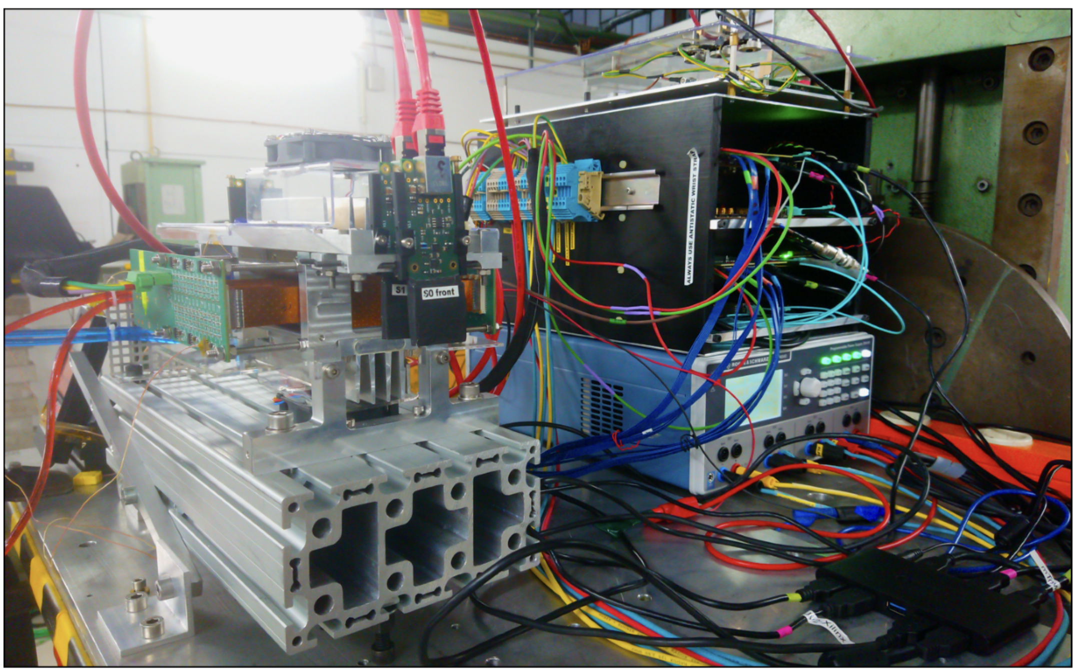
Universiteit Utrecht

Opleiding Natuur- en Sterrenkunde

Event selection to reduce pileup in data from a pixel calorimeter prototype

Bachelor Thesis

Aart van Bochove



Supervisors:

Dr. ing. N. van der Kolk
Subatomic Physics (SAP)

Prof. Dr. T. Peitzmann
Subatomic Physics (SAP)

June 12, 2020

Abstract

To learn more about matter at high density, the installation of a forward calorimeter (FoCal) is proposed as an addition to the ALICE detector at CERN. A small prototype for this calorimeter is build, the mTower. In February 2020, measurements with electron beams of a few GeV have been performed at the DESY test beam facility in Hamburg with this prototype to test its performance. The measurements suffer from pileup, where multiple electrons enter the detector during one event. To assess the performance, measurements of single electrons in the prototype have to be selected. A set of 5 criteria has been developed to make this selection of events without pileup. These 5 criteria reduce the amount of pileup substantially, but there is still an asymmetry present in the final distributions of the number of hits, which suggests that not all of the events with pileup are removed. However, it is not possible to reduce the pileup further with the current available data. The selection criteria are applied to the current data set, resulting in a sufficiently clean sample. The distributions of the number of hits are shown for several electron beam energies. For future test beam measurements, the amount of pileup should be reduced by for example lowering the electron beam rate or by changing the trigger setup to create a new source of data which can be used to identify pileup.

The title page picture shows the mTower calorimeter prototype during the test measurements at the DESY test beam facility in Germany in February 2020.

Contents

Acronyms

ALICE A Large Ion Collider Experiment. i, 2, 4, 5, 7, 15

CGC Color-Glass Condensate. 2, 6

DCR Dark count rate. 18

DGLAP Dokshitzer, Gribov, Lipatov, Altarelli, Parisi. 5, 6

FoCal Forward Calorimeter. i, 2, 7, 15

LHC Large Hadron Collider. 2, 4, 41

MAPS Monolithic Active Pixel Sensor. 16, 18

mTower mini Tower. i, 2, 13, 16–19, 21, 38, 44, 46

PDF Parton Distribution Function. 5–7, 15

QCD quantum chromodynamics. 4, 5

QGP quark-gluon plasma. 2, 4, 7

SiPM Silicon photomultiplier. 17, 18, 38

1 Introduction

In physics, an important research area consists of the study of extreme physical phenomena, for example extremely high or low temperatures, very strong magnetic fields, high velocities, high or low densities or materials which have a very high conductivity. Some of those phenomena meet in high-energy nuclear collisions. In these collisions a *quark-gluon plasma* (QGP) is produced, which we assume existed also in the early stages of the universe, just after the big bang. The QGP can still be found in the dense core of neutron stars. A QGP is a state of matter at extremely high temperatures or densities, in which the constituents of particles, the quarks and gluons, are not confined. Another interesting state of matter theoretically postulated in these collisions is the Color-Glass Condensate (CGC). These high-energy collisions take place in particle accelerators, for example in the Large Hadron Collider (LHC) at CERN.

The LHC is the largest and most powerful particle accelerator in the world. In a circular ring with a circumference of 27 km, two particle beams are accelerated into opposite directions, after which they are brought into a collision with each other. These collisions take place between protons and/or lead-ions at several locations in the ring where detectors are positioned, one of which is ALICE (A Large Ion Collider Experiment). This detector is specialized in heavy ion collisions [1].

A *forward calorimeter* (FoCal) is proposed to be installed for the ALICE experiment, during the Long Shutdown 3 of CERN, scheduled between 2025 and 2027 [2]. It will most likely be positioned at a distance from the interaction point of 7 m. The electromagnetic calorimeter must be able to discriminate decay photons from direct photons at very high energy, which requires a very high granularity [3]. In order to build such a detector, research has to be done to test several aspects of the design. As part of this research, a small prototype is build using the foreseen technology, the mTower. In February 2020, measurements were performed at the DESY test beam facility in Germany with the mTower. Electron beams were fired at the mTower to investigate its response to electrons, the performance of the readout electronics and the quality of the data it produces.

Sometimes multiple electrons entered the detector during one readout cycle of data (also called an *event*). This is called pileup. In order to assess the response to a single electron, such events must be filtered out. Usually this pileup is removed by looking at time between consecutive triggers. During this test beam the time information was not recorded, which is exceptional. For that reason, other sources of information must be investigated to try to remove this pileup. In this thesis selection criteria are investigated with the aim of creating a clean single electron data sample. These selection criteria are based on the location of the energy deposits (they are called *hits*) in the detector.

In chapter 2 the theoretical background is explained, looking at the motivation for developing this calorimeter, how electromagnetic showers behave and how calorimeters work. The design of the mTower prototype and the setup of the test beam with this prototype is discussed in chapter 3. In chapter 4 the developed criteria are laid out and in chapter 5 their effectiveness is evaluated. A conclusion and discussion can be found in chapter 6. Chapter

7 contains a summary in Dutch for people with less experience in physics. Appendix A is a description of the code developed during the research project and appendices B and C contain additional figures and a table supporting the content of this thesis.

2 Theoretical Background

2.1 Motivation for the proposed calorimeter

A forward calorimeter is proposed to be installed for the ALICE experiment. This section will give the theoretical motivation for installing this calorimeter.

2.1.1 Quarks and gluons at high density

Atoms consist of protons, neutrons and electrons, where the first two make up the atomic nucleus. Both protons and neutrons consist of up and down quarks. Apart from these two types of quarks, there are also the charm and strange quarks and the top and bottom quarks. In normal situations quarks do not exist on their own, they always appear in combination with other quarks, forming hadrons. This is called *confinement*. Inside such a hadron, the quarks are held together by the strong nuclear force. The strong interaction is carried by massless particles called gluons. The field of study of the strong interaction is called *quantum chromodynamics* (QCD). There are many open questions about some properties of QCD, for example about how confinement works and what the distribution of constituents (i.e. quarks and gluons) inside protons or nuclei is.

When matter is at high temperatures and at high densities, the quarks are freed from the confinement by the strong nuclear force. When matter is in this state of deconfinement, this is called the *quark-gluon plasma* (QGP). To get a better understanding of how the strong nuclear force works, it is very insightful to see what happens when matter exists in the form of a QGP. About 10^{-12} seconds after the big bang, matter started forming in the form of quarks and gluons. Because of the high densities and temperatures, they existed as a QGP. Only after about 10^{-6} seconds, densities and temperatures were low enough to allow for the creation of hadrons [4]. This means that learning more about the QGP can teach us a lot about the early stages of our universe.

2.1.2 ALICE

The high densities and temperatures needed for the existence of a QGP can be reached when ions are made to collide with each other at very high energies. That is why particle accelerators such as the LHC are very useful for getting a better understanding of the QGP and quark confinement. The ALICE detector at CERN is designed to primarily study heavy ion (Pb-Pb nuclei) collisions. During those collisions, many particles are formed in the hot QGP. Some of the created particles leave the QGP and live long enough to reach the detectors. The goal for ALICE is to identify as much properties of those particles as possible. To do this, different layers of detectors are positioned around each other. Starting from the interaction point and going outwards there is first the inner tracking system (ITS). It identifies the decay of heavy particles with a small lifespan. Behind that there is the time projection chamber (TPC) which tracks charged particles. Then follows the transition radiation detector (TRD) which identifies electrons and particles, time of flight (TOF) which measures the velocity of particles and calorimeters (EMCal and DCAL) which measures the energy of particles. All of these detectors are surrounded by a large solenoid magnet of 0.5 Tesla. The detector size

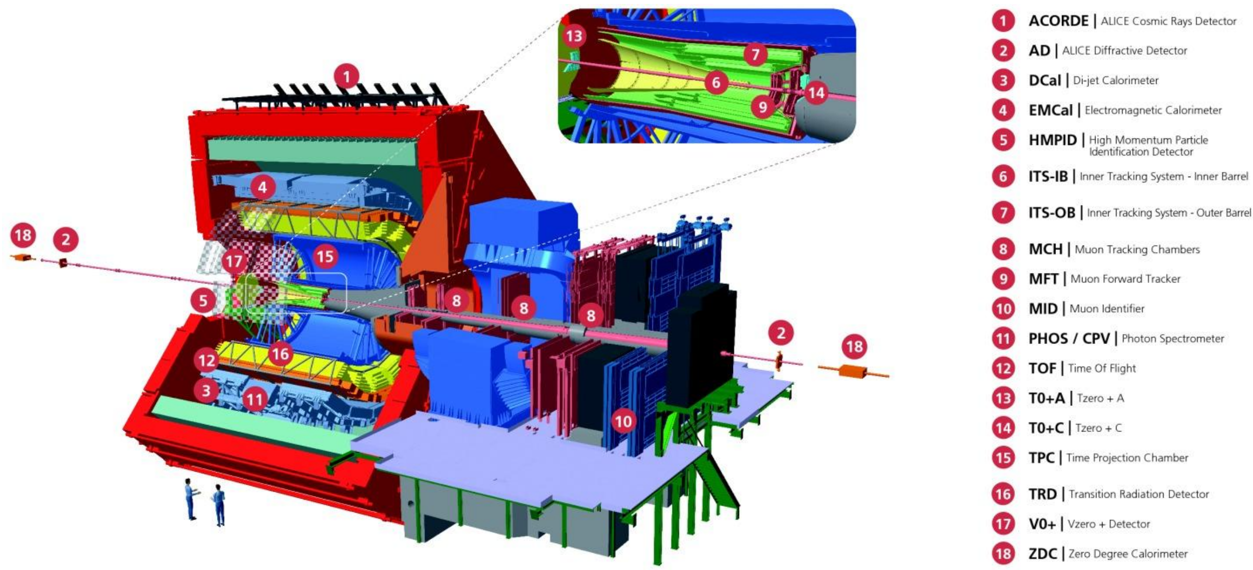


Figure 1: A schematic of the ALICE detector as during run 3 [7].

is $26 \times 26 \times 16$ m^3 [5, 6]. A schematic of the ALICE detector is shown in figure 1. The proposed calorimeter would be situated at about 7 m from the interaction point, on the left side of the figure.

2.1.3 Parton Distribution Functions

In the described collision experiments, the incoming protons and nuclei move with very high momentum. In that case the momenta of the quarks and gluons (together called partons) are approximately collinear with the momentum of the hadrons. The momentum fraction x of a parton is the fraction of the total momentum of the hadron which is carried by that parton. The momentum distribution functions of the partons in the hadron are called *Parton Distribution Functions* (PDFs). They describe the probability density to find a parton with a momentum fraction x at energy scale Q^2. Q^2 is the squared momentum transfer to the hadron in the collision [8, 9].

Unfortunately QCD does not predict the shapes of these PDFs, they can only be determined by experiment. At low x, not much is known about the PDFs, but this knowledge is very important for understanding collisions at high energy. Especially for moderate and low Q^2, it is unknown how the PDFs evolve towards low values of x. This results in an uncertainty in the interpretation of data gathered in a particle collision [10]. In figure 2 an example of the uncertainty in the description of the PDFs in nuclei via the nuclear modification factor R_g^{Pb} is shown for intermediate and low Q^2 at low x.

Although QCD does not predict the shapes of the PDFs, it does give a quantitative description of the rate of change of parton distribution functions when Q^2 is varied. This rate of change is given by the DGLAP differential equations. These equations can be used to find PDFs at low Q^2 when they are (better) known at high Q^2. There is a problem however,

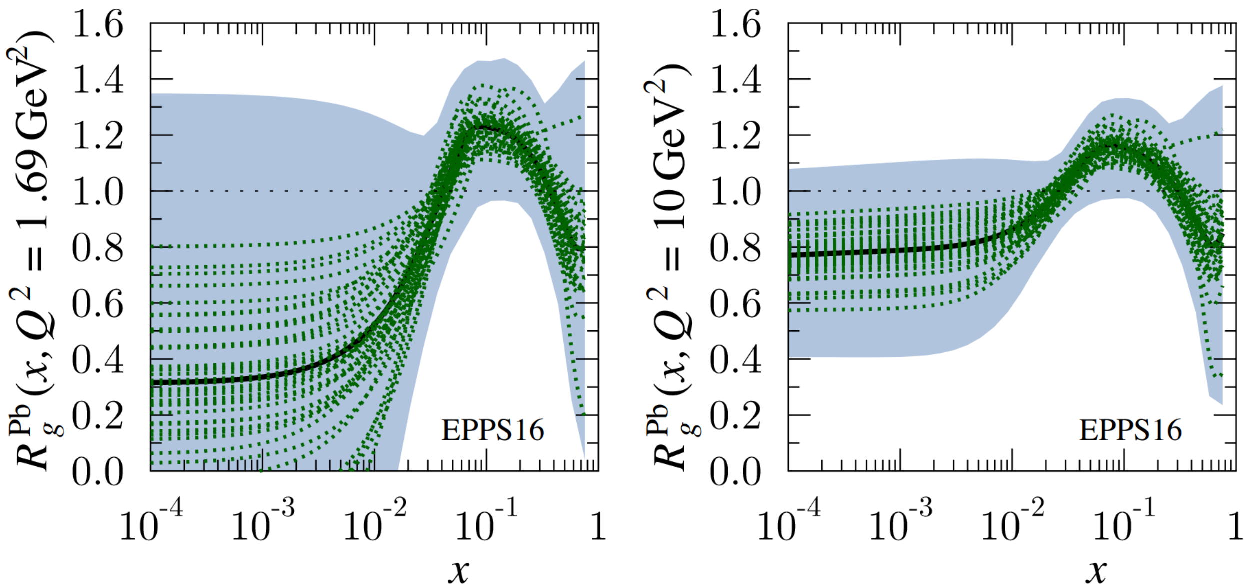


Figure 2: The large uncertainty of the nuclear modification factor R_g^{Pb} of gluon PDFs for low Q^2 on the left and moderate Q^2 on the right [11].

because these equations lead to a strong increase of the gluon density at low x, which cannot continue indefinitely. In fact, these equations use only linear terms for the evolution of the PDFs, so for the high parton density at low x, the DGLAP equations are not valid anymore. One has to use non-linear evolution equations, which lead to a limitation of the gluon density, called *gluon saturation*. This state of saturated gluon matter can most likely be described as the classical limit of the strong interaction, also called the *Color-Glass Condensate* (CGC) [12].

Particle production processes in hadronic collisions are sensitive to a minimally reachable value of x which can be estimated as:

$$x_{\min} \approx \frac{2p_{\mathrm{T}}e^{-y}}{\sqrt{s}}. \tag{1}$$

In this equation p_{T} is the transverse momentum of the outgoing particle, i.e. the momentum perpendicular to the direction of the incoming particle. $\sqrt{s}$ is the center of mass energy of the collision particles. y is the *rapidity* of outgoing particle [13].

A particle with velocity $\vec{v}$ has a rapidity $\vec{y}$, given by $\vec{y} = \hat{v}\,\mathrm{arctanh}\,\frac{v}{c}$ with $\hat{v}$ a unit vector in the direction of $\vec{v}$ and v the length of vector $\vec{v}$. With the definitions $E = \gamma mc^2$ and $|\vec{p}| = \gamma mv$ this can be rewritten as $\vec{y} = \hat{v}\,\mathrm{arctanh}\,\frac{|\vec{p}|c}{E}$. In particle physics many times only the longitudinal component of the rapidity is used. This gives us a rapidity

$$y = \mathrm{arctanh}\,\frac{p_{\mathrm{L}}c}{E}, \tag{2}$$

with p_{L} the longitudinal momentum of the outgoing particle, i.e. the momentum in the direction of the incoming particle. This is the rapidity y which is used in equation (1).

Sometimes the approximation $|\vec{p}| \approx \frac{E}{c}$ is used to define the *pseudorapidity* η as

$$\eta = \operatorname{arctanh} \frac{p_{\mathrm{L}}}{|\vec{p}|}. \tag{3}$$

This approximation holds for particles with $|\vec{p}|c \gg mc^2$ with m the mass of the particle.

As stated earlier, we are interested in the low x region of the PDFs, so the $x_{\min}$ given in equation (1) should be as small as possible. This can be done very effectively by investigating particles with a high rapidity y. From equation (2) we can learn that this involves particles whose momentum is pointed very close to the beam axis. This high rapidities are called *forward rapidities*, particles who travel with those rapidities end up in the *forward region* of the detector. The proposed calorimeter would be placed in this forward region of the ALICE detector.

2.1.4 Forward direct photons

In heavy ion collisions - among other particles - photons are created. A distinction is made between *direct* and *decay* photons. Direct photons are photons created directly in scattering processes, while decay photons are produced from hadronic decays of particles like neutral pions. Direct photons can be further classified as *thermal* or *prompt*. Thermal photons are created inside the QGP and tell us about its temperature. Prompt photons come from hard scattering of partons and carry information about the low x structure of the PDFs.

The measurement of direct photons is complicated by the presence of decay photons. The neutral pion, for example, decays into two photons that, at high energy and forward rapidity, will travel very close to each other. A differentiation between direct and decay photons is needed. To be able to do these measurements, an upgrade is proposed for ALICE. The proposal is the installation of an *electromagnetic calorimeter* in the forward region which can measure the energy of photons. To make sure the photons of different origins can be differentiated from each other, this forward calorimeter (FoCal) will be able to measure the position of photons very accurately, such that it can recognise two very close by photons. For this the calorimeter needs to have a very high granularity and a small Molière radius[1]. A prototype is developed to help with the final design of this calorimeter. Details on the design of the prototype can be found in section 3.1.

2.2 Electromagnetic showers

In a calorimeter, an incoming particle interacts with the calorimeter material. In these interactions, part of the energy of the particle is deposited in the material, which can be measured. The interactions induce an *electromagnetic shower*. To get a better understanding of how an electromagnetic calorimeter works, first an explanation about electromagnetic showers will be given, mostly based on [14–16].

[1] The Molière radius is a measure for the width of an electromagnetic shower and will be discussed in section 2.2.3.

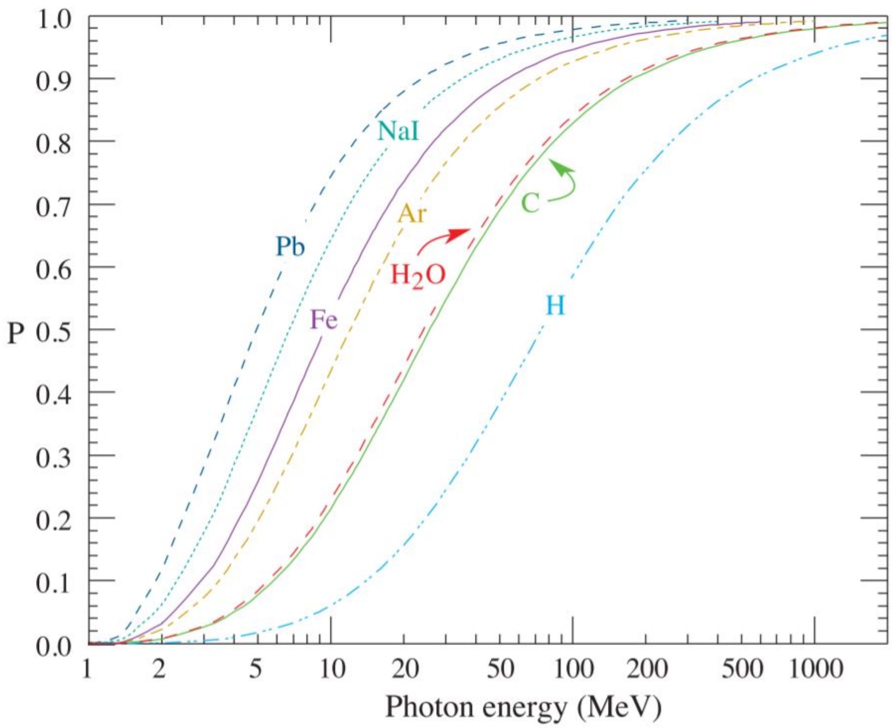


Figure 3: Probability that a photon interaction will result in conversion to an electron-positron pair, given for photons traveling through different materials [17].

2.2.1 Energy loss processes

When an electron, positron or photon interacts with matter, different processes take place, depending on the energy of the particle. High-energy electrons or positrons will mostly produce photons in the form of *bremsstrahlung*, while high-energy photons will mostly induce *pair production*.

Bremsstrahlung When an electron of positron gets close to the nucleus of an atom, it is decelerated due to the electric field of the nucleus. During this deceleration a photon is produced.

Pair production When a photon gets close to the nucleus of an atom, the photon creates an electron-positron pair. In figure 3 the probability of a photon creating an electron-positron pair is shown. This can only happen if a photon is close to the nucleus of an atom, to ensure conservation of momentum and energy.

These two processes are dominant at energies above 10 MeV. At lower energies other processes take place, such as *Compton scattering* and the *photo-electric effect* with photons, and *ionisation* with electrons and positrons. Figure 4 shows the fractional energy loss in lead of electrons and positrons due to different processes, such as ionisation and Bremsstrahlung, showing clearly that at higher energy Bremsstrahlung is the dominant source of energy loss.

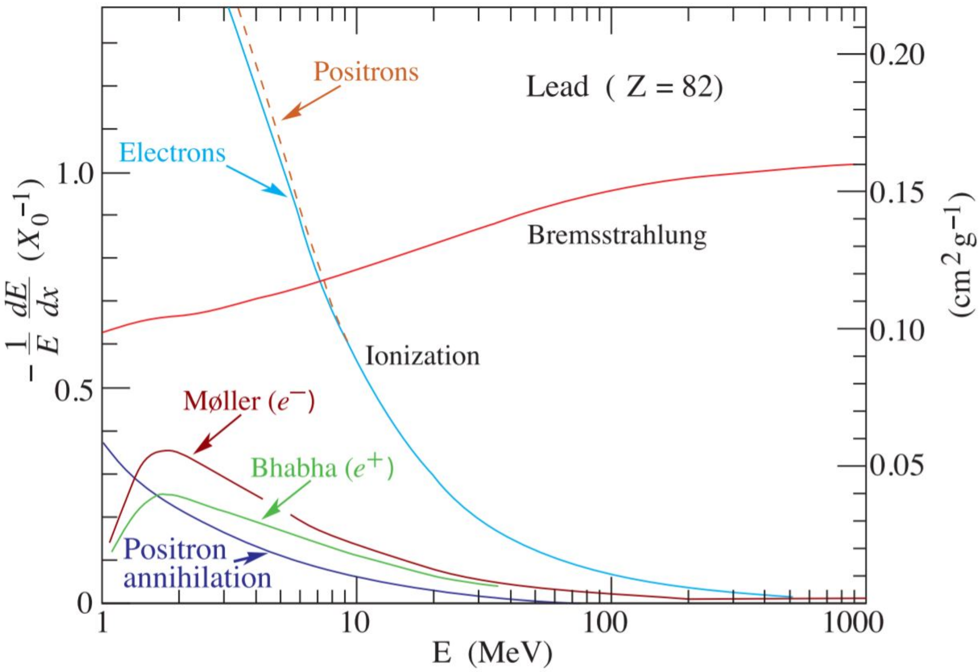


Figure 4: Fractional energy loss per radiation length in lead as a function of electron or positron energy [17].

Compton-scattering A photon can give part of its energy to an electron. By doing that, the electron gets freed from the atom.

Photo-electric effect A photon can also give all of its energy to an electron of an atom to free it.

Ionisation Positrons and electrons can come into collision with an atom. An electron will be kicked out of an electron shell, which leaves the atom with a positive charge.

Most of the processes mentioned above, especially the high-energy processes, double the number of particles. A photon will create an electron and a positron, which will both create photons, which will create electron-positron pairs, and so forth. This cascade of particles is called a *particle shower*. Since it involves positrons, electrons and photons it is called an *electromagnetic shower*. A simplified visualisation of such a particle shower is shown in figure 5. When the particles have lost a lot of energy they cannot produce new particles anymore because other energy loss processes become dominant, the shower comes to an end.

Critical energy Every time when a particle interacts through pair production or bremsstrahlung, the energy is divided among the resulting particles. Due to this effect, bremsstrahlung and pair production will appear less as the particle energies get smaller, while there will be more interactions due to Compton-scattering, the photo-electric effect and ionisation. The energy at which an electron loses as much energy due to radiation as due to ionization is called the *critical energy* ϵ_c. This is given by

$$\epsilon_c \approx 2.66 \cdot \left(\frac{ZX_{g0}}{A}\right)^{1.11} \text{ [MeV]}. \tag{4}$$

2.2.2 Longitudinal shower development

Radiation length When high-energy electromagnetic particles interact with a material, the cross sections of the pair-production and Bremsstrahlung processes are almost energy-independent. The cross section is a measure of probability that those processes take place in a collision between the electromagnetic particle and an atom in the material. In this situation particles are considered to have high energy, if the electron energy E satisfies

$$E \gg \frac{m_e c^2}{\alpha Z^{1/3}}, \tag{5}$$

with m_e the electron mass, c the speed of light, α the fine-structure constant and Z the atomic number of the medium. The fact that the cross sections can be considered energy-independent for high-energy particles, points to the existence of a natural length scale which describes the mean free path length of an electron while traveling in the material. In the longitudinal direction (along the travel direction of the incoming particle) this length scale is called the *radiation length* X_{g0}, given by

$$X_{g0} \approx \frac{A}{4\alpha NZ(Z+\varsigma)r_e^2 \ln \frac{183}{Z^{1/3}}} \text{ [cm}^2\text{ g}^{-1}], \tag{6}$$

with A the atomic weight of the medium, N Avogadro's constant, r_e the classical electron radius and ς a correction term, to take into account the contribution of atomic electrons to the overall bremsstrahlung process. When discussing the detector geometry, it is more useful to use a different representation of the same quantity, namely X_0, which is given in units of length. It is defined as

$$X_0 = \frac{X_{g0}}{\rho} \text{ [cm]}, \tag{7}$$

with ρ the density of the medium. This length scale can be used to calculate a normalised depth parameter t, which is more easily comparable for different materials. It is defined as:

$$t = \frac{x}{X_0}, \tag{8}$$

with x the distance traveled through the medium. Another useful quantity is the *total track length* T, given by

$$T = \frac{E_0}{\epsilon_c} \; [X_0], \tag{9}$$

with E_0 the energy of the incoming particle and ϵ_c the critical energy, given by equation (4). It represents the summed length of all individual tracks of showering charged particles, in units of radiation lengths.

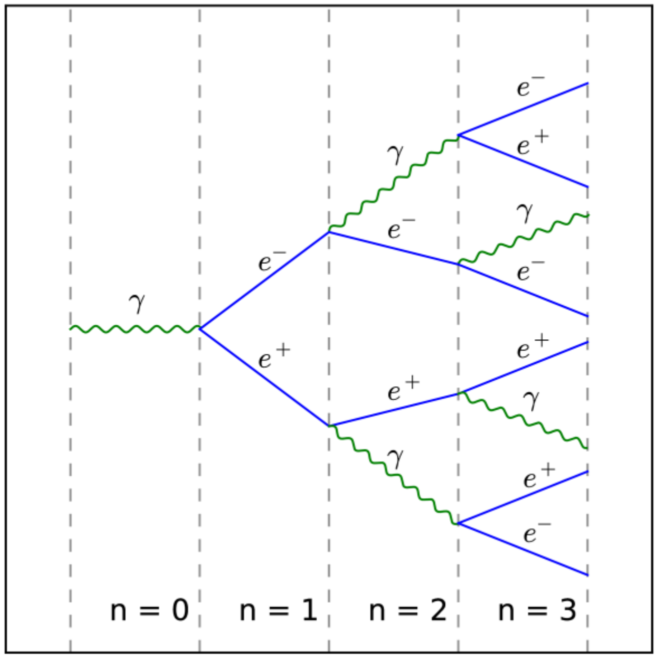


Figure 5: A visualisation of the simple model of an electromagnetic shower, where the number of particles doubles every radiation length [18].

Simplified model A very simple model describing the shower development, makes the assumption that every radiation length X_0, the number of particles doubles. This means that at depth t, the number of particles is given by

$$N(t) = 2^t. \tag{10}$$

Furthermore, the assumption is made that each time the number of particles is doubled, the energy is divided evenly over the particles, i.e. the amount of energy per particle halves. This leads to

$$E_p(t) = \frac{E_0}{N(t)} = E_0 2^{-t}, \tag{11}$$

with E_0 the energy of the incoming particle and $E_p(t)$ the energy per particle at depth t. A visualisation of this model is shown in figure 5. As discussed in the previous sections, this multiplication cannot continue indefinitely. In this simple model it is assumed that the doubling stops when the energy per particle E_p reaches the critical energy ϵ_c. When this happens, the number of particles is at its maximum, because after that less and less new particles are created because energy loss through ionisation becomes dominant. This is called the *shower maximum*. The depth of this shower maximum t_{max} can be found by evaluating equation (11) at $E_p(t_{max}) = \epsilon_c$, which gives us

$$t_{max} = \log_2\left(\frac{E_0}{\epsilon_c}\right). \tag{12}$$

This logarithmic dependence shows why calorimeters are very effective for measuring energy of high energy particles. If the energy of a particle doubles, the length of the detector has

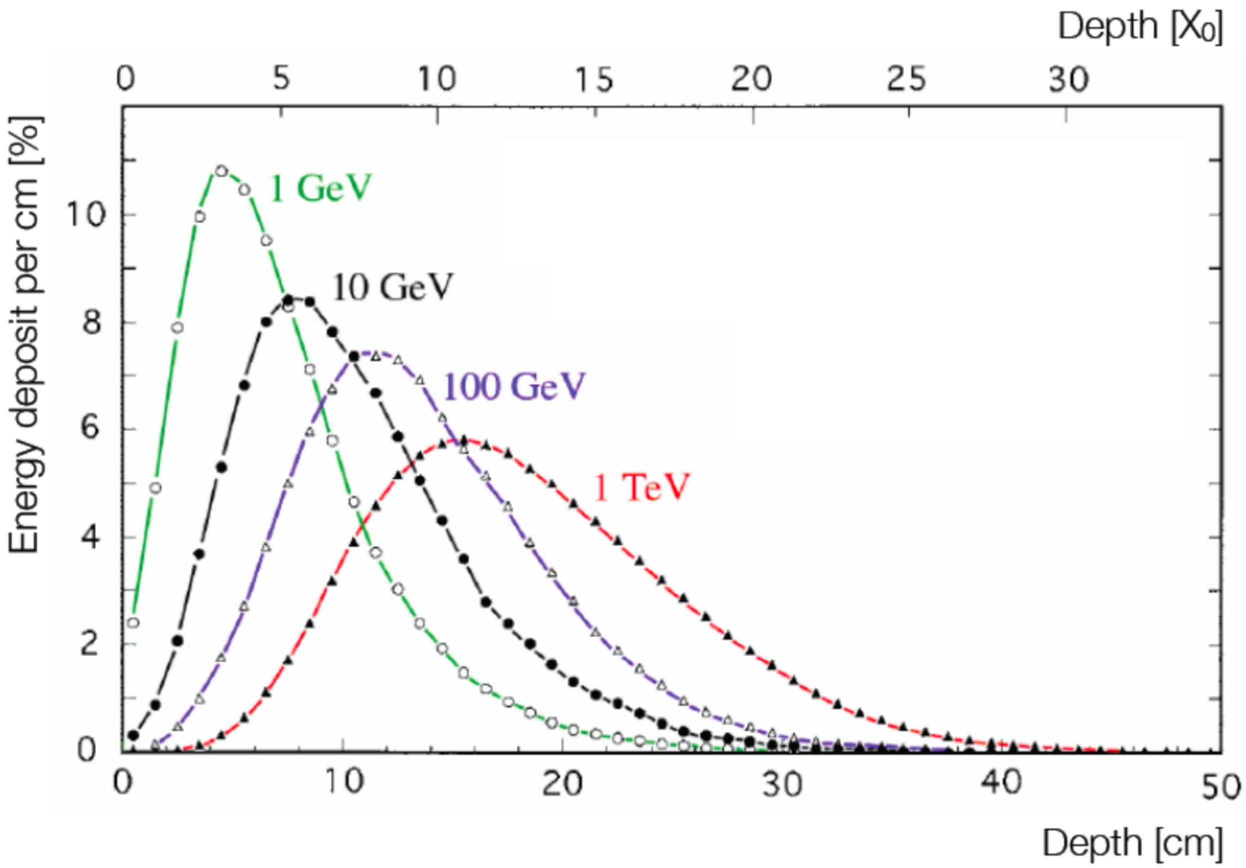


Figure 6: The longitudinal shower shape in copper for different energies according to theory. The integrals are normalized to the same value [20].

to increase with just a constant value. This means that particles with very high energy can still be measured with relatively short detectors.

Realistic model A more realistic distribution of the longitudinal shower shape is given by

$$\frac{\mathrm{d}E}{\mathrm{d}t} = E_0 \frac{b^{\omega+1}}{\Gamma(\omega+1)} t^{\omega} e^{-bt}, \tag{13}$$

with b and ω fit parameters and $\Gamma(\omega+1)$ the gamma distribution function. This model is proposed by Longo and Sestili based on Monte Carlo simulations of electromagnetic showers in lead glass [19]. It does a decent job at describing the longitudinal shower shape, but it is still an approximation. An example of the result of this model for copper is shown in figure 6.

According to this model, the shower maximum is located at

$$t_{max} = \frac{\omega}{b} = \ln\left(\frac{E_0}{\epsilon_c}\right) + f, \tag{14}$$

with $f = -0.5$ if the incoming particle is an electron, and $f = +0.5$ if the incoming particle is a photon.

In the longitudinal direction 95% of the shower is contained at a length

$$L(95\%) \approx (t_{max} + 0.08Z + 9.6)\ [X_0]. \tag{15}$$

2.2.3 Lateral shower development

Molière radius An electromagnetic shower does not only develop in the longitudinal direction (along the travel direction of the initial particle), but also in transverse directions (perpendicular to the travel direction of the initial particle). The development in the transverse direction is called the *lateral* development. This is mostly caused by lower energy particles, due to the photo-electric effect and Compton-scattering. Similar to the radiation length, it is also possible to define a transverse length parameter. This is the *Molière radius* R_M, given by

$$R_M = \sqrt{\frac{4\pi}{\alpha}} \left(\frac{m_e c^2}{\epsilon_c}\right) X_0, \tag{16}$$

with α the fine-structure constant, m_e the mass of the electron, c the speed of light, ϵ_c the critical energy defined in equation (4) and X_0 the radiation length defined in equation (7). In the radial direction 95% of the shower is contained within a radius

$$R_e(95\%) \approx 2R_M. \tag{17}$$

Lateral distribution The lateral distribution depends on the depth t. To describe this distribution, at least two components are needed. Close to the shower core the lateral distribution scales approximately with R_M. This part of the shower is dominated by scattering high-energy electrons. At a larger distance from the shower core, the share of photons increases. Several parametrizations are used to describe this distribution, there is no generally accepted model for the lateral distributions. An example of a parametrization is

$$f(r) = \frac{2rR^2}{(r^2 + R^2)^2}, \tag{18}$$

with r the radial distance to the shower axis and R a free parameter in units of R_M. This can be used as long as the shower is measured with a calorimeter with a resolution of at least order $1R_M$. The mTower prototype can help with discovering more about the lateral distribution of electromagnetic showers.

2.3 Calorimeters

A calorimeter is a measuring device, which can measure the energy of particles by fully absorbing them. They contain a dense material, in which a particle loses all of its energy in a shower, as described in the previous section. The information in this section is mostly based on [14, 21].

Based on measurements of the energy deposited in the dense material, the total energy of the incoming particle is calculated. Calorimeters are generally categorised based on two characteristics.

Hadronic vs. electromagnetic The first attribute based on which they are categorised is the kind of incoming particle that they are designed to measure. Calorimeters which are optimised to measure hadrons are called *hadronic*, and calorimeters which are optimised to measure electrons, positrons and photons are called *electromagnetic*. Hadronic showers develop in a different way than electromagnetic showers. There is no simple model possible to describe them, contrary to electromagnetic showers, as seen in section 2.2.2. Since the calorimeter discussed in this thesis is an electromagnetic calorimeter, hadronic calorimeters and hadronic showers will not be discussed.

Homogeneous vs. sampling Calorimeters are also categorised as either a *homogeneous* or a *sampling* calorimeter. In a homogeneous calorimeter the absorption material (in which the shower is induced) is also the sensitive material. In a sampling calorimeter, the absorption material is different from the sensitive material. The detector is consists of alternating layers of these materials. The benefit of a sampling calorimeter is that each material can be optimised to fit its goal. The drawback is that the energy which is deposited in the absorption material will not be measured, so only a fraction of the total energy is measured. The sampling fraction f_{samp} of a sampling calorimeter is given by

$$f_{samp} = \frac{E_S}{E_S + E_A}, \tag{19}$$

With E_S the amount of energy deposited in sensitive layers and E_A the amount of energy deposited in absorption layers. To calculate the total amount of deposited energy in the calorimeter, the measured energy should be divided by f_{samp}. The calorimeter discussed in this thesis is a sampling calorimeter, so after this only sampling calorimeters will be discussed.

2.3.1 Energy resolution

Equation (9) shows that the total track length is proportional to the energy of the incoming particle. The measured energy in the detector is proportional to the track length in the sensitive material, which is a fraction of the total track length. That means that the total measured energy E is proportional to the energy E_0 of the incoming particle. This very simple relation shows why an electromagnetic calorimeter is so effective in measuring the energy of electromagnetic particles. Note that this linear relation does not hold for hadronic showers in an electromagnetic calorimeter.

How well the energy of the incoming particle is estimated in a calorimeter is called the *energy resolution*. The shower development is a stochastic process. That means that the signal is governed by number fluctuations, i.e. Poisson statistics, so the size of variations is related to the square root of the signal:

$$\frac{\sigma(E)}{E} \propto \frac{1}{\sqrt{E_0}}. \tag{20}$$

In reality the energy resolution is influenced by more than just the stochastic uncertainty. For example, electronics create a constant electronics noise. As seen in equation (20), the energy

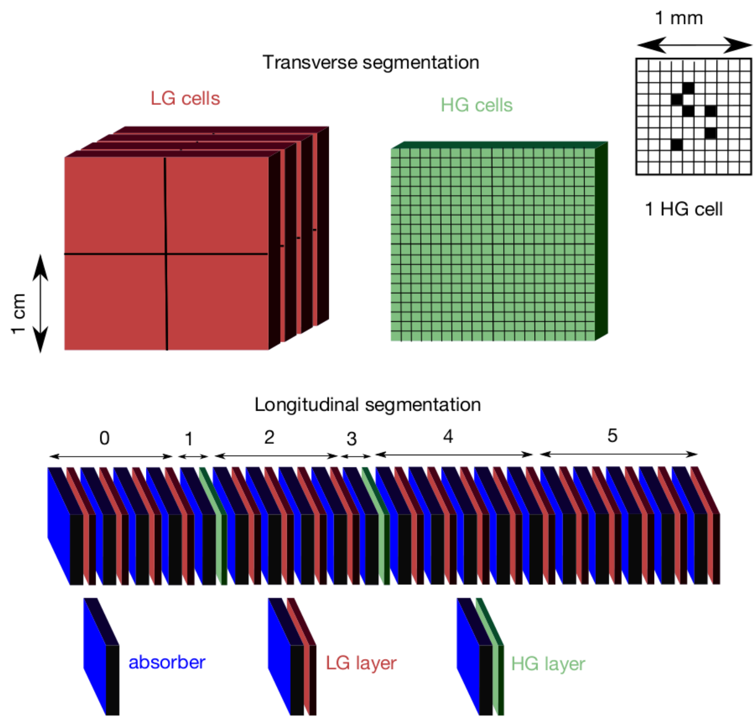


Figure 7: The design of the proposed FoCal design [13].

resolution is defined as the relative uncertainty, so this contributes with a factor $1/E$. Other uncertainties are created by for example nonuniformities in either the absorbing material or the sensitive matarial. This contributes to a constant term in the energy resolution. If a calorimeter is too short or has a too small width, part of the shower will leave the detector, this is called *leakage*. Usually sensitive layers of calorimeters consist of different components. There can be gaps between the sensitive parts of these different components. No particles will be measured in those regions. All of these effects deteriorate the energy resolution. This more realistic energy resolution can be described by

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{21}$$

with a, b and c the constants described above. $\oplus$ indicates a quadratic sum.

3 The mTower

As stated in section 2.1.4, the installation of an electromagnetic forward calorimeter (FoCal) is proposed for the ALICE experiment. This calorimeter needs to be able to differentiate between direct and decay photons. To be able to do this, a calorimeter with high granularity and small Molière radius is required. The calorimeter would be situated at about 7 m from the detection point. This will allow measurements of particles with pseudorapidities of $3.2 < \eta < 5.3$. This allows for discovering more about the PDFs for x down to 10^{-5} [3]. FoCal would consist of an electromagnetic calorimeter, with a hadronic calorimeter behind it.

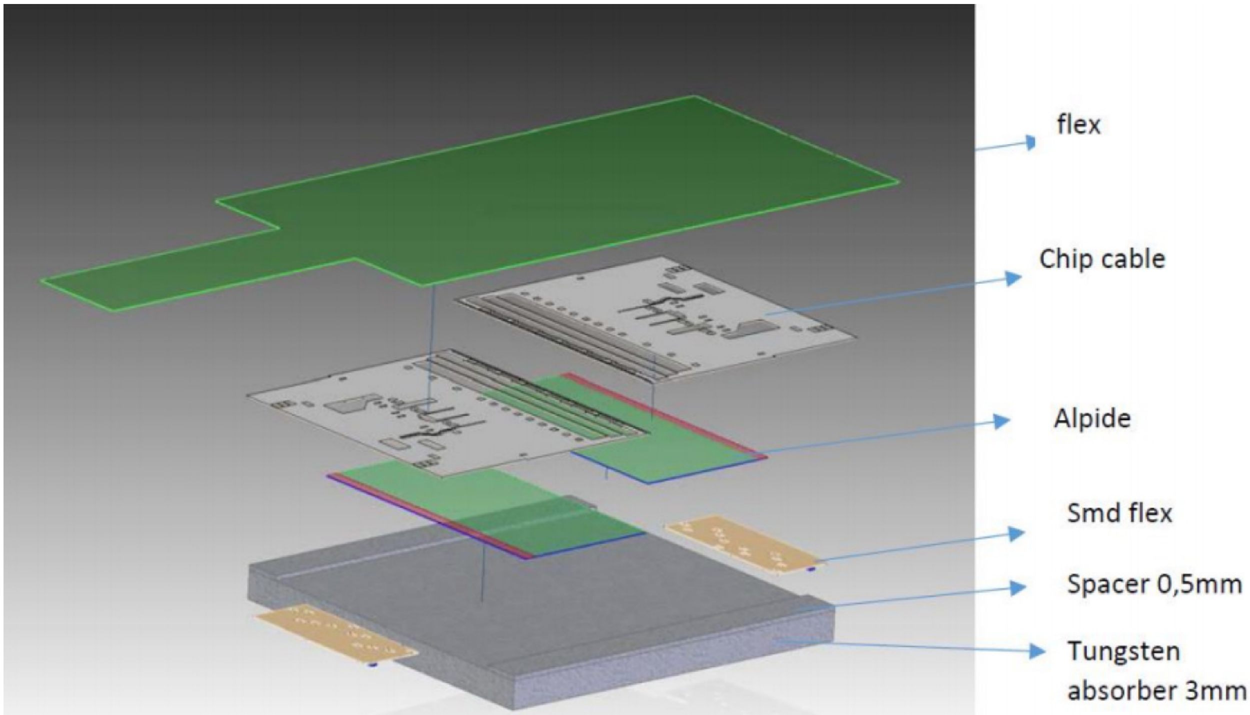


Figure 8: The design of a single mTower layer [3].

The electromagnetic calorimeter would be a sampling calorimeter with two low granularity layers using silicon pad detectors and 18 high granularity layers using monolithic active pixel sensors (MAPS). A schematic overview of the proposed detector is depicted in figure 7. A prototype has been built to research the use of high granularity layers. This prototype is called the mTower (mini Tower). In this section the design, properties and test setup of this prototype will be discussed.

3.1 Design

This calorimeter is a sampling calorimeter with 24 layers. Each layer consists of a 3 mm tungsten absorber on which silicon sensors and thin cabling is glued. As the silicon sensors are at the front of the layers (only covered by the cabling) the last tungsten absorber is effectively not used. In section 2.2.2 the radiation length is discussed as a useful quantity when particles have sufficient energy, as stated in equation (5). Tungsten has atomic number $Z = 74$. This gives $\frac{m_e c^2}{\alpha Z^{1/3}} = 17$ keV. Energies of initial particles discussed in this thesis are between 1 and 6 GeV, so these are definitely high-energy particles. Tungsten has a radiation length of 3.5 mm [22], so each absorber is 0.86 radiation lengths. