

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.

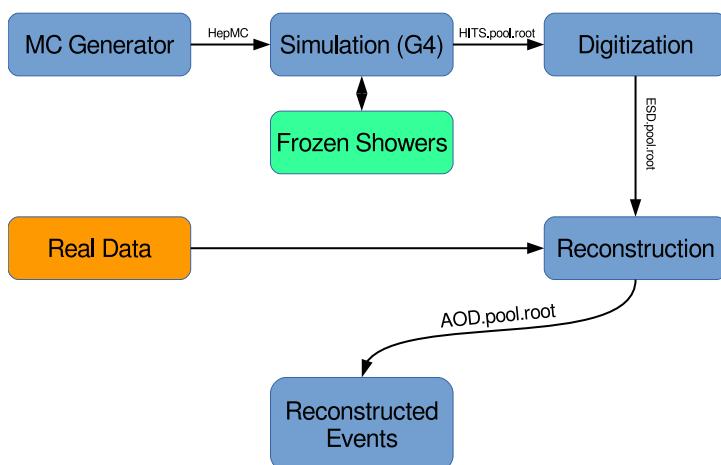
28 **8.1 Monte Carlo simulation at ATLAS experiment**

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**

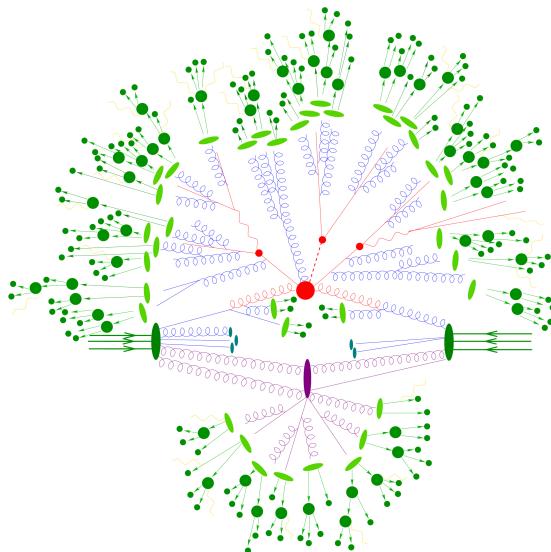


Fig. 8.2: Schematic view of a $t\bar{t}H$ event produced in a pp -collision: the hard scattering is shown as a red blob with the solid and dashed lines as the resulting three particles. Independently happening multi-particle interactions are indicated by the violet blob. Parton showers are shown with 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 [6] 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 [7] Powheg is generator with NLO ME [2], that can be interfaced with other generator(such
77 as Pythia or Herwig) for higher precision of showering.

78 Pythia [8] 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 [3] for hadronisation.

82 Herwig [5] 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 [1]. 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

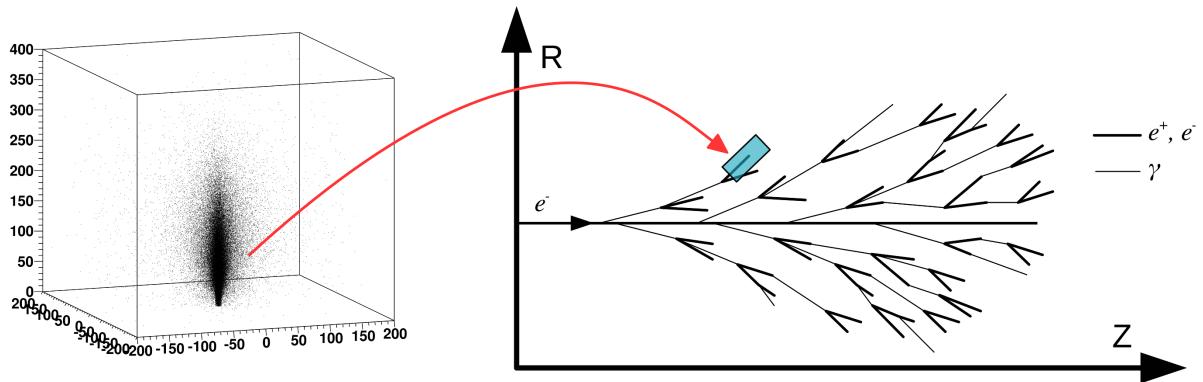


Fig. 8.3: Diagram showing the shower substitution of the low-energy particle, during the high-energy particle simulation.

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.

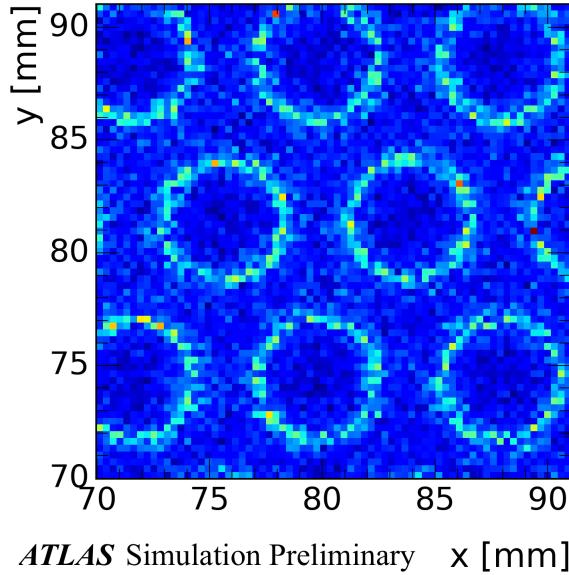


Fig. 8.4: Average energy response in a calorimeter vs x,y for electrons with energy less than 1 GeV

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:

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

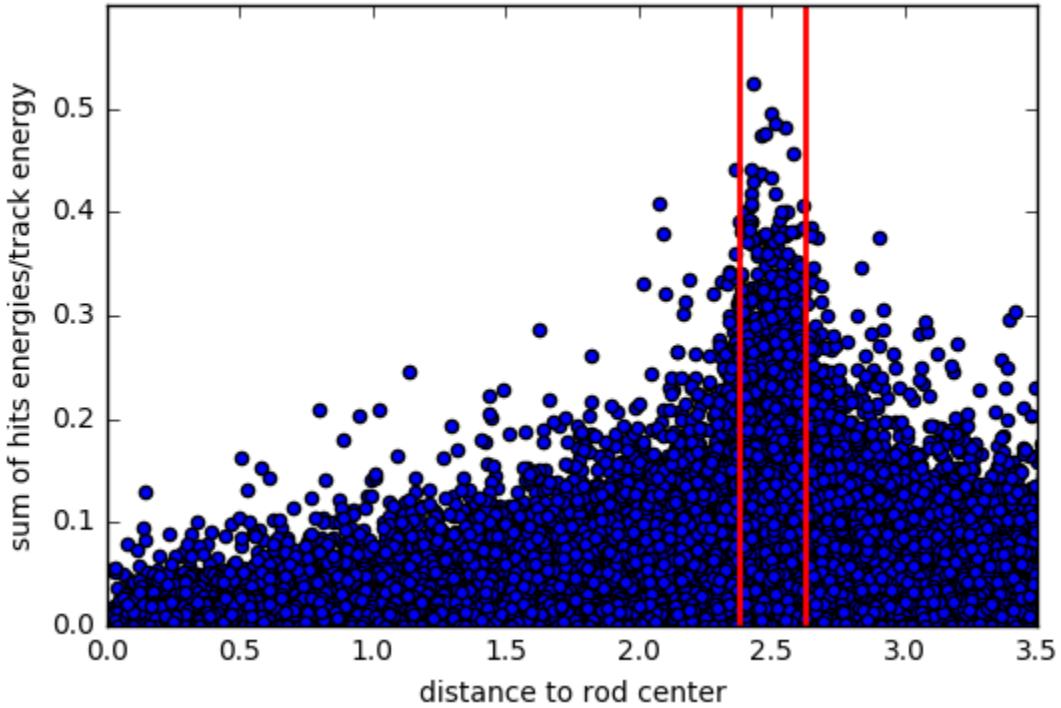


Fig. 8.5: Energy response for electrons in a calorimeter for all electrons in a library

