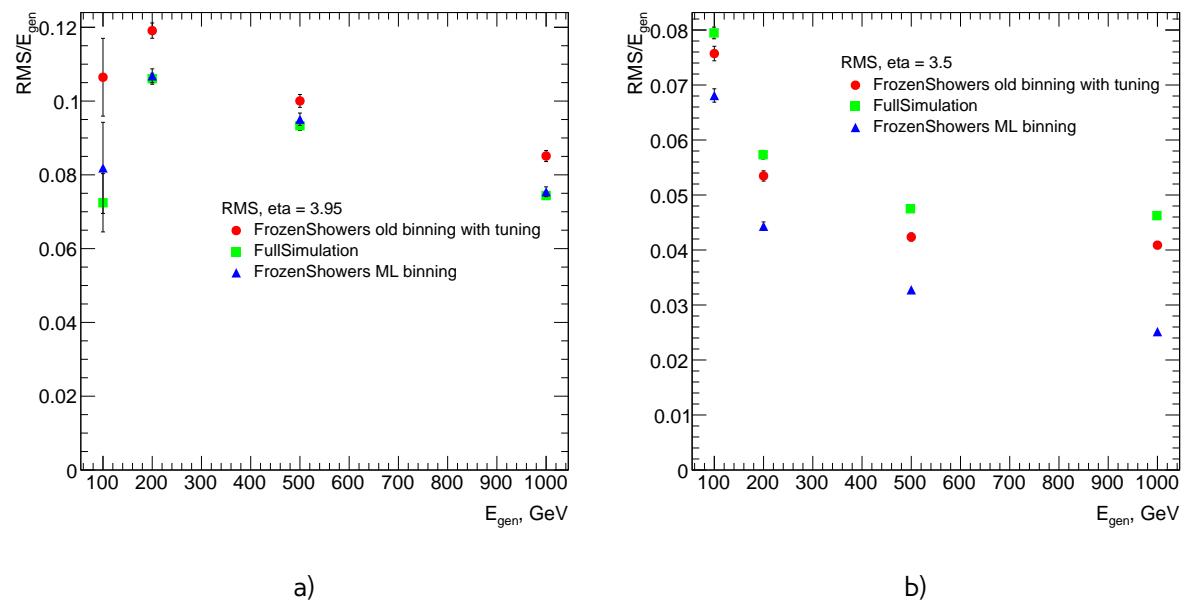


Fig. 8.11: Resolution of reconstructed electrons for full simulation, new libraries with ML binning and old tuned libraries with original binning for a) eta = 3.95 b) eta = 3.5

218

Chapter 9

219

DataSample

220

Part III

221

Measurement of cross-section

222 Chapter 10

223 Selection

224 Selection criteria is the set of requirements, that is applied both on data and MC. Analysis is de-
225 pending on a selection, that can separate process of interest (signal) from other processes. For
226 $pp \rightarrow W \rightarrow ev/\mu\nu$ and $pp \rightarrow Z/\gamma^* \rightarrow ee/\mu\mu$ selection criteria can be divided into 3 groups: data
227 quality, lepton and boson cuts. In this chapter all of them will be discussed and a cut flow presented
228 In this chapter selection criteria for $pp \rightarrow W \rightarrow ev/\mu\nu$ and $pp \rightarrow Z/\gamma^* \rightarrow ee/\mu\mu$ are presented.

229 **10.1 Data quality cuts**

Table 10.1: Analysis selection

Event selection	
Single lepton trigger	
Good Run List	
Reject events with LAr errors	
Number of tracks at primary vertex ≥ 3	
Electron 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	
$\text{EtMiss} > 25 \text{ GeV}$	$\text{EtMiss} < 20 \text{ GeV}$
$M_T > 45 \text{ GeV}$	$66 < M_{ee} < 116 \text{ GeV}$
Z boson selection	

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

237 **10.2 Lepton quality cuts**

238 Online selection of events is based on a single lepton trigger, depending on a flavor of analysis.
239 For electron analysis it is required to have EF_e15_loose1 trigger, which records electrons with $E_T >$
240 7GeV . This trigger is also using additional "loose" isolation requirements to exclude jets, that are
241 misidentified as electrons. In muon channel lowest single lepton trigger available used (EF_mu10). It
242 records events with muons $E_T > 10\text{GeV}$. Moreover, matching between trigger and lepton is required.

243 All of the analysis are using similar selection criteria, applied on a leptons. All of the leptons must
244 satisfy requirement $P_T > 20\text{GeV}$ Electron candidates are required to be within pseudorapidity range
245 $|\eta| < 2.47$. Candidates within the transition region between the barrel and endcap electromagnetic
246 calorimeters, $1.37 < |\eta| < 1.52$, are removed. Additionally, for better multijet background rejection
247 medium identification and PtCone20 < 0.1 criterias are applied.

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

251 **10.3 Boson selection**

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

256 The reconstructed lepton pair in case of Z boson analysis is required to invariant mass between
257 66 and 116 GeV. Both upper and bottom limits allows to exclude regions with high background
258 contamination and low statistics.

259 Full set of cuts is summarized in a table 10.1.

260 **10.4 Cut flow**

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

262

Chapter 11

263

Monte Carlo corrections

264 Monte Carlo plays important role in cross-section measurement. It is constantly undergoing correction
 265 to data, in order to obtain a required precision. Part of this corrections have been described in
 266 a chapter 8. Unfortunately, not everything can be taken into account during simulation itself. This
 267 leads to a differences between data and monte carlo, that needs to be accounted for. There are two
 268 possible methods to correct monte carlo without regenerating it. First on is to apply event weight, so
 269 what each mc event can contribute to non 1 entries in a histogram. This is called reweighting. Second
 270 one is to smear MC. It is using random number to alter reconstructed 4-vectors. This chapter de-
 271 scribes all additional corrections, what have been applied on MC in this analysis. All of this correction
 272 are introducing additional systematic error, that will be discussed in the chapter ??

273 11.1 Lepton efficiency corrections

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

- 275 • The reconstruction efficiency ϵ_{rec} is a probability to reconstruct lepton as a lepton of this
 flavor.
- 277 • The identification efficiency $\epsilon_{id|rec}$ is the probability that a reconstructed lepton survives iden-
 tification requirements.
- 279 • The trigger efficiency $\epsilon_{trig|rec,id}$ is the probability, that lepton satisfy trigger requirements.

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

$$\epsilon_{total} = \epsilon_{rec} \times \epsilon_{id|rec} \times \epsilon_{trig|rec,id} \quad (11.1)$$

280 All of this efficiencies are measured using Tag and Probe method in $Z \rightarrow ll$ decays. This is allowing
 281 to insure, that all of the reconstructed lepton candidates are coming from an actual leptons. One of
 282 the leptons from Z boson, called "probe", is initially selected with all of the cuts, minus one under
 283 study. Second one, called "probe" satisfies more tighter selection with additional cut, such as, for
 284 example, trigger matching.

Reconstruction efficiency is assosiated with algorithm used to perform reconstruction. This is causing difference between electrons and muons efficiencies. In electron case it is a probability to reconstruct an elec tron with an electromagnetic calorimeter as an electron. Muon reconstruction efficiency is given by:

$$\epsilon_{reco,muon} = \epsilon_{reco,muon|ID} \cdot \epsilon_{ID} \approx \epsilon_{reco,muon|ID} \cdot \epsilon_{ID|MS}, \quad (11.2)$$

