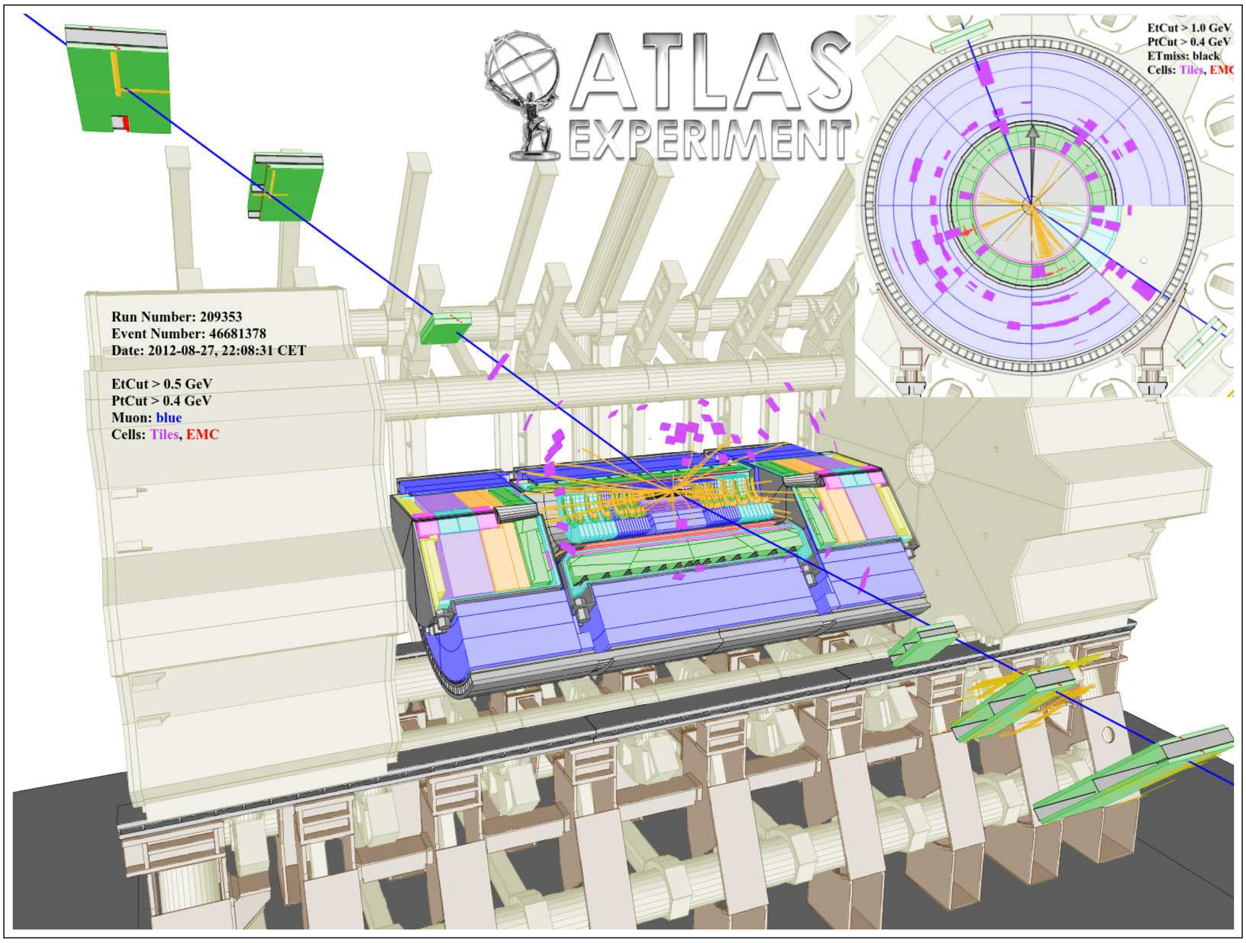
Studying the Z Boson with the ATLAS Detector at the LHC

Version 0.5 (December 21, 2018)



Falk Bartels, Julia I. Djuvsland, Manuel P. Geisler, Martin Wessels

Contact: martin.wessels@cern.ch

Advanced Physics Lab for Physicists
Heidelberg University

Contents

1 Preliminary Remarks

In this lab course we will study the Z boson using proton-proton collision data recorded with the ATLAS detector at the CERN Large Hadron Collider. For this, a set of real data and several supporting simulated Standard Model processes are available. Central techniques for performing a typical analysis in modern particle physics will be acquired. This lab course was introduced first in the winter semester 2018/19.

1.1 Brief Overview

The goal of this lab course is to receive an impression of a data analysis as performed in modern particle physics. Particle physics [1] is concerned with the fundamental constituents of the Universe, the elementary particles, and the interactions between them, the forces. A modern particle physics experiment, such as the ATLAS experiment [2], requires years of design and planning, research and development, and finally construction of a detector the size of a house. And of course it needs a particle accelerator such as the Large Hadron Collider, LHC, at CERN (Conseil Européen pour la Recherche Nucléaire – European Council for Nuclear Research) located astride the Franco-Swiss border near Geneva.

Due to the complexity of the machines and detectors, research in particle physics nowadays is performed in large collaborations. ATLAS for instance comprises about 3000 scientific authors from 181 institutions around the world, representing 38 countries from all the world's populated continents.

The ATLAS collaboration has released an official dataset [3], [4], open to the public for educational use only, which will be used in this lab course. The dataset consists of real proton-proton collision data which were recorded with the ATLAS detector at the LHC in 2012. While these exact data were also part of the discovery of the famous Higgs boson that same year[1] the released statistics corresponding to an integrated luminosity of 1 fb^{-1} is not sufficient to extract the signal of the Higgs decay. Instead, we are using another boson for our investigations in this lab course, the Z boson.

The Z boson is a neutral elementary particle which – along with its electrically charged cousin, the W – carries the weak force. Same as the Higgs boson, it was also discovered at CERN, in 1983 at the Super Proton Synchrotron (SPS), operated as proton-antiproton collider. Owing to its electrical neutrality, the Z boson must decay into a particle-antiparticle pair. In this lab course we are making use of the clean detector signatures generated by Z bosons decaying into electron-positron and muon-antimuon pairs, as illustrated in figure 1. We will use the same methods and tools as used in a "real analysis" in order to extract these events out of the bulk of recorded data, reconstruct the Z boson candidates and determining its mass with fitting technique.

Given this lab course is a data analysis, all work will be done with a computer. The main tasks will consist in the development and coding of an analysis code to extract the Z candidate events and perform the various measurements and plots. For this we will be using the Python programming language, interfacing the data analysis package ROOT. Both belong to the standard tools in modern physics data analysis, not only in the field of particle physics.

[1] On 4 July, the ATLAS and CMS collaborations announced they had each observed a new particle in the mass region around 125 GeV consistent with a Higgs boson, the particle linked to the mechanism proposed in the 1960s to give mass to the W, Z and other particles.

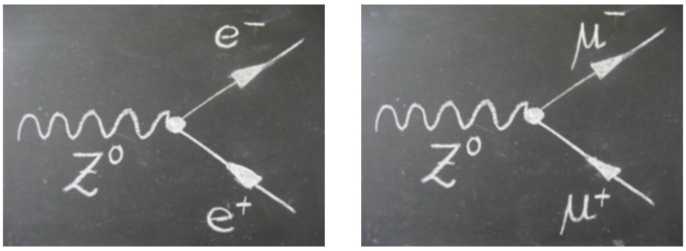


Figure 1: Leading order Feynman diagrams of the Z boson decay into electron and muon pairs [5].

1.2 Prerequisites

Um diesen Versuch sinnvoll bearbeiten zu können, werden Grundkenntnisse der Teilchenphysik benötigt. Daher wird der Stoff der PEP4 Vorlesung als bekannt voraus gesetzt.

Die Arbeit bei diesem Versuch besteht im Wesentlichen in der Programmierung eines Analysecodes mit Python. Daher sollte grundsätzliches Interesse an einer solchen Tätigkeit vorhanden sein. Grundkenntnisse in der Programmierung mit Python sollten vorhanden bzw. vor dem Beginn des Versuchs selbständig erlernt werden (siehe unten).

The course can in principle be booked without these prerequisites, but we expect the students to acquire the missing knowledge elsewhere. It is not covered in these instructions, but will be needed to successfully complete this course.

1.3 Preparation

Neben der allgemeinen Wiederholung des PEP4 Stoffes zur Teilchenphysik sollten insbesondere die Eigenschaften des Z-Bosons bekannt sein. Da Kollisionsdaten des LHC untersucht werden, spielen auch Streureaktionen von Quarks eine vordergründige Rolle. Die grundlegende Funktionsweise des LHC Beschleunigers sowie des ATLAS Detektors sollten vorbereitet werden.

We are using the programming language Python in this lab course. Basic programming skills in Python have to be taken for granted and cannot be taught within the limited time available in this lab course. It is urgently required that you refresh or learn basic commands and coding elements in Python before starting this lab course.

An interactive introductory course into Python is offered within the framework of the beginner's lab and can be found here:

`https://www.physi.uni-heidelberg.de/Einrichtungen/AP/Python.php` .