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

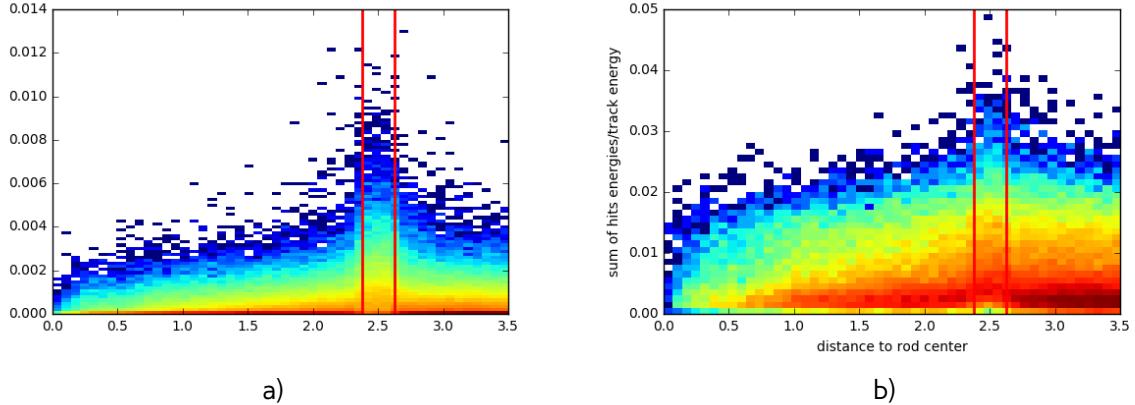


Fig. 8.6: Energy response for electrons in a calorimeter for a) electrons with energy less than 100 MeV b) electrons with energy bigger, than 100 MeV

| 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.

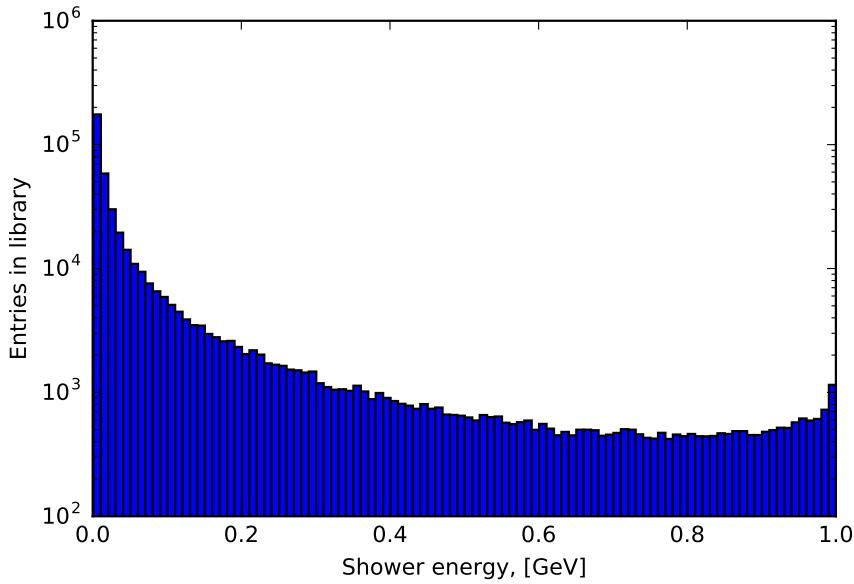


Fig. 8.7: Distribution of shower energy used in production of 1000 GeV electrons.

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.

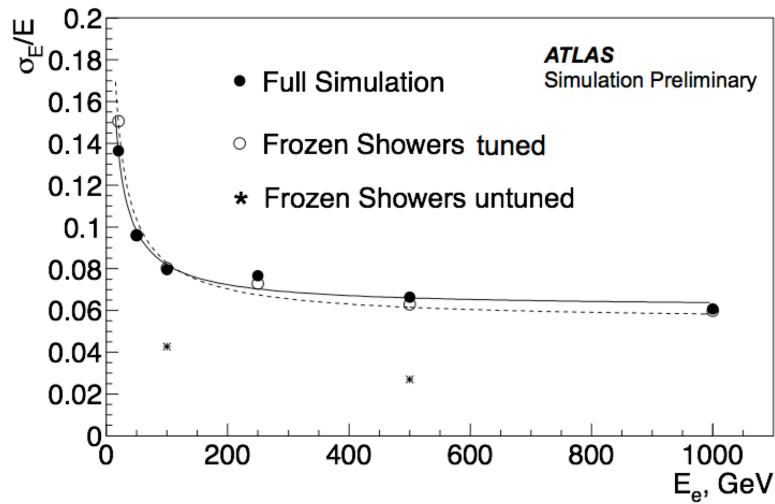


Fig. 8.8: Electron resolution for full simulation, tuned and untuned frozen showers

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.