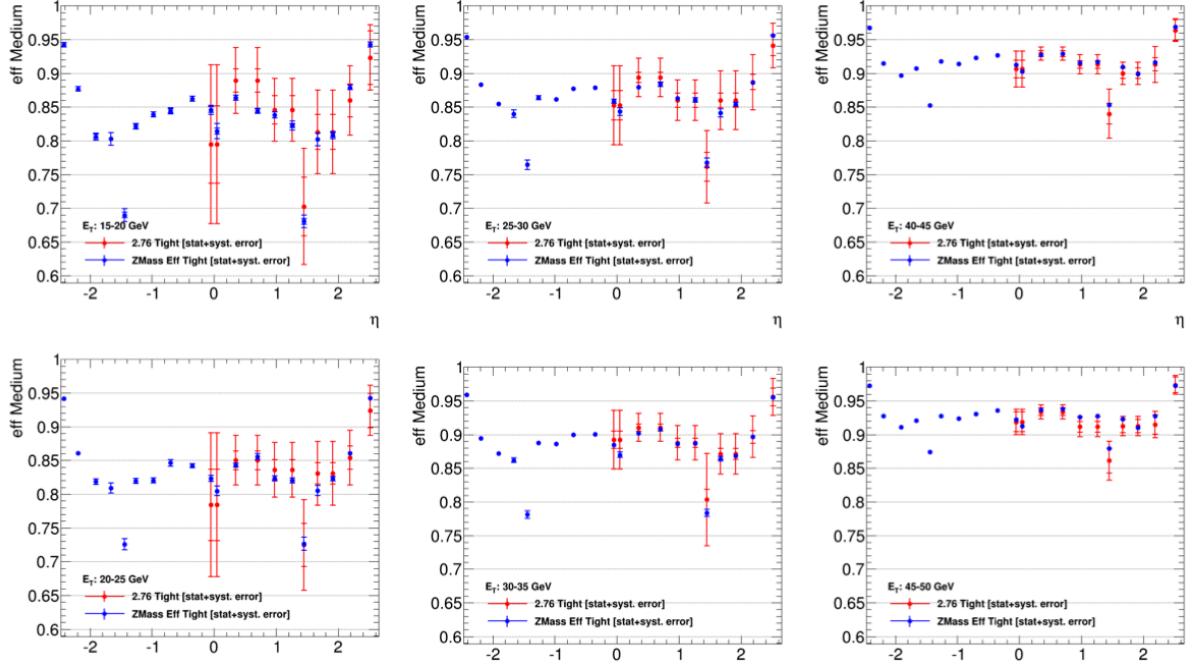


Fig. 11.1: Comparison of electron efficiencies as calculated for 8TeV (blue points) and 2.76TeV (red points) for MC simulation. Efficiencies are shown as a function of pseudorapidity (η) for different electron E_T bins. Both statistical and systematic uncertainties are shown.

where $\epsilon_{reco,muon|ID}$ is a conditional probability that muon reconstructed in ID is also reconstructed using MS as a combined muon, and ϵ_{ID} is a probability that muon is reconstructed as an ID track. This quantity cannot be measured directly and therefore is replaced by $\epsilon_{ID|MS}$, that can be measured by tag-and-probe method. uncertainty in this analysis.

Simulation samples are corrected to match data efficiencies by a scale-factor :

$$SF_{reco,id,trig} = \frac{\epsilon_{reco,id,trig}^{data}}{\epsilon_{reco,id,trig}^{MC}} \quad (11.3)$$

Each of the scale factors calculated in a p_t and η bins and has an associated statistical and systematical uncertainty component. Statistical component is connected to a size of $Z \rightarrow ll$, which is in our case is around 500 event per each lepton flavor. This means that precise calculation of scaling factors based on this data is difficult.

It is possible to use scale factors for 8 TeV 2012 data. The main difference between this data samples are center of mass energy and a pile-up conditions (10 in 2012 and less than 1 in 2013). This effects have been studied on a $Z \rightarrow ee$ sample. Fig. 11.1 shows that all of the differences in a scale factors are negligible and fully covered by the statistical error. This justifies the usage of 8 TeV scalling factors with increased

Unfortunately, single muon trigger haven't been presented in a 2012 data, so muon trigger scale factor needed to be derived from a 2.76 TeV data. Different configurations is calculated as a single number in order to minimize systematics contribution.

301 11.2 Electron energy scale and resolution

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

- 304 • Electronic calibration, that transfers a raw signal from a readout to a cluster energy deposit.
- 305 • MC based calibration. It corrects effects of energy loss in the material in front of calorimeter
 306 and leakage into the hadronic calorimeter. This calibration is applied on both data and MC.
- 307 • Correction of calorimeter cell responce in data. This is allowing to get right responce in non-
 308 optimal HV-regions and exclude biases in a calorimeter electronics reconstruction.

Energy shift is parameterised, as:

$$E^{data} = E^{MC}(1 + \alpha_i), \quad (11.4)$$

where E^{data} and E^{MC} are the energies in data and simulation, respectivelly and α_i is a mean shift in a given bin i in η . Effect of this miscalibration on a reconstructed mass of Z boson is:

$$m_{i,j}^{data} = m_{i,j}^{MC}(1 + \alpha_{i,j}), \quad \alpha_{i,j} \sim \frac{\alpha_i + \alpha_j}{2} \quad (11.5)$$

309 neglecting second order terms. $m_{i,j}^{data}$ and $m_{i,j}^{MC}$ are reconstructed mass of Z boson in a i and j bins
 310 of η for data and MC respectivelly.

There is also a need to correct difference in a electron resolution. It can be desribed by a formula
 8.1. It is assumed, that sampling and noise terms are moddeled well by MC and the main diiference
 is coming from a constant term. So, the electron resoultion correction then can be written as:

$$\frac{\sigma_E^{Data}}{E_i} = \frac{\sigma_E^{MC}}{E_i} \oplus c_i \quad (11.6)$$

311 where c_i is η dependent relative resolution correction. Similarly to a energy scale correction it is
 312 possible to derive resolution correction factor by a comparing $m_{i,j}^{data}$ and $m_{i,j}^{MC}$ distribution.

313 Correction values of α_i and c_i are obtained via χ^2 fit on a invariant mass electrons for data and
 314 MC. Resulting energy scale is applied on a data, while resolution is corrected for MC. The resulting
 315 scale is validated on a $J/\psi \rightarrow ee$ and $Z \rightarrow ee\gamma$

316 11.3 Muon momentum correction

Muon momentum resolution is depending on a η , ϕ and p_T of the muon [?]. There is an empirical formula to describe it inside the detector (ID or MS):

$$\frac{\sigma_{Det}(p_T)}{p_T} = \frac{r_0^{Det}(\eta, \phi)}{p_T} \oplus r_1^{Det}(\eta, \phi) \oplus r_2^{Det}(\eta, \phi) \cdot p_T \quad (11.7)$$

317 The first term origins from fluctuations of energy loss in transversed material. Second r_1^{Det} is com-
 318 ing from magnetic field inhomogenities and local displacements. Third term r_2^{Det} describes intrinsic
 319 resolution effects.

Similarly to electrons, overall energy scale shift between data and MC parameterised as:

$$p_T^{data} = p_T^{MC} + s_0^{Det}(\eta, \phi) + s_1^{Det}(\eta, \phi) \cdot p_T^{MC}, \quad (11.8)$$

where $s_0^{Det}(\eta, \phi)$ is coming from the imperfect knowledge of energy losses for muons passing through detector.

This leads to a total correction formula:

$$p_T^{Cor,Det} = \frac{p_T^{MC,Det} + \sum_{n=0}^1 s_n^{Det}(\eta, \phi) (p_T^{MC,Det})^n}{1 + \sum_{m=0}^2 \Delta r_m^{Det}(\eta, \phi) (p_T^{MC,Det})^{m-1} g_m}, \quad (11.9)$$

where g_m are normally distributed random variables with mean 0 and width 1. Because small amount of material between interaction point and the ID, $\Delta r_0^{ID}(\eta, \phi)$ and $s_0^{ID}(\eta, \phi)$ are set to 0. Missalignment effect for an MS is corrected on a simulation level by adding a random smearing to an alignment constants. This is allowing to set $\Delta r_2^{MS}(\eta, \phi)$ to 0 during a fit.

The correction factors are extracted using $Z \rightarrow \mu\mu$ candidates events with requirement on a two combined muons. For correction invariant mass distribution $m_{\mu\mu}^I D$ and $m_{\mu\mu}^{MS}$ are considered individually within a specific $\eta - \phi$ region of fit. Combined muon parameters are used to obtain angles η, ϕ . The correction extraction is performed first for an ID and then for MS with addition of the fit variable:

$$\rho = \frac{p_T^{MS} - p_T^{ID}}{p_T^{ID}}, \quad (11.10)$$

which represents p_T imbalance in ID and MS.

In a second step corrections are propagated to the combined momentum, using a weight average:

$$p_T^{Cor,CB} = f \cdot p_T^{Cor,ID} + (1 - f) \cdot p_T^{Cor,MS}, \quad (11.11)$$

where weight f is derived from mc.

11.4 Number of vertexes reweighting

Another way is to apply weight on number of vertexes. This variables have a connection between each other, because each pp interaction produces separate vertex. Fig ?? shows a comparison of number of vertexes before and after reweighting.

This also allows to improve agreement for a E_T^{miss} distribution.

11.5 Hadron recoil correction

Remaining differences between Data and MC in E_T^{miss} are assumed to be due to a $\sum E_T$ mismodelling. This quantity is defined as a scalar sum of all active topo-clusters in the event:

$$\sum E_T = \sum_{i=0}^N |\vec{E}_T| \quad (11.12)$$

This variable is connected with a underlying event activity. However, $\sum E_t$ is correlated with a p_T^W , what is assumed to be simulated well in a MC. Correction procedure should leave this quantity untouched. In order to do this, correction factors are derived separately in a different p_T^W bins and flavours, since p_T^W is reconstructed differently in these channels. Size of this bin is chosen, so what each bin is having around 600 events in a data. Due to a limited statistics, data is approximated by a polynomial fit. Example of this fit is shown on a Figure ???. After this, weights are obtained by dividing resulting polynomial by a MC. Weights scale is corrected, so what number of events leaves unchanged in a MC:

$$\int \sum E_T^{MC} = \int \sum E_T^{MC} \cdot w, \quad (11.13)$$

³³⁴ where w are the weights. Total correction factors are shown on a Figure ?? The resulting $\sum E_T$
³³⁵ distribution shows good agreement with data, while p_T^W leaves unchanged. This correction is also
³³⁶ helping to improve E_T^{miss} distribution.

337

Chapter 12

338

Missing Transverse Energy reconstruction and 339 correction

340 Atlas detector has almost 4π coverage. This is allowing to calculate imbalance of energies inside
341 calorimeter, especially transversal part of it called E_T^{miss} . Neutrino from a $W \rightarrow l\nu$ decay is leaving
342 detector, without interacting with it, that is causing large energy imbalance in a detector.

Standard reconstruction of E_T^{miss} at ATLAS experiment uses transverse energy deposits in the calorimeter, energy losses in cryostat and reconstructed muons for a calculation:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,calo} + E_{x(y)}^{miss,cryo} + E_{x(y)}^{miss,muon}. \quad (12.1)$$

Calorimeter term is using information from reconstructed physics objects for calibration of cell response. The total transverse energy in calorimeter is defined as:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,\tau} + E_{x(y)}^{miss,jets} + E_{x(y)}^{miss,SoftTerm} + E_{x(y)}^{miss,\mu}. \quad (12.2)$$

343 where each term is calculated as the negative sum of the calibrated reconstructed objects, projected
344 onto the x and y directions. Each jet with energy $P_T > 20$ GeV is corrected for a pile-up and a jet
345 energy scale is applied. Soft term is calculated from topoclusters and tracks, that are not associated
346 with high-pt objects. To avoid double counting muon energy loss in calorimeter is subtracted
347 from E_T^{miss} . The E_T^{miss} muon term is calculated from the momenta of muons measured in a range
348 of pseudorapidity. Since pileup gives a significant effect on a E_T^{miss} performance several methods of
349 pileup suppression are used. This procedure was optimised for 8 TeV runs and using a calibration
350 constants from it. This can cause problems with 2.76 TeV low pileup run. As a examine showed this
351 is not optimal procedure in this case. Control plots for W production in electron and muon channels
352 are shown on a Fig. 12.3. Where a big discrepancies in a muon and electron channel, that cannot be
353 accounted to multijet background.

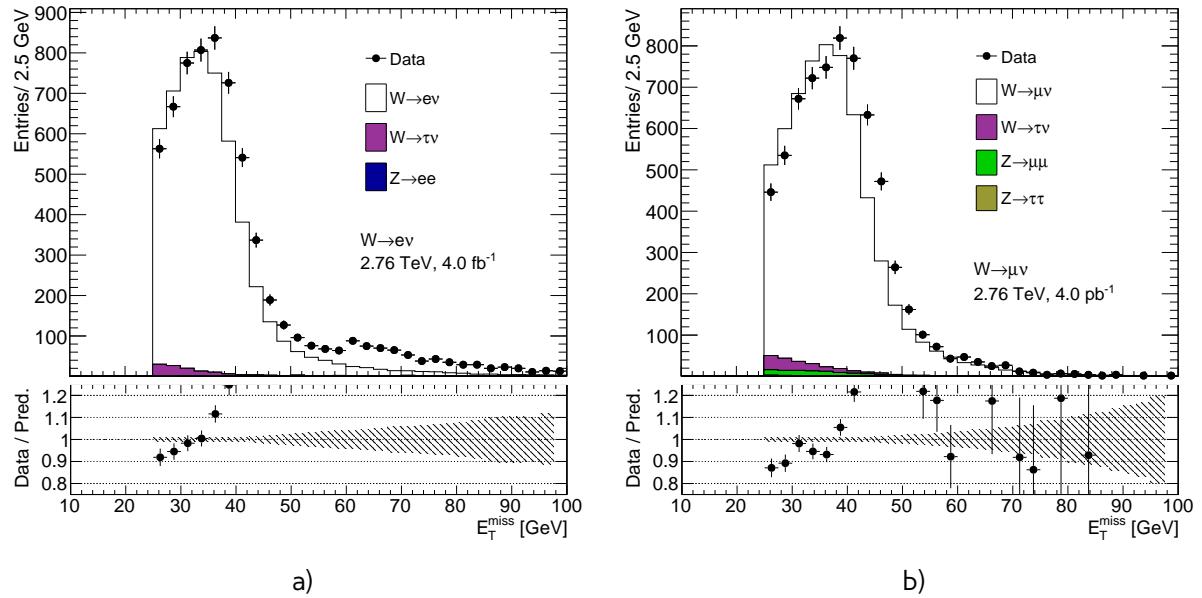


Fig. 12.1: Data and MC comparison for E_T^{miss} calculated by standard ATLAS algorithm for a) $W \rightarrow e\nu$ b) $W \rightarrow \mu\nu$ events

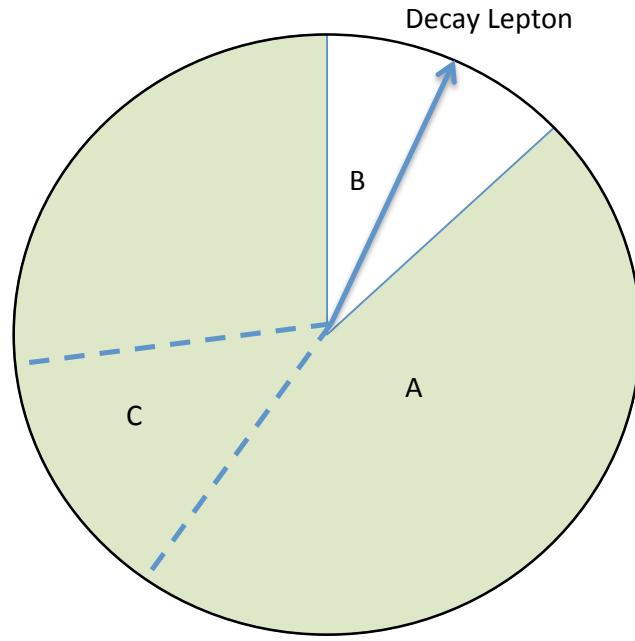


Fig. 12.2: Definition of different zones in the calculation of the cluster-based hadronic recoil.

354 12.1 Hadron Recoil Calculation

Second way of calculating E_T^{miss} was developed specifically for a W and Z decays by W mass measurements group. This procedure is using this fact, that a transverse momentum of a W-boson has to be balanced with initial (quark-gluon) state radiation, because initial sum of transverse momentum is zero:

$$\vec{P}_T^W = \vec{P}_T^{lep} + \vec{P}_T^\nu = \sum \vec{P}_T^{ISRquarks,gluon}, \quad (12.3)$$

where $\sum \vec{P}_T^{ISRquarks,gluon}$ is a transverse momentum of partons from initial state radiation, also called hadronic recoil (HR). Therefore, E_T^{miss} can be determined as:

$$E_T^{miss} = P_T^\nu = -HR + p_T^l \quad (12.4)$$

This procedure assumes, that recoil is arises from one single leading jet, and the rest is coming from a soft hadronic activity. This hadron recoil is computed as a vector sum of calorimeter clusters:

$$HR = \sum_{i=0}^{N_{topo}} \vec{p}_T^{topo} \quad (12.5)$$

while a scalar sum of all transverse energies is corresponding to the hadronic activity of the event:

$$\sum E_T = \sum_{i=0}^{N_{topo}} E_T^{topo} \quad (12.6)$$

355 To avoid double counting of lepton energy losses in calorimeter, the clusters inside cone with radius
 356 $dR = 0.2$ are excluded from this calculation. To compensate soft activity inside this cone, clusters
 357 are then compensated by replacement cone (Fig. 12.2). This cone is defined as cone at the same
 358 pseudorapidity, but different ϕ . It should be far from any other lepton and hadron recoil direction.
 359 Each cone is then rotated to a direction of the original lepton direction. This definition is not taking
 360 into account jet reconstruction aspects. This is allowing to get a better data MC agreement (Fig. ??).

361 where \vec{v} is the unit vector along boson p_T direction. Parallel component is sensitive to a bias
 362 relative to a truth p_T of the boson, while perpendicular component is 0 on average and affected just
 363 by a resolution effects. Momentum of Z boson can be precisely determined by a measurement of
 364 its decay products. This is allowing to determine data driven hadron recoil corrections. Due to a low
 365 cross-section of Z production, these corrections are mostly statistically dominated. On another hand,
 366 it is also possible to use W boson decays for a not so systematically precise, but statistically better.

367 However, due to a limited statistics of a Z decays in this data, it is better to combine this with a
 368 corrections derived from W decays da

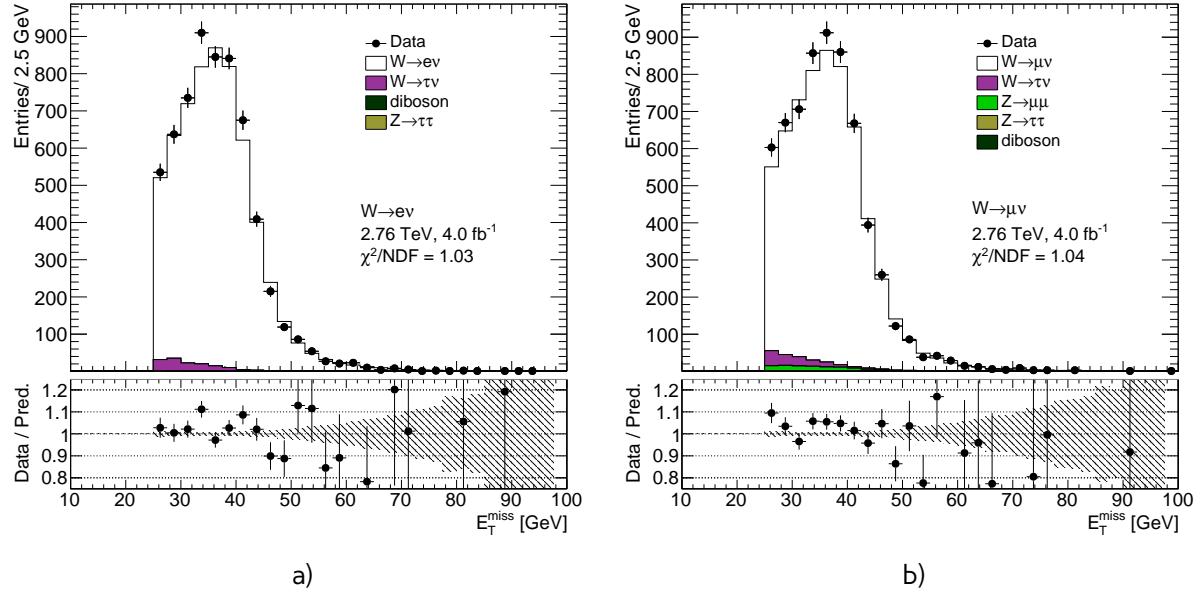


Fig. 12.3: Data and MC comparison for E_T^{miss} calculated from hadron recoil for a) $W \rightarrow e\nu$ b) $W \rightarrow \mu\nu$ events

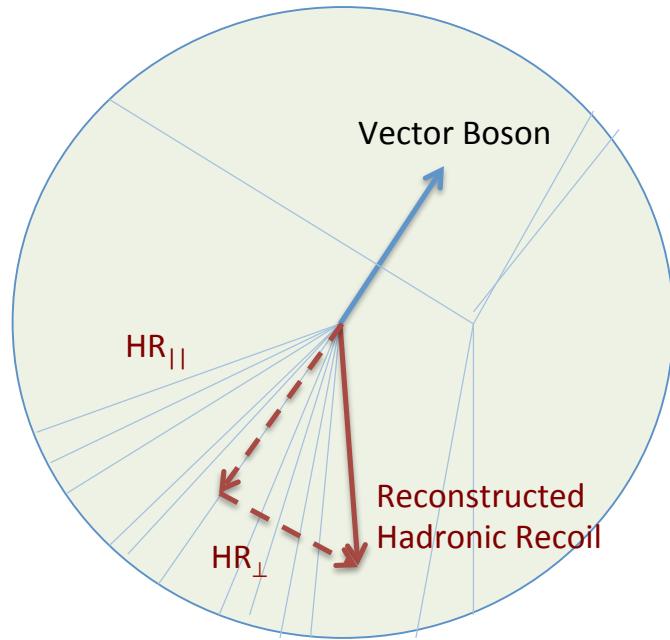


Fig. 12.4: Parallel and perpendicular projection of the hadronic recoil with respect to the transverse momentum of the vector boson

369 12.2 Hadron Recoil calibration

370 E_T^{miss} affects significantly on a W boson measurement, so its important to have good understanding
 371 of sources of a possible differences in a hadron recoil reconstruction in a data and monte carlo.

This differences can be paramerised at the respect of the truth boson Pt as follows:

,

(12.7)

372 where is a projection to the direction of the truth boson pt, and is the projection on a perpendicular
 373 plane. On average should be match absolute value of boson pt and thus this quantity is sensitive to
 374 a relative bias of hadron recoil reconstruction. On another hand comes from calorimeter resolution
 375 and should have 0 mean. This component is sensitive to a discrepancies in data/mc resolutions.
 376 Standard procedure of calibrating hadron recoil uses Z boson, since its transverse momentum can
 377 be determined not only by a hadron recoil, but also from its decay products. Zpt resolution coming
 378 from lepton reconstruction is around 10 GeV, while hadron recoi resolution is around 30 GeV, what
 379 is alowing to have a precise determination of a calibration constants. On another hand, size of the Z
 380 sample in 2.76 TeV data is small, so this constants will be statistically dominated. On another hand, in
 381 W decays hadron recoil corrections cannot be determined through boson pt directly, but this channel
 382 has a better statistics. The procedure is more complicated, because it should not change truth
 383 boson spectrum and not correct additionally boson pt mismodelling. At the end, the combination of
 384 2 determination results have been used. This section describes a procedure of calibrating bias and
 385 resolution mismodelling in a hadron recoil, that was apapted for 2.76 TeV data.

386 12.2.1 Underlying event mismodelling correction

One of the possible sources of diffetences in data and MC is a modelling of underlying event. Soft interactions cannot be computed precisely by a pertrubative quantum chromodynamics methods, so it is usually described by a phenomenological models, that can be not so precise. Pile-up mismodelling usually accounted by correcting average number of interaction per bunch crossing to match a data. However, ATLAS simulation is suited for an ordinary high pile-up runs, so this quantity is not modelled well in a case of 2.76 TeV analysis ???. Another quantity that depends on a underlying event is a $\sum E_T$, defined in a Eq. 12.6. There is a visible shift between data and monte carlo in a both channels. This is causing differences in a hadron recoil resolution, since this variables are highly correlated (add some figure). Size of the Z sample is not sufficient to determine the correction factors, so the W boson was used. To leave boson pt spectrum untouched correction factors are determined inside pt bins as:

$$SF = \frac{\sum E_T^{data}(p_T^{W,rec})}{\sum E_T^{MC}(p_T^{W,rec})} \quad (12.8)$$

387 12.2.2 Hadron recoil bias correction

388 One of the possible sources of hadron recoil bias could be subtraction of the lepton induced energy
 389 in the calorimeter.

390 Chapter 13

391 **Background estimation**

392 After the event selection described in chapter 10 the background contribution is around % for W
393 and % for Z analysis (which is with this statistics is negligible). Main backgrounds for W analysis
394 are coming from:

- 395 • Processes with τ lepton, misidentified as a electron or muon + missing energy from neutrino
396 • Z decays with one missing lepton.
397 • QCD processes. In electron channel this is mostly coming from jets faking electrons, while in a
398 muon channel it consists mostly of a real muons produced in decays of heavy-flavor mesons.
- 399 Most of the backgrounds are estimated using MC. They are normalized using highest cross-section
400 order available. The total list of simulated backgrounds and its cross-section is shown in a Table 13.1

Table 13.1: Background processes with their associated cross sections and uncertainties. The quoted cross sections are used to normalise estimates of expected number of events

Process	$\sigma \cdot BR$ [pb]	Order
$W^+ \rightarrow e\nu$	2116(1.9)	NNLO
$W^+ \rightarrow \mu\nu$	2116(1.9)	NNLO
$W^+ \rightarrow \tau\nu$	2116(1.9)	NNLO
$W^- \rightarrow e\nu$	1267(1.0)	NNLO
$W^- \rightarrow \mu\nu$	1267(1.0)	NNLO
$W^- \rightarrow \tau\nu$	1267(1.0)	NNLO
$Z \rightarrow ee$	303(0.2)	NNLO
$Z \rightarrow \mu\mu$	303(0.2)	NNLO
$Z \rightarrow \tau\tau$	303(0.2)	NNLO
$t\bar{t}$	7.41	LO
WW	0.6	LO
ZZ	0.7	LO
WZ	0.2	LO
$DYee$	2971	LO
$DY\mu\mu$	2971	LO

401 Due to a large cross-section uncertainty and high amount of events needed to be generated, QCD
402 background is estimated using a data driven method.

403 **13.1 Data-driven method**

404 **13.2 Uncertainties estimation**

405 Chapter **14**

406 **Uncertainties**

407 **14.1 Toy MC method**

408 **14.2 Experimental systematic uncertainties**

409 table about methods

410 **14.2.1 Hadron Recoil correction uncertainty**

411 **14.3 Theoretical uncertainty**

412 • PDF

413 • Parton shower and matrix element

414

Chapter 15

415 **Control distributions**

416 Chapter **16**

417 **Results of the Cross Section Measurement**

418 **16.1 Cross-Section measurement definition**

419 **16.2 Fiducial phase-space**

420 **16.3 Comparation with Theoretical Predictions**

421

Part IV

422

PDF fits