Another excellent tutorial is being provided by the Python Software Foundation itself and can be accessed through following link:

`https://docs.python.org/3/tutorial/index.html` .

Neben Python wird das Datenanalyse Framework ROOT benutzt. Dokumentation sowie Tutorials können unter http://root.cern.ch gefunden werden. Dort befindet sich auch ein vollständiges (und sehr ausführliches) Benutzerhandbuch … aber die meisten lernen ROOT "'by doing"'.

Leptons			Quarks		
Particle	Q	Mass [GeV]	Particle	Q	Mass [GeV]
electron (e^-)	-1	0.0005	down (d)	-1/3	0.003
neutrino (ν_e)	0	$< 10^{-9}$	up (u)	+2/3	0.005
muon (μ^-)	-1	0.106	strange (s)	-1/3	0.1
neutrino (ν_μ)	0	$< 10^{-9}$	charm (c)	+2/3	1.3
tau (τ^-)	-1	1.78	bottom (b)	-1/3	4.5
neutrino (ν_τ)	0	$< 10^{-9}$	top (t)	+2/3	174

Table 1: The twelve fundamental particles of the SM and their charges (Q) and masses.

Boson	Mediated Force	Q	Spin	Mass [GeV]
Gluon (g)	Strong	0	1	0
Photon (γ)	Electromagnetism	0	1	0
W Bosons ($W^\pm$)	Weak	$\pm$1	1	80.4
Z Boson (Z)		0	1	91.2

Table 2: The four force-carrying bosons of the SM.

2 Theoretical Foundations

This chapter provides a very brief introduction into the underlying physics concepts of this lab course.

2.1 The Standard Model of Particle Physics

Today, the knowledge gained by scientists about the structure of matter is summarised in the Standard Model of particle physics, which has been developed in the second half of the last century and has proved very successfully in describing all experimentally results in the field of high energy physics. The basic assumptions of the Standard Model are simple and can be written down in a few lines [6].

Matter is composed of elementary fermions with spin $\frac{1}{2}$, quarks and leptons, each occurring in three families consisting of two particles. Each of the 12 particles has a corresponding anti-particle with opposite charge but otherwise identical properties. A summary of the twelve fundamental particles of the Standard Model is given in table 1.

These elementary particles are subject to three fundamental forces, which are the strong, the electromagnetic and the weak force. Gravity, the forth known fundamental interaction, has not yet been included in the Standard Model. The interactions between the particles composing matter are mediated by bosons carrying spin 1. These exchange-particles are the massless photon for the electromagnetic force, the massive Z and $W^\pm$ bosons for the weak force and eight massless gluons mediating the strong interaction. A summary is shown in table 2. The third kind of particle in the Standard Model is the Higgs boson, which is responsible for the creation of particle masses.

Towards a grand unification of the fundamental forces, the electromagnetic and weak force are combined in the Standard Model into the electroweak interaction. The gauge theory describing the strong interaction is the Quantum Chromodynamics.

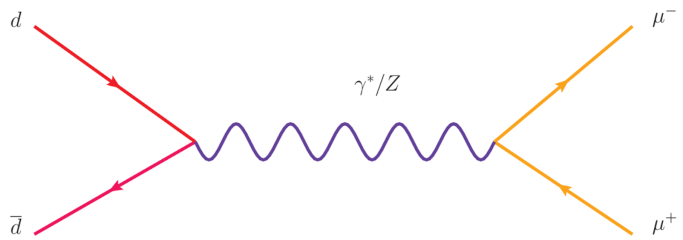


Figure 2: A simple Drell-Yan diagram.

2.2 A Short History of the Z Boson

Sorry, still to come.

2.3 The Drell-Yan Process at the LHC

Inspired by https://www.quantumdiaries.org/2015/05/18/dy-resummation/

Particle colliders like the Large Hadron Collider (LHC) are, in a sense, very powerful microscopes. The higher the collision energy, the smaller distances can be studied.

One of the most important processes that occurs in proton collisions is the Drell-Yan process. When a quark (e.g. a down quark d) from one proton and an anti-quark (e.g. an anti-down quark $\bar{d}$) from an oncoming proton collide, they can annihilate into a virtual photon (γ) or Z boson if the net electric charge is zero (or a W boson if the net electric charge is one). After briefly propagating, the photon/Z can split into a lepton and its anti-particle partner, for example into a muon and anti-muon or electron–positron pair. In pictures, the quark anti-quark annihilation into a lepton anti-lepton pair looks like figure 2. By the conservation of momentum, the sum of the muon and anti-muon momenta will add up to the photon/Z boson momentum.

This is nicely illustrated in figure 3 where the invariant mass distribution for any muon-anti-muon pair produced in proton collisions at the 7 TeV LHC is plotted, as measured by the CMS experiment. The rightmost peak at about 90 GeV corresponds to the production Z bosons. The other peaks represent the production of similarly well-known particles in the particle zoo that have decayed into a muon anti-muon pair. The clarity of each peak and the fact that this plot uses only about 0.2% of the total data collected during the first LHC data collection period (Run I) means that the Drell-Yan process is a very useful tool for calibrating the experiments. If the experiments are able to see the Z boson, the rho meson, etc., at their correct energies, then we have confidence that the experiments are working well enough to study nature at energies never before explored in a laboratory.

However, the Drell-Yan process is not as simple as drawn in figure 2. Real collisions include the remnants of the scattered protons. An example for a more realistic Drell-Yan interaction is depicted in figure 4, in which in addition to the proton remnants, gluon radiation processes are included. This will lead to more particles in the final state, in particular jets of hadrons produced by the hadronization of quarks and gluons.

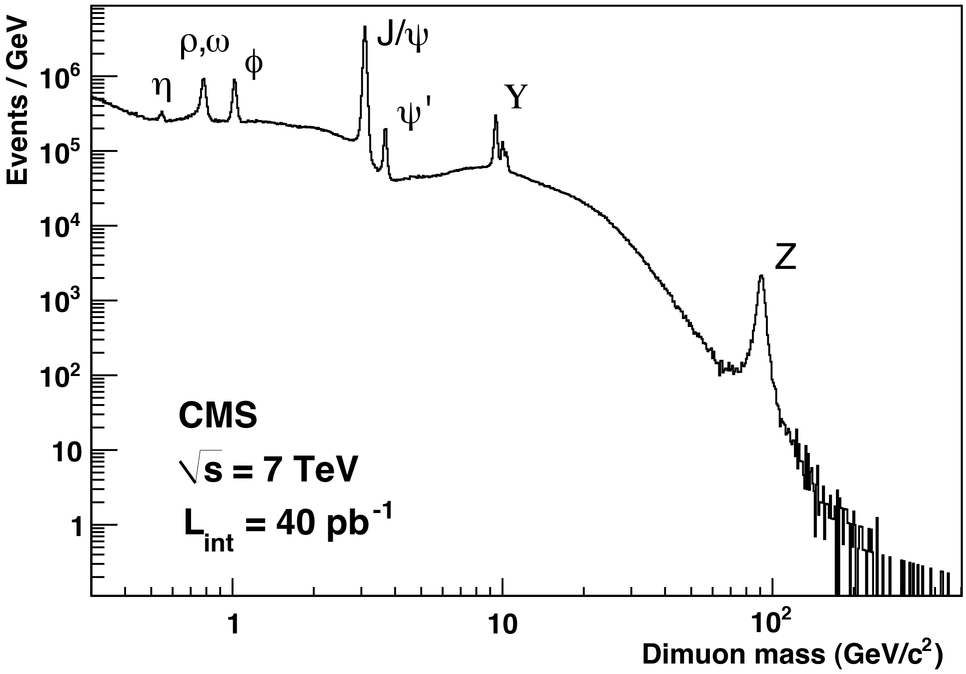


Figure 3: CMS invariant-mass spectra of opposite-sign muon pairs using 2010 and 2011 data.

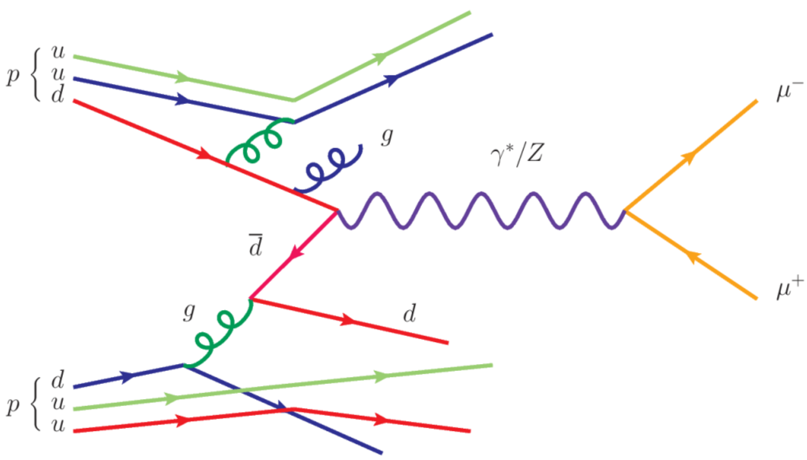


Figure 4: More realistic Drell-Yan diagram.

3 The Large Hadron Collider and the ATLAS Experiment

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the world's current most powerful particle collider. It is located up to 175 m underground at the European Centre for Nuclear Research (CERN) in Geneva, Switzerland. Inside the accelerator ring with its almost 27 km circumference and state-of-the-art magnets, protons[2] are accelerated to energies[3] of $\sqrt{s} = 13$ TeV. The design luminosity of the LHC is 10^{34}cm^{-2}s^{-1}.

A chain of accelerators, mostly stemming from previous CERN experiments, pre-accelerates the protons to 450 GeV before injecting the particles into the LHC. The LHC hosts two beams in separate beam pipes, one circulating clockwise the other counter-clockwise. At interaction points (IPs) the beams are crossed, particles collide and experiments measure the processes.

The beams are structured into bunches. Each beam $i = 1, 2$ has $n_b = 3564$ possible bunch slots and every filled bunch slot contains typically $N_i = \mathcal{O}(10^9)$ protons each. Successive bunch slots are temporally spaced by 25 ns, which is a result of the circumference of the ring, the total number of bunch slots and the speed of light. The revolution frequency f_{rev} of a bunch around the whole ring is 11.2 kHz.

The lattice[4] uses high-end technologies. Superconducting magnets with a temperature of roughly 2 K are used to keep the protons on track. The large energies involved require dipole magnets of up to 8 T and currents of about 11 kA. Beam bending is accomplished by 1232 dipole magnets, beam focusing by 382 quadrupole magnets and 3700 additional magnets take care of various further tasks, such as the beam dump at the end of a run.

Four main experiments are built at different IPs: ATLAS and CMS are multi-purpose experiments with a broad physics program ranging from precision measurements of Standard Model (SM) interactions to searches for Dark Matter. LHC-b is a smaller experiment specialized in measurements of the flavour sector of the SM. Lastly, ALICE focuses on the mode of operation when heavy ions rather than protons are injected into the LHC. A schematic illustration of the accelerator ring and the four experiments can be found in Figure 5.

A quantity of special relevance for collision experiments is the instantaneous luminosity $\mathcal{L}$, which is a measure for the abitility of a particle accelerator to produce the required number of interactions. It is defined as

$$\mathcal{L} = \frac{N_1 N_2 f_{\mathrm{rev}} n_b}{4\pi\sigma_x\sigma_y}\,. \tag{1}$$

The subscripts 1 and 2 refer to the two beams. σ_x and σ_y denote the horizontal (x) and vertical (y) beam spread and are a measure for how focused the beams are[5]. A related quantity is the integrated luminosity, which yields the size of a collected data set by a run. It is defined as

$$\mathcal{L}_{int} = \int \mathcal{L}\; dt\,. \tag{2}$$

---

2 Heavy ions can be injected into the machine, too. For this lab course this is however irrelevant.

3 This energy has been achieved during Run-2 of operation of the LHC, which lasted from 2015 to 2018. In the previous Run-1 energies of up to $\sqrt{s} = 8$ TeV were reached. The design centre-of-mass collision energy is 14 TeV and is expected to be reached in the future Run-3.

4 The optics setup of an accelerator.

5 This formulation assumes a Gaussian intensity profile of the beam

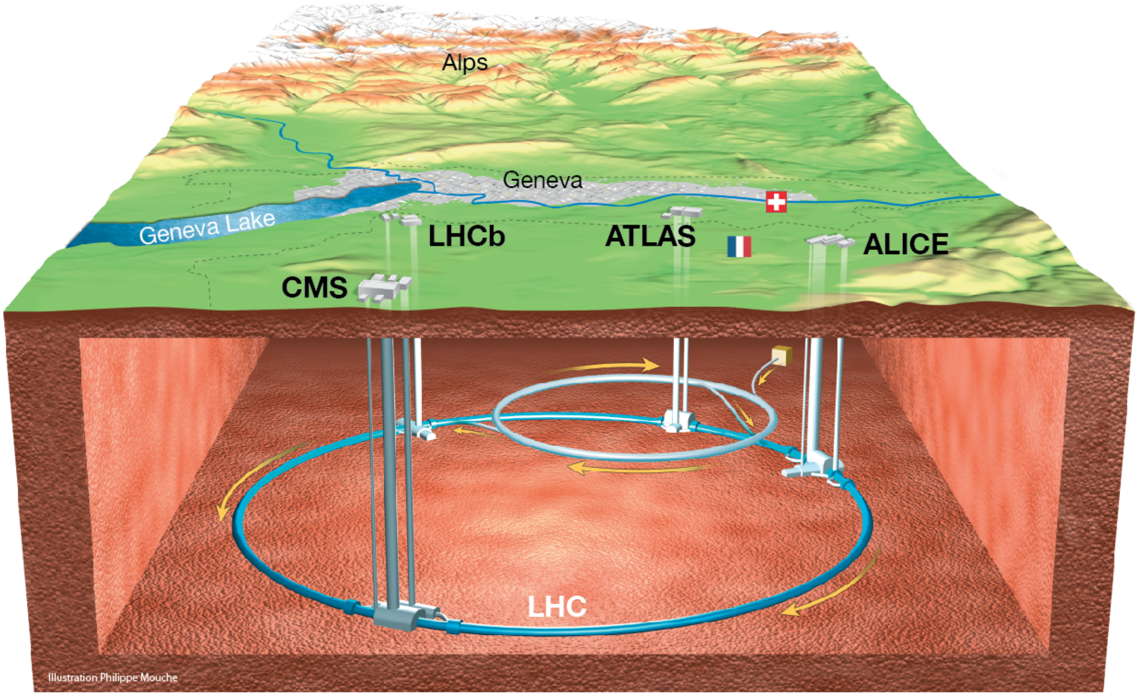


Figure 5: A schematic view of the LHC ring with the four main experiments and the Geneva region.

The number of events N which involve a specific transition $pp \to X$ of protons into a particle X, for instance a Z boson or a Higgs, in a given run with $\mathcal{L}_{int}$ can be calculated via

$$N = \sigma_{pp\to X} \cdot \mathcal{L}_{\text{int}} \,, \tag{3}$$

where $\sigma_{pp\to X}$ is the production cross-section of the proton-proton initial state into the final state X.

3.2 The ATLAS Experiment

The ATLAS experiment is an approximately cylindrically shaped huge general-purpose detector with a length of 44 m, a width of 25 m and a mass of 7000 t. The experiment consists of three main sub-detectors in an onion-like arrangement around the beam pipe and a magnetic system. The latter bends trajectories of electrically charged particles and allows for the determination of their momenta and charges. The sub-detectors are the so-called inner detector for tracking (closest to the beam pipe), a calorimeter system for energy measurements and shower shape recognition (surrounding the tracker), and a muon spectrometer (farthest distance to the beam pipe). A schematic view can be seen in Figure 6. The sub-detectors are described in more detail in the following.

The Inner Detector

The inner detector's (ID) main objective is the reconstruction of tracks of charged particles. It yields important information necessary for particle identification and p_T measurements. It also plays a

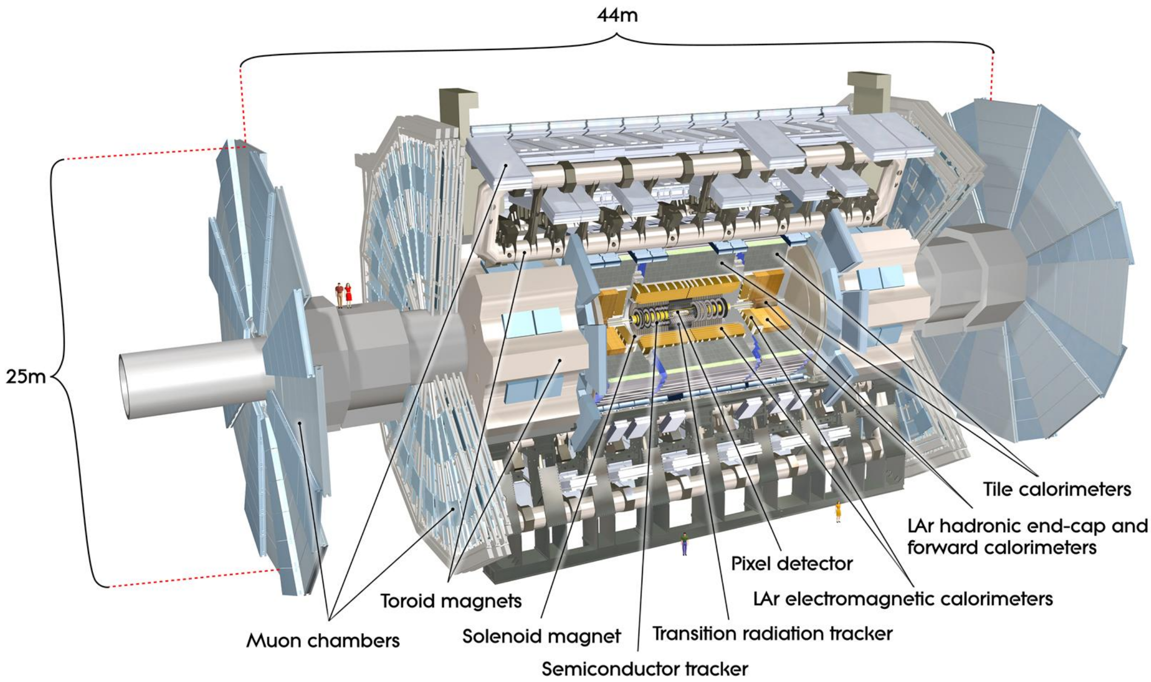


Figure 6: A schematic view of the ATLAS detector.

crucial role in the reconstruction of primary vertices and thus helps in assigning particles to their vertex of origin. It measures the energy loss of charged particles traversing a medium via ionization as described by the Bethe-Bloch equation. Various tracking systems measure the products of this ionization, i.e. electrons or electron-hole pairs, or make use of transition radiation. The inner detector as a whole can reliably reconstruct tracks of charged particles with a $p_T > 400$ MeV. It covers a forward angle of $|\eta| < 2.5$.[6]

The ID consists of three tracking detectors themselves arranged in an onion-like structure. From innermost to outermost these are the pixel detector (PD), the semi-conductor tracker (SCT), and the transition radiation tracker (TRT). The radial coverages of each are 3-15 cm (PD), 30-56 cm (SCT), and 56-107 cm (TRT).

The Calorimeters

The ATLAS calorimeter system consists of two systems, the electromagnetic (EM) and the hadronic (HAD) calorimeter. They are built to fully contain the showers of electromagnetically and strongly interacting particles, allowing for precise energy measurements. In addition, the fine segmentation of the calorimeters into *calo cells* allows for the reconstruction of shower shapes, which aids the discrimination between various kinds of particles and improves particle identification. ATLAS uses *sampling calorimeters* meaning that the calorimeters are made of two kinds of materials: an active material in which the energy measurements take place and a dense absorbing material which fuels

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

particle showering and leads to containment of the shower. The whole calorimeter system covers $|\eta| < 4.9$.

The EM calorimeter uses liquid argon (LAr) technology for the active material and lead absorbers. In the active material ionization takes place and copper-etched electrodes extract the signals. The HAD calorimeter makes mostly use of scintillating tiles for the active material and steel for the absorber. The active material measures the nuclear interactions that take place during hadronic showers.

The Muon Spectrometer

As muons are relatively long-lived particles they are able to traverse and escape the detector. Being charged they leave products of ionization. Hence the muon spectrometer (MS) is an additional tracking system enveloping the other systems to allow for the extended reconstructions of muon tracks. Several types of so-called muon chambers exist which measure spatial and temporal components of muon tracks using different techniques and various ionization gases. Muon tracks can be reconstructed within the MS's coverage, which is $|\eta| < 2.7$.

Trigger and Data Acquisition

ToDo: L1Calo, HLT, Grid

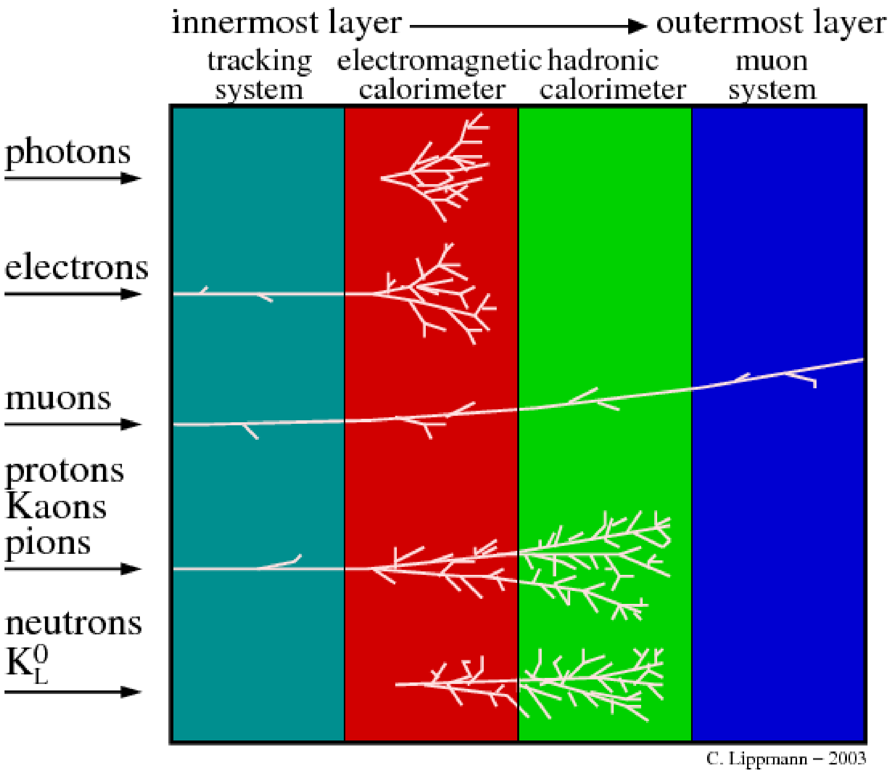


Figure 7: A schematic view of the traces various particles leave in different parts of a multi-purpose detector [7]. Each particle type has its own signature in the detector.

4 Particle Identification

Particles are reconstructed and identified based on the signatures and energy deposits they leave in the various sub-detectors of ATLAS. Three major steps are involved in particle detection: reconstruction, calibration, and identification. Reconstruction algorithms yield a sample of candidates for particles types, for instance electron candidates, based on typical signatures of these particles. Calibration algorithms correct the measured quantities to their appropriate scales depending on the type of particle. The sample of particle candidates after reconstruction and calibration typically is very impure. For instance, the sample of calibrated electron candidates also contains many photons and hadronic jets. Hence the last step, identification, takes care of curating the sample by dismissing particle candidates which likely have a different origin. For this often multivariate techniques such as neural networks or boosted decision trees are used.

Broadly speaking, two kinds of information play the central roles for particle detection: tracking information from the ID and MS and calorimeter clusters. Before briefly discussing the signatures of particles relevant to this lab course, a short overview on tracks and clusters is provided.

Inner Detector tracks are created by charged particles which ionize the active material of sensor modules. For track reconstruction two kinds of algorithms exist: an inside-out and an outside-in algorithm. The inside-out algorithm aims at tracks from primary particles, defined as final state particles with a mean-lifetime of $t_p > 3 \cdot 10^{-11}$ s. Its starting point are three-hit seeds in the SCT and pixel detector. After resolving potential track ambiguities it extends the track candidate outwards into the TRT. The outside-in algorithm's focus is finding tracks of secondary particles, which are produced in interactions of primary particles and typically have an origin different from the primary vertex of the event. As the name suggests, the algorithms starts with the outermost sub-detector of the ID and moves inwards. Found tracks have an associated *quality*. The quality of a track depends on how many total hits in all parts of the ID it has (the more the better) and on how well a χ^2 fit can be performed. The higher (tighter) the quality, the more certain the track has resulted from ionization of a charged particle that originated in a relevant vertex.

Track reconstruction in the **Muon Spectrometer** follows a similar general idea but a different algorithm is used. The algorithm starts with muon track candidates in the middle layer of the MS. Additional segments from other chambers inwards or outwards are added based on hit multiplicity, angles, positions and fit quality.

Calorimeter clusters are built from individual calorimeter cells and are indicative of energy losses of particles in the active medium. There are two main algorithms ATLAS uses for clustering cells: the sliding window algorithm and the topological algorithm. Electron and photon candidates are built from the sliding window algorithm. Hadronic jets are built from topological clustering. Since electrons play a central role for this lab course, the former algorithm will be briefly described in the following. The sliding window algorithm segments the calorimeter into 200 x 256 calorimeter towers in the $\eta - \phi$ plane. A *window* of fixed size, for instance 3 x 5 calorimeter towers, is slid across the entire plane. For each possible position of the window the energies of the constituent towers are summed. If a local energy maximum is found and if it is above a certain noise threshold, a so-called seed for a cluster is found. The seed is turned into a final cluster by iterating over all calorimeter layers and by including additional cells within layer-dependent windows. The position of the cluster is defined as the energy-weighted barycentre of the cluster.

4.1 Electrons and Photons

Electron reconstruction starts with calorimeter clusters. The window size for the sliding window algorithm, see above, is 3 x 5 towers in units of 0.025 x 0.025 in $\eta - \phi$. This cluster reconstruction yields a high efficiency for electrons: 95% (99%) for electrons with p_T = 7 GeV (15 GeV). Since electrons are charged particles, they also leave tracks in the ID. Tracks that might belong to the cluster are fit with an electron hypothesis and if the extrapolated track leads to the cluster the two are matched. In the case of several valid tracks, the ambiguity is solved using information such as the track-cluster distance or the number of hits of the tracks in the pixel detector. The energy measurement of the electron is based on the cluster information whereas the position measurement is based on the track information. The reliance on the ID for electron reconstruction means that in can only be performed in the region that is covered by the ID, i.e. within $|\eta| < 2.47$.

Electron candidates are calibrated to an appropriate electromagnetic energy scale. For this, a mixture of simulated data, cf. section 5.3, and measured reference events of the process $pp \to Z \to e^+e^-$ are used.

The sample of electron candidates is still impure and an identification algorithm is required to resolve this. In order to discriminate real electrons from photons or hadronic jets which just look a bit like electrons, a likelihood (LH) method is used. The LH is a multivariate analysis technique that scores electron candidates according to certain metrics. Variables that are considered for this are for instance shower shape information using the calorimeter clusters and track-cluster matching. Based on the LH score three working points are defined: *loose, medium,* and *tight*. An electron candidate labelled *tight* has the highest probability of in-fact corresponding to an electron. An electron candidate labelled *loose* on the other hand still has a fair chance of in reality corresponding to a photon or a hadronic jet. In addition to the identification algorithm isolation requirements can be applied. An isolation requirement enforces that there is little to no other significant activity in close vicinity around the electron candidate. This requirement helps discriminating *prompt* electrons, such as in

$pp \rightarrow Z \rightarrow e^+e^-$, against electrons originating for from instance photon conversions or the decays of heavy quarks.

Photon reconstruction, calibration, and identification follows a similar strategy as for electrons. However, with photons being neutral they do not leave tracks in the ID. Hence clusters without associated tracks are classified as photon candidates. In about 40% of the cases however, a photon conversion takes place and the photon converts into two electrons which hit the calorimeter in close vicinity. In this case two tracks can be found in the ID which often lack hits in the first layer.

4.2 Muons

Muon reconstruction relies on tracks in the ID and the MS. Calorimeter information is usually not used as muons are minimally ionizing. Hence they do not leave large deposits in the calorimeters. For forming muon candidates tracks from the ID and MS are fit using a muon hypothesis. If the fit quality is good, a muon candidate is stored. If it is bad, a different combination of ID and MS tracks is fit. Calibration, identification, and isolation follow similar trains of thought as above.

5 Practical Information and Tools

5.1 Computing and Python Comments

We are using the programming language Python in this lab course. Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.

As already pointed out in section 1, basic programming skills in Python have to be taken for granted and cannot be taught within the limited time available in this lab course. It is urgently required that you refresh or learn basics commands and coding elements in Python before starting this lab course. Links to two Python tutorials can be found in subsection 1.2.

The personal computers in this lab course are running Ubuntu 18.04 LTS as operating system. Ubuntu is a free and open-source Linux distribution based on Debian. As most modern operating systems, Ubuntu utilizes a graphical desktop environment, GNOME, which is simple and easy to use such that sophisticated knowledge in Linux is not required.

In order to run a Python script, open a terminal by right-clicking onto the desktop and choosing ”Open Terminal” from the drop-down menu. In the terminal window type:

```
$ python name_of_script.py
```

where name_of_script is the actual file name without the .py extension.

5.2 The Data Analysis Framework ROOT

Originally ROOT started as a plotting program, which nowadays has matured into a multi-purpose data analysis framework which is used by most particle physicists. It can be used in different ways: From the command line with the build-in C++ interpreter Cling, with interpreted C++ macros or C++ compiled code, or with Python scripts. For smaller data samples, the latter becomes more and more popular. Also in this lab course Python is used for both the data analysis as well as post-processing the results.

ToDo: Maybe shortly introduce ROOT command line, TBrowser, TTree, histograms, TTStack, fitting, adding legends.

5.3 Monte Carlo Simulation

ToDo: Explain better and more details.

Monte Carlo simulations play a central role in particle physics analysis, and are used for various purposes:

- Detector design and optimization (complicated, huge and very expensive detectors).
- Simulation of particle interactions with the detector material.
- Physics analysis (e.g. signatures of new physics models, background estimation, efficiency determinations).

The Monte Carlo simulation chain consists of the event generation, detector simulation and reconstruction. Events are generated according to a chosen model (i.e. calculation of four-vectors of semi-stable particles). In the simulation step, the full detector response is evaluated, including analogue signals, digitization and trigger (time expensive). The output format of the simulation is identical to that of the real data and the same software is used to reconstruct particle candidates (electrons, photons, muons, etc.) in both cases.

More details can be found in following reviews of the Particle Data Group:

`http://pdg.lbl.gov/2018/reviews/rpp2018-rev-mc-event-gen.pdf` ,
`http://pdg.lbl.gov/2018/reviews/rpp2018-rev-monte-carlo-techniques.pdf` .

Monte Carlo Particle Numbering Scheme

The Monte Carlo particle numbering scheme, also referred to as Particle Data Group Identifiers, PDGID, is intended to facilitate interfacing between event generators, detector simulators, and analysis packages used in particle physics. It is maintained by the Particle Data Group in cooperation with Monte Carlo authors. The general form is a 7-digit number which encodes information about the particle's spin, flavour content, and internal quantum numbers. Particles are given positive numbers, antiparticles negative numbers.

Table (still to be included) lists the PDGID for the six quarks and leptons of the Standard Model. More details can be found in this PDG review.

5.4 Pseudorapidity

ToDo: Expand to hadron collider event kinematics in general.

The pseudorapidity η (eta) describes the polar angle of a particle relative to the beam axis. It is defined as

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

In hadron collider physics, particle production is normally constant as a function of pseudorapidity.

Figure 8 shows the cross-section of the ATLAS Liquid Argon Calorimeter system through the beam axis. The ATLAS detector design is symmetrical in pseudorapidity. The detector region $|\eta| > 3$ is particularly challenging since particle densities and energies are at their highest.

5.5 The Tag-and-Probe Method

The precise knowledge of the particle detection efficiencies plays a crucial role when e.g. measuring cross sections or searching for new physics signatures (limit setting). Measuring the efficiencies requires a clean and unbiased sample of the particles under investigation. For electrons, the method of choice is the tag-and-probe method, which makes use of the characteristic signatures of $Z \to ee$ decays.

In this method, strict selection criteria are applied on one of the two decay electrons, called tag, and the second electron, the probe, is used for the efficiency measurements. Standard event selection criteria are applied to further reject background. The tag-and-probe pairs must also pass requirements on their reconstructed invariant mass.

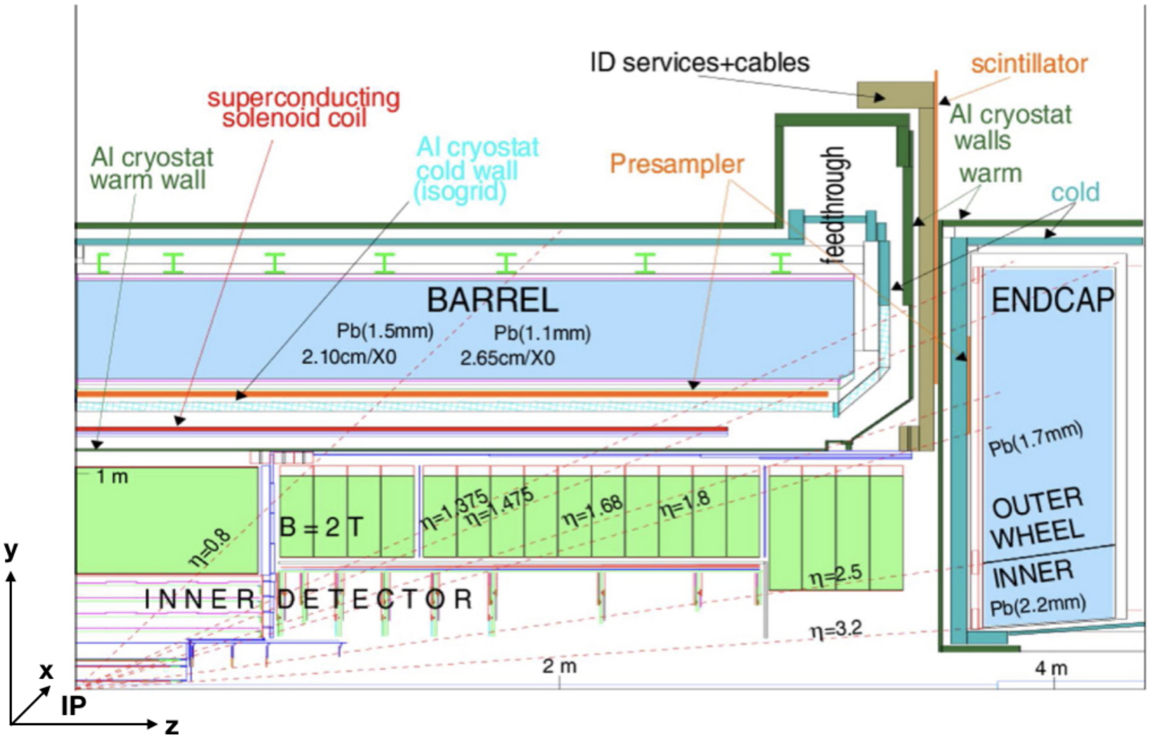


Figure 8: Longitudinal cut-out of the ATLAS Liquid Argon Calorimeter system in the $y - z$ plane [8]. Shown is one quarter of the cross section through the calorimeter.

The probe samples might be contaminated by background objects (for example, hadrons misidentified as electrons, electrons from semileptonic heavy flavour decays or from photon conversions). This contamination can in principle be determined using either background template shapes or combined fits. However, the background contribution is reasonably small such that it can be neglected within the scope of this lab course.

The tag-and-probe selection makes sure, that the probe electron is most likely a real electron and thus can be used to determine detection efficiencies. For this the number of electrons is independently estimated at the probe level and at the level where the probe electron candidate satisfies in addition the tested criteria. The efficiency ϵ is then defined as the fraction of probe electrons satisfying the tested criteria.

Usually the total efficiency to detect an electron in ATLAS is divided into different components, namely trigger, reconstruction and identification efficiencies, as well as the efficiency to satisfy additional analysis criteria, like isolation. The full efficiency ϵ_{total} for a single electron can be written as:

$$\epsilon_{total} = \epsilon_{reconstruction} \times \epsilon_{identification} \times \epsilon_{trigger} \times \epsilon_{additional} \quad .$$

5.6 ATLAS Open Data

ATLAS released an official dataset, open to the public for educational use only [3], [4]. The dataset consists of real data recorded with the ATLAS detector in 2012, with an integrated luminosity of (1.0007 ± 0.0019) fb^{-1} and a centre of mass energy of 8 TeV. Matching simulated Monte Carlo data for various physics processes have been made public too.

Both real and simulated data are subjected to a stringent preselection, both on event as well as on physics object level. In particular, a primary vertex is required, with more than four tracks pointing to it, and event cleaning as well as background rejection cuts are applied. A single lepton trigger must be satisfied, and the event must contain at least one reconstructed lepton with $p_T > 25$ GeV.

On object level, medium++ identification quality is required for all electrons (c.f. section 4.1), + and tight quality for all muons. In addition, electrons and muons must be well measured in the central part of the ATLAS detector with a minimum p_T of 5 GeV.

The released data is provided in a simplified format to reduce the complexity of a full scale analysis. The resulting format is a ROOT TTree with 45 branches as detailed in table 3.

In addition to the datasets themselves, ATLAS released analysis tools as well as actual example analyses. Note however, that these are not needed for this lab course in which we will develop our own tools and analysis. Keeping this in mind, further general information can be found on the public website of the ATLAS Open Data project where also a Get Started GitBook is provided.

5.7 Uncertainties of the Measurement

In natural science, each measurement is performed with a certain precision, and the corresponding uncertainty or error must be indicated with the result. This is also – or in particular – true for experimental particle physics analyses in which it is not uncommon that a significant fraction of the effort is dedicated to the proper evaluation of the experimental errors.

One has to distinguish between statistical and systematic errors. Statistical errors are due to statistical uncertainties and arise from stochastic fluctuations. Important examples for measurements in experimental particle physics are finite statistics (Poisson distribution) and measurement resolutions.

Another class of errors, usually much harder to detect, are systematic errors which arise from uncertainties in the apparatus or model. Examples are calibration uncertainties, detector acceptance and poorly-known theoretical parameters. In general, systematic errors are the uncertainties in systematic correction factors.

”The treatment of systematic errors is one of the most demanding parts of experimental physics. It requires ingenuity, experience, cunning, and a fair degree of paranoia. These are well provided in particle physics. It also requires honesty and bravery, which is scarcer” (Roger Barlow).

6 Instructions

This chapter provides the actual instructions for the daily work during this lab course, which you should follow in the given order. After some general information about the files needed for the data analysis, we will start with a few exercises using ROOT interactively. While this is useful to get familiar with the structure of the underlying data – or more general for something "fast and simple" – the ROOT browser and command line are not suitable for systematically processing larger data sets. Therefore we'll be moving on to using Python scripts rather quickly, in order to automate the work. When this is accomplished, we can dive deeper and deeper into the actual particle physics analysis.

6.1 Information about the Files in this Lab Course

All files which are needed for this lab course are stored in the directory `/fp/data`. These are the ATLAS data and simulated MC files in form of ROOT trees, as well as example Python code to analyse them. As in the real experiment, the data files are divided into so-called streams, depending on which trigger selected the event. For the MC files, the subdivision is based on the underlying physics process which was chosen for the event generation.

Hint: The actual directory holding the files resides on a server and will be mounted automatically with AutoFS when accessed. Hence it is likely that the directory is not visible when you start your daily work. In this case, open a terminal (right-click with the mouse on the desktop and choose "Open Terminal" from the drop-down menu) and type `ls /fp/data`. This will cause the directory to be mounted, and also list its content.

The ROOT files for the ATLAS data trees are stored in the directory `/fp/data/Data`. The `DataEgamma.root` file contains events selected by electron or photon triggers, while the `DataMuons.root` file contains events selected by muon triggers.

In the `/fp/data/MC` directory, various ROOT files for simulated MC processes are contained. Only three files are needed for this lab course: `mc_147770.Zee.root`, `mc_147771.Zmumu.root`, and `mc_147772.Ztautau.root`. As indicated by the names, these ROOT files contain events with Z bosons decaying into electron pairs, muon pairs, and tau pairs, respectively.

Finally, the folder `/fp/data/Code` contains example Python code for the analysis event loop, plotting as well as fitting the final results. These files should be copied into the user's home area first day of the lab course, and can be modified there. Contrary, the ROOT files for the ATLAS data and MC simulations should remain in their original locations. Instead, the full path to the location of the files needs to be given to the Python analysis scripts when looping over the events.

6.2 Getting to Know the Basics

In this subsection you will be using ROOT interactively in order to make yourself familiar with the structure of the data files in the form of ROOT trees; and of course to get a first impression of ROOT itself. For this you will be using ROOT's `TBrowser` which allows to browse ROOT objects in graphical way.

ROOT can be started from the command line, issuing the command `root` in a terminal window. The result is the output of the ROOT banner for a few seconds and the prompt of Cling which is an interactive C++ interpreter. User input is typed just after the ROOT prompt "root[n]" where n is counting the accepted commands. In order to quit this interactive ROOT session type ".q".

1. Get familiar with the open data files provided for this course. In a terminal, open the `DataEgamma.root` file with ROOT and use the ROOT browser to look at its content:

   ```
   $ root /fp/data/Data/DataEgamma.root
   root [] TBrowser t
   ```

   A short comment for each variable in the ROOT file can be found in Appendix 3 or online at https://cheatham1.gitbooks.io/openatlasdatatools/variable_names.html .

2. Plot the `vxp_z` variable. Make yourself clear what is shown, and why the distributions looks as it does. What does the statistics box in the upper right corner tell you? What are the units of the axis? Label them correctly by right-clicking with the mouse at the proper elements. Save and print the plot using the TBrowser: `File` $\rightarrow$ `Save` (or `Save As`). png or pdf are the file types of choice.

3. What are LHC bunches, and how are they related to the `vxp_z` plot in question 2? How many peaks do you see in the `vxp_z` distribution? Explain possible reasons. What type of plot could help you answering this question? (We might come back to this later).

4. Plot the `lep_n` variable. What is shown? Unfortunately this distribution has too coarse binning out of the box. We need to use ROOT magic in the command line to show $n = 1$ binning. Type

   ```
   root [] mini -> Draw("lep_n >> htemp(10,0,10)")
   ```

   Then press `enter` and click with the middle mouse button on the canvas to refresh it. (To do: Find better procedure here.)

5. How many leptons do you expect in the final state of a Z decay? Why do you see events here with more and less than two leptons? What is the dominant process for the 1-lepton final state? Draw Feynman diagrams for 1-lepton, 2-lepton and 3-lepton final states.

6. Plot the `lep_pt`, `lep_eta` and `lep_phi` variables. What do you see? What does the statistics box in the upper right corner tell you now? Why are there more entries than in question 2? How many entries do you expect and why? What are the general features of the distributions? Why are there gaps in the η distribution? *Hint: Consult figure 8 and have a closer look on the tracker coverage as well as dead material distributions.* Where does the steep rise around 25 GeV in the p_T spectrum come from? Explain the concept of a trigger.

6.3 Automating Things

While interactive ROOT sessions are useful for quick checks and small operations, they are not suitable for systematic data analysis. Not only will the usually large size of the data sets become problematic, but more importantly, in data analysis we want to apply specific operations on certain observables (in the form of variables) for each event. Therefore a loop over all events in the ROOT tree is needed, allowing to program the desired operations at event level. For this we are using Python scripts in combination with PyROOT which is an extension module that allows the user to interact with any ROOT class from the Python interpreter.

1. Before you begin, create a work directory in your home area and copy all Python examples files from `/fp/data/Code` into that directory.

2. Look at the provided Python script `eventloop.py` and make yourself familiar with its content. What does it do? Run the script on the `DataEgamma.root` file and limit the number of events to a reasonable amount. Practise the handling of primitive versus complex variable types such as

arrays and lists by printing the content of various variables in the event loop, e.g. `vxp_z`, `lep_n`, `lep_pt`, `lep_eta`.

3. Compute the invariant mass of the two leading leptons in the event and plot it. Pay attention to the sign of M_{ll}^2 and restrict to a physical result.
   Hint: The leptons can be treated as massless. Recall the trigonometric relations between p_x, p_y, p_z and p_T. In particular show that $p_z = p_T \cdot \sinh(\eta)$ by first evaluating its dependence on the polar angle θ, and then rewriting as function of the pseudorapidity η.

4. Now use ROOT's TLorentzVector class to calculate the invariant mass and compare with your results by overlaying the two histograms. Make sure all plots have a legend and proper axis labels. *Hint: Use the histogram draw option "same" and change the plotting style of the curves either by using the ROOT editor, or directly in the Python script.*

5. How do you expect the distribution to look like for the decay of a Z boson – and why might it not look like your expectation when running over the `DataEgamma.root` file? How many peaks can you identify, and what do they belong to? *Hint: Try single and/or double logarithmic axis scales.* Now run your analysis over the Monte Carlo file `mc_147770.Zee.root` and compare the results for the invariant mass distributions.

6.4 Selecting Events

How can we improve the invariant mass distribution and extract a better Z boson peak? – By applying a proper event selection! In general one has to distinguish between cuts on event level, and selection criteria applied on particle level. Usually both are needed.

1. It is interesting to keep track of how many events are discarded by criteria. This is called a "cut flow" histogram and very useful to have. Hence create a histogram with 20 bins that will show the total number of events remaining for each consecutive event selection criterion that be introduced in the following.

   ROOT allows to label the axis for each bin individually. So let's label each bin with the selection criterion it belongs to. Figure 9 shows as example the cut flow histogram for the analysis of the `DataEgamma.root` data file. Which are the two most effective cuts?

2. In the following, a brief description of the applied selection criteria is given. Implement them in your analysis code to select a clean sample of Z boson candidates. What do you think is the reason that the cuts on event level play only a minor role here? Validate the improvements of the cuts by monitoring the changes in the invariant mass distribution.

Weights: This is not actually an event selection cut, but rather to keep track of the impact of event weights in the Monte Carlo files. While in real data a single event is unique and counts for one, event weights might have been applied in the MC generation (c.f. section 5.3).

Trigger: In order to preselect events in which a Z boson decayed into an electron or muon pair, the corresponding trigger must be required to have fired the event. On the level also the luminosity accounting takes place.

GRL: The Good Run List. After the data acquisition of a so-called ATLAS run is completed, the data are investigated by the data quality group, and harsh criteria are applied in order make sure that only meaningful data enter the physics analysis. This way whole runs, or in most cases parts of runs, in which for example important parts of the ATLAS detector were malfunctioning, will be flagged and can be excluded on analysis level (the subunit of an ATLAS run is called luminosity block and lasts approximately one minute).

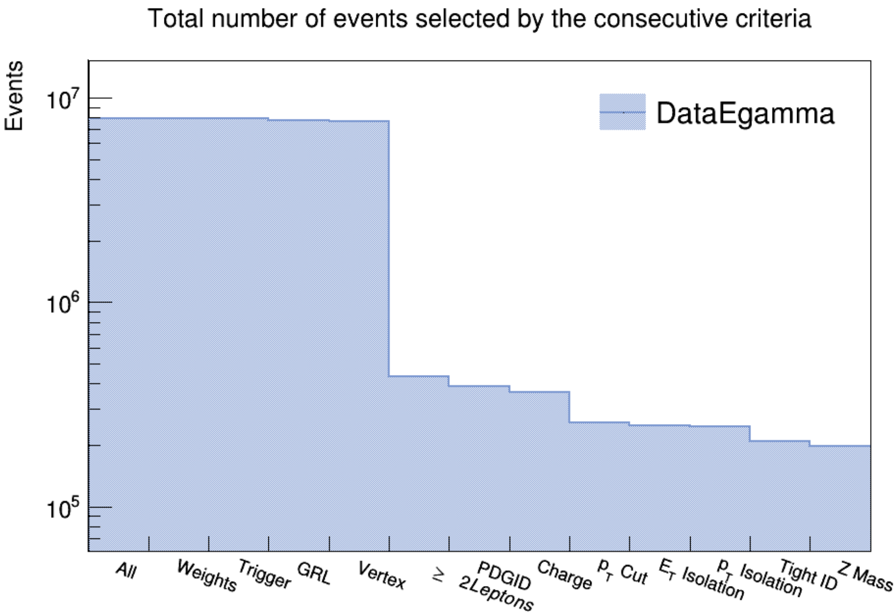


Figure 9: The cut flow for the selection of Z boson candidates for the `DataEgamma.root` file.

Vertex: In order to separate background from physics events, a well measured primary vertex is required.

2 Leptons: Given that the Z boson decays into lepton pairs, at least two leptons are required.

PDGID: Particle Data Group Identifiers. As described in section 5.3, each particle in Monte Carlo event generation is assigned to a unique identifier. Therefore also after the simulation of detector effects, the type of the particle is always known on MC level. For real data however, the identification of particles is more complicated and utilises the signatures as measured in the various detector components (c.f. section 4). To facilitate the particle identification, the Open Data project provides a generalised PDGID, `lep_type`, which is valid both for MC as well as data. It can be used to uniquely identify the leptons as electrons and muons. *Hint: Pay attention of the sign difference for particles and antiparticles*.

Charge: The Z bosons as electrically neutral particle has to decay into a lepton pair with opposite charges. This can be taken into account in the event selection.

p_T Cut: Due to the high mass of the Z boson, the two decay leptons in the final state are expected to carry sizeable transverse momenta, with distributions peaking around half the Z mass. This can be utilised effectively in the event selection by requiring for the leptons p_T values of at least 25 GeV.[7]

As can been seen by monitoring the changes in the invariant mass distribution, the largest part of the background should be removed after having applied the lepton p_T cut (recall that the Open Data preselection applies rather harsh quality criteria on the leptons candidates, cf. section 5.6). However, as statistics is not of concern, it is still worth to increase the purity of the event sample by further tightening the lepton identification criteria. This can be achieved by introducing isolation cuts and increasing the lepton quality, as discussed in section 4.1.

[7] Also the thresholds for ATLAS' main single lepton triggers are placed such that they become fully efficient only in this p_T regime, making such transverse momentum cut advisable. Turning the argument around, the reason for this particular trigger setup is the aim to still be able to efficiently record the decays of the electroweak bosons.

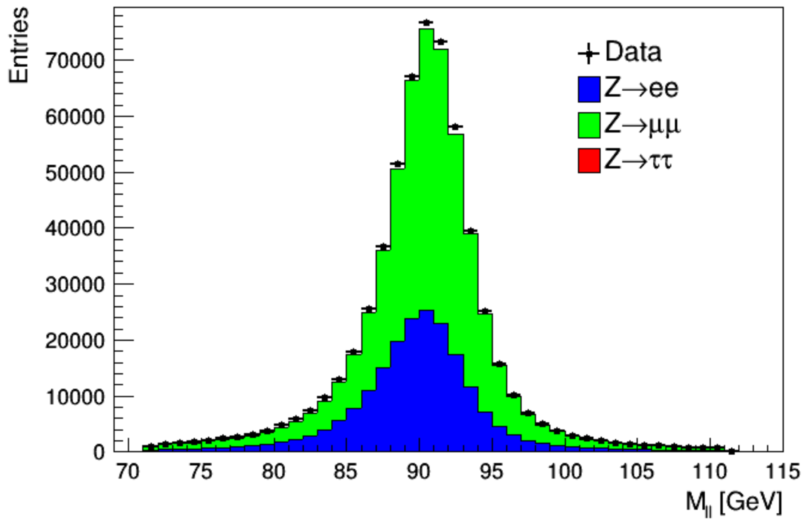


Figure 10: Invariant mass of the Z candidate.

Isolation: The isolation in the $\eta - \phi$ space can be used to improve the selection of leptons originating from Z boson decays. Calculate the *relative* E_T and p_T isolation of the leptons and investigate their distributions. *Hint: The amount of transverse energy (momentum) deposited in a cone around the lepton candidate should be compared to the transverse energy (momentum) carried by the lepton candidate itself.* What are reasonable cut values? Discuss your findings with the tutor.

Tight ID: The lepton reconstruction algorithms on detector level provide the possibility to ask for looser or harsher identification criteria. These are bit-encoded in the `lep_flag` variable. Bit number 9 (when counting from zero) indicates a "tight" selection and should be required in the analysis. *Hint: Recall bitwise operations and bit-testing.*

Z Mass: Finally, the known mass of the Z boson can be used in the event selection, by restricting the invariant mass of the di-leptons to an allowed range.

3. Test the implementation of the event selection criteria continuously by processing a reasonably small sample of `DataEgamma.root` events. After completion, discuss your low-statistics invariant mass histogram with the tutor. If satisfactory, process the complete statistics of both data ROOT files and all three Monte Carlo files. This will take several hours and should be done over night.

6.5 Comparing Data and MC Simulation

As described in section 5.3, Monte Carlo simulations play an important role in particle physics analysis. Amongst other they are used to prove that all detector effects are understood, and that all physics processes contributing to the final state under study are correctly modelled and included.

A simple but descriptive way of demonstrating the level of agreement between data and simulation is the comparison of the measured and simulated distributions in "stacked plots". Figure 10 shows such a plot for the invariant mass of the Z candidate as measured in this lab course with the full statistics.

1. In figure 10 the data histogram corresponds to the sum of the Z boson candidates decaying into electron and muon pairs. To merge ROOT files containing histograms the utility `hadd` can be used. At the shell command line, simply type `hadd` to get online help.

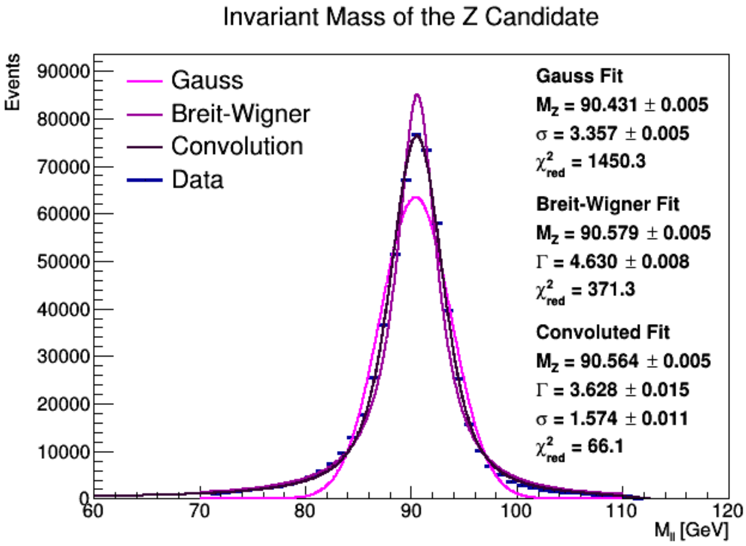


Figure 11: Fitting the invariant mass of the Z candidate.

```
$ hadd result.root file1.root file2.root ...  filen.root
```

Use the `hadd` command to merge your histogram result files of the `Egamma` and `Muon` analyses into a combined `Data` results histogram file.

2. Compare your final invariant mass distribution of the merged data file to the Monte Carlo simulation. For this use and complete the provided script `plot.py` to make the same plot as shown in figure 10. With the THStack class, ROOT provides a simple way for plotting histograms stacked on top of each other. Note that proper scaling factors have to be applied to the Monte Carlo distributions to take into account the target data luminosity of $1\ \mathrm{fb}^{-1}$. Those are provided in the `plot.py` script.

   ToDo: Better explanation of the scaling factors, maybe include a table.

6.6 Fitting the Z Mass

The selected event sample can be used to study properties of the Z boson. Obviously, the invariant mass distribution provides a direct way for the determination of the Z boson mass and its decay width. For best results, the mass and width are extracted from a fit to the invariant mass spectrum, using the convolution of a Gauss and Breit-Wigner distribution, as shown in figure 11.

1. Why is the convolution of a Gauss and Breit-Wigner line shape expected to yield the best fit results? Which effects are taken into account by each of the two contributions?

2. Use and complete the provided script `fit.py` to extract the mass and width of the Z boson from the measured data sample. Compare your results with the PDG values.

6.7 Determining Efficiencies

The precise knowledge of efficiencies plays an important role in many areas of physics data analysis. Hereby not necessarily the question stands in the foreground, if or when a selection or measurement is fully efficient, but also how precise the efficiency is known. If errors are small, inefficiencies can be corrected for in the measurement. In this lab course, we determine the efficiency for the ”tight” electron identification criteria.

1. Use the tag-and-probe method as described in section 5.5 to determine the detection efficiency for ”tight” electrons in data as function of p_T. Fill a numerator and denominator histogram separately in the event loop. The final efficiency can be derived by dividing both histograms. *Hint: ROOT provides a TH1::Divide() method.*

2. Compare the ”tight” electron detection efficiencies in data and Monte Carlo simulation. Is the detector simulation able to describe the data correctly?

3. What type of error determination need to be applied correctly account for the statistical uncertainties in the efficiency determination?

6.8 Systematic

So far there is no proper treatment of statistical and especially systematic errors implemented yet in this lab course – lucky you! However, considering the general comments in section 5.7 and recalling the individual analysis steps needed for the derivation of the Z boson mass (beginning with the calculation on event level up to the fitting of the invariant di-lepton mass distribution) you should be able to identify a series of error candidates. Discuss with your tutor.

Appendix

Branch Name	Type	Description
runNumber	int	Run identifier
eventNumber	int	Event identifier
channelNumber	int	Data sample ID eg WW sample 105985
mcWeight	float	Weight of a simulated event
pvxp_n	int	Number of primary vertices
vxp_z	float	z-position of the primary vertex
trigE	bool	Boolean whether a standard trigger has fired in the egamma stream
trigM	bool	Boolean whether a standard trigger has fired in the muon stream
passGRL	bool	Signifies whether event passes the Good Run List may be put in isGoodEvent
hasGoodVertex	bool	Signifies whether the event has at least one good vertex where Ntracks > 4
lep_n	int	Number of preselected leptons
lep_truthMatched	vector<bool>	Boolean indicating whether the lepton is matched to a simulated lepton
lep_trigMatched	vector<bool>	Boolean signifying whether the lepton is the one triggering the event
lep_pt	vector<float>	Transverse momentum of the lepton
lep_eta	vector<float>	Pseudorapidity of the lepton
lep_phi	vector<float>	Azimuthal angle of the lepton
lep_E	vector<float>	Energy of the lepton
lep_z0	vector<float>	z-coordinate of the track associated to the lepton wrt. the primary vertex
lep_charge	vector<float>	Charge of the lepton
lep_type	vector<int>	Number signifying the lepton type (e, mu, tau) of the lepton
lep_flag	vector<int>	Bitmask implementing object cuts
lep_ptcone30	vector<float>	Scalar sum of track pTs in a cone of R=0.3 around lepton, not including lepton pT
lep_etcone20	vector<float>	Scalar sum of track ETs in a cone of R=0.2 around lepton, not including lepton ET itself
lep_trackd0pvunbiased	vector<float>	d0 of the track associated to the lepton at the point of closest approach (p.o.a.)
lep_tracksigd0pvunbiased	vector<float>	d0 signifcance of the track associated to the lepton at the p.o.a.
met_et	float	Transverse energy of the missing momentum vector
met_phi	float	Azimuthal angle of the missing momentum vector
jet_n	int	Number of selected jets
alljet_n	int	Total number of jets in event
jet_pt	vector<float>	Transverse momentum of the jet
jet_eta	vector<float>	Pseudorapidity of the jet
jet_phi	vector<float>	Azimuthal angle of the jet
jet_E	vector<float>	Energy of the jet
jet_m	vector<float>	Invariant mass of the jet
jet_jvf	vector<float>	Jet vertex fraction of the jet
jet_flag	vector<int>	Bitmask implementing object cuts of the top group
jet_trueflav	vector<int>	Flavor of the simulated jet
jet_truthMatched	vector<int>	Information whether the jet matches a simulated jet
jet_SV0	vector<float>	Weight from algorithm that reconstructs Secondary Vertices associated with a jet
jet_MV1	vector<float>	Weight from algorithm based on Multi-Variate technique
scaleFactor_PILEUP	float	Scalefactor for pileup reweighting. It effectively reweights the profile of average interactions per bunch crossing so that simulated data is the same as measured data.
scaleFactor_ELE	float	Scalefactor for electron efficiency
scaleFactor_MUON	float	Scalefactor for muon efficiency
scaleFactor_BTAG	float	Scalefactor for btagging algorithm. Should only be applied if analysis is specifically using b-tagging
scaleFactor_TRIGGER	float	Scalefactor to account for the different operating efficiencies of the used triggers
scaleFactor_JVFSF	float	Scalefactor for jet vertex fraction
scaleFactor_ZVERTEX	float	Scalefactor to reweight the distribution of the z position of the primary vertex

Table 3: Branches of the trees in the ATLAS Open Data ROOT files.

References

[1] M. Thomson, *Modern Particle Physics*. New York: Cambridge University Press, 2013, ISBN: 9781107034266. [Online]. Available: http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.2.T46::2013.

[2] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, S08003, 2008. DOI: 10.1088/1748-0221/3/08/S08003.

[3] "Review Studies for the ATLAS Open Data Dataset," CERN, Geneva, Tech. Rep. ATL-OREACH-PUB-2016-001, Aug. 2016. [Online]. Available: https://cds.cern.ch/record/2203649.

[4] "Review of ATLAS Open Data 8 TeV datasets, tools and activities," CERN, Geneva, Tech. Rep. ATL-OREACH-PUB-2018-001, Jun. 2018. [Online]. Available: https://cds.cern.ch/record/2624572.

[5] M. Jende *et al.* (). International Masterclasses - Hands On Particle Physics, [Online]. Available: http://atlas.physicsmasterclasses.org/en/index.htm (visited on 12/21/2018).

[6] M. Wessels, "General Search for New Phenomena in ep Scattering at HERA," PhD thesis, RWTH Aachen, Jul. 2004.

[7] C. Lippmann, "Particle identification," *Nucl. Instrum. Meth.*, vol. A666, pp. 148–172, 2012. DOI: 10.1016/j.nima.2011.03.009. arXiv: 1101.3276 [hep-ex].

[8] M. Aharrouche *et al.*, "Study of the response of atlas electromagnetic liquid argon calorimeters to muons," *Nucl. Instrum. Meth.*, vol. A606, pp. 419–431, 2009. DOI: 10.1016/j.nima.2009.05.021.