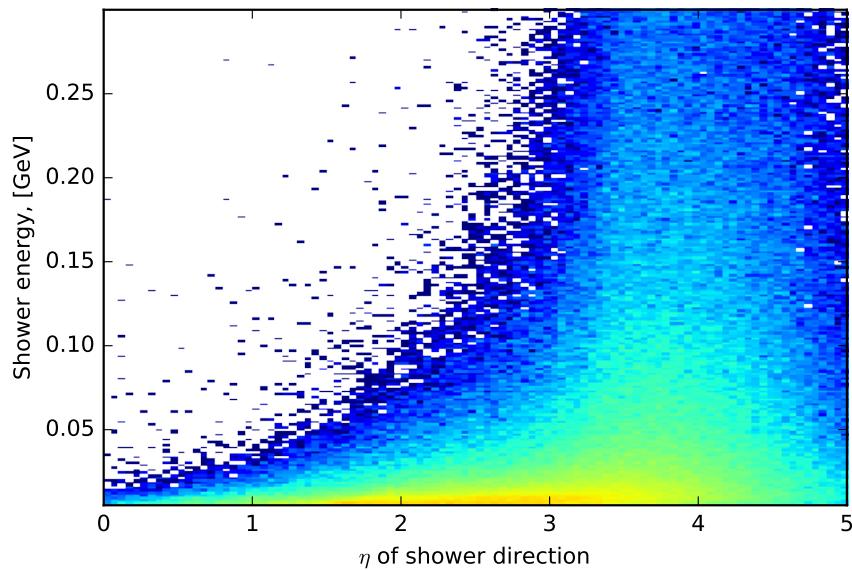


Fig. 8.9: Distribution of showers used in production of 1000 GeV electrons on shower energy vs $\eta_{momentum}$ plane.

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.

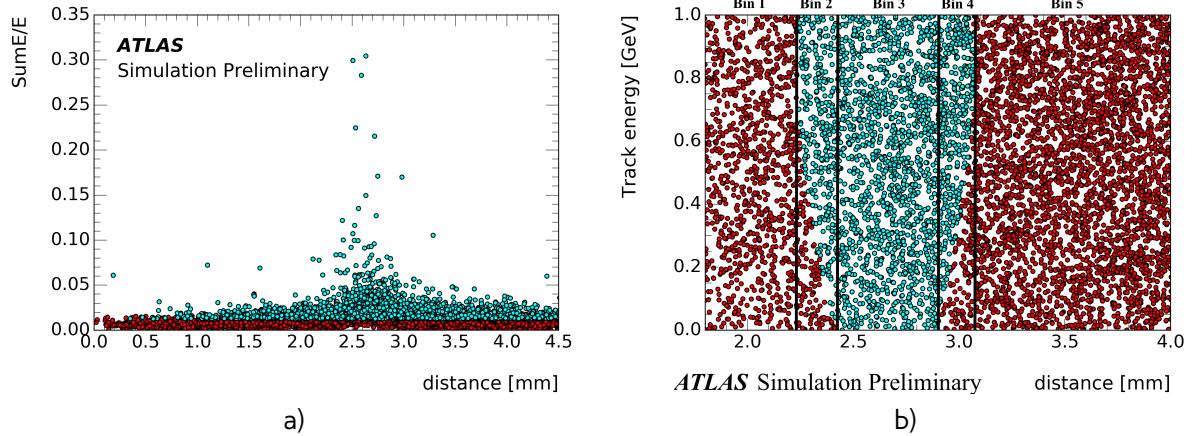


Fig. 8.10: Results of machine learning for a) first classifier b) second classifier. Cyan dots are corresponding to sensitive material showers, red - dead material showers

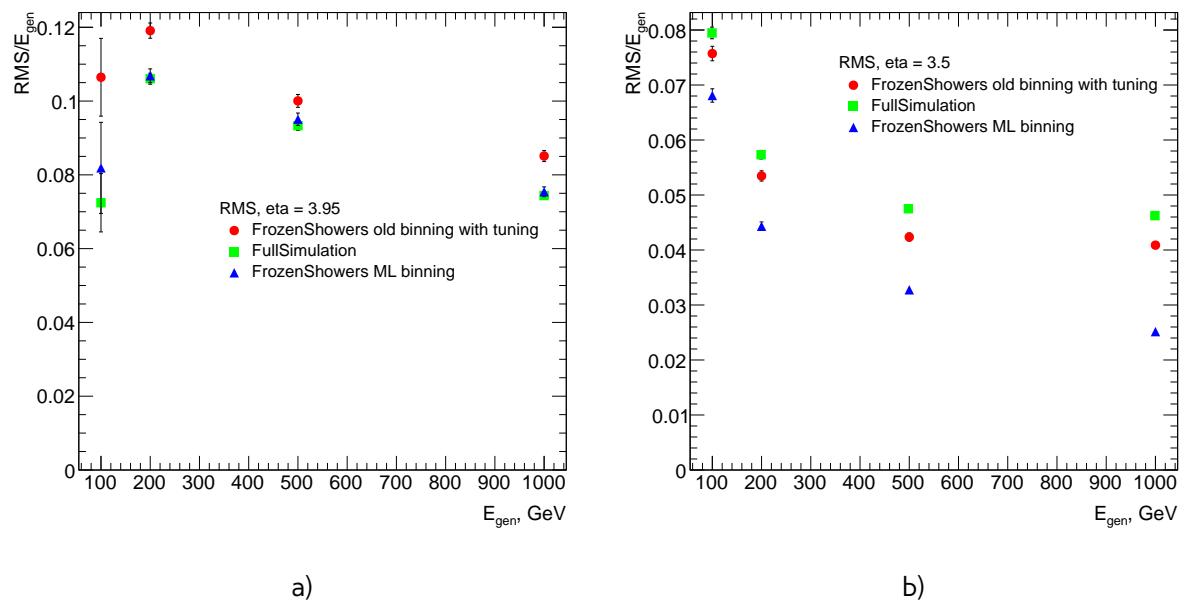


Fig. 8.11: Resolution of reconstructed electrons for full simulation, new libraries with ML binning and old tuned libraries with original binning for a) eta = 3.95 b) eta = 3.5

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:

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

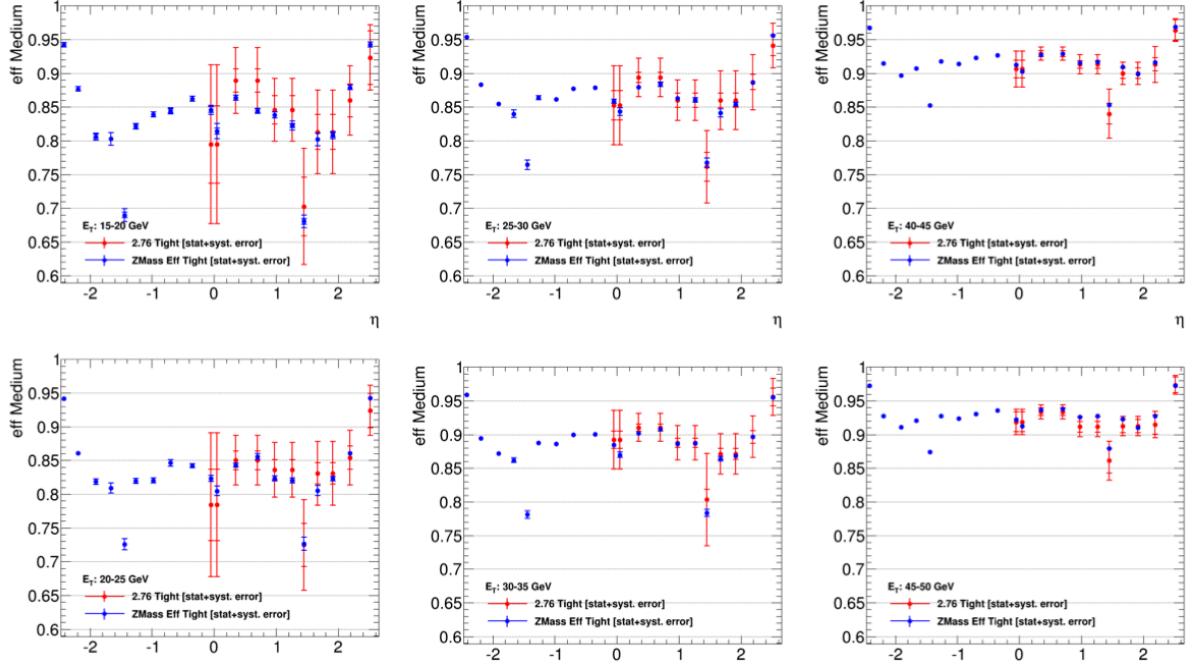


Fig. 11.1: Comparison of electron efficiencies as calculated for 8TeV (blue points) and 2.76TeV (red points) for MC simulation. Efficiencies are shown as a function of pseudorapidity (η) for different electron E_T bins. Both statistical and systematic uncertainties are shown.

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 is calculated as a single number in order to minimize systematics contribution.

300 11.2 Electron energy scale and resolution

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

- 303 • Electronic calibration, that transfers a raw signal from a readout to a cluster energy deposit.
- 304 • MC based calibration. It corrects effects of energy loss in the material in front of calorimeter
 305 and leakage into the hadronic calorimeter. This calibration is applied on both data and MC.
- 306 • Correction of calorimeter cell response in data. This is allowing to get right response in non-
 307 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, respectively 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:

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

308 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
 309 of η for data and MC respectively.

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

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

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

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

315 11.3 Muon momentum correction

Muon momentum resolution is depending on a η , ϕ and p_T of the muon [4]. There is an empirical formula to describe it inside the detector (ID or MS):

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

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

Additional particle interactions can affect many different distributions, including E_T^{miss} . Pile-up modelling 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 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.

336 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 this 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 a dividing resulting polynomial by a MC. Weights scale is corrected, so what number of events leaves unchanged in a MC:

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

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

340

Chapter 12

341

Background estimation

342

12.1 Data-driven method

343

12.2 Uncertainties estimation

344 Chapter **13**

345 **Uncertainties**

346 **13.1 Toy MC method**

347 **13.2 Experimental systematic uncertainties**

348 table about methods

349 **13.2.1 Hadron Recoil correction uncertainty**

350 **13.3 Theoretical uncertainty**

351 • PDF

352 • Parton shower and matrix element

353

Chapter 14

354

Control distributions

355

Chapter 15

356 Results of the Cross Section Measurement

357 15.1 Cross-Section measurement definition

358 15.2 Fiducial phase-space

359 15.3 Comparation with Theoretical Predictions

360

Part IV

361

PDF fits

³⁶² Bibliography

- ³⁶³ [1] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl.Instrum.Meth.*, A506:250–303, 2003.
- ³⁶⁴ [2] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO vector-boson production matched
³⁶⁵ with shower in POWHEG. *JHEP*, 0807:060, 2008.
- ³⁶⁶ [3] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String
³⁶⁷ Dynamics. *Phys.Rept.*, 97:31–145, 1983.
- ³⁶⁸ [4] ATLAS collaboration. The ATLAS experiment at the CERN Large Hadron Collider. *JINST*, 3, 2008.
- ³⁶⁹ [5] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour, and
³⁷⁰ B.R. Webber. HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering
³⁷¹ Gluons (including supersymmetric processes). *JHEP*, 01:010, 2001.
- ³⁷² [6] S.D. Drell and Tung-Mow Yan. Massive Lepton Pair Production in Hadron-Hadron Collisions at
³⁷³ High-Energies. *Phys.Rev.Lett.*, 25:316–320, 1970.
- ³⁷⁴ [7] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*,
³⁷⁵ 0411:040, 2004.
- ³⁷⁶ [8] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*,
³⁷⁷ 0605:026, 2006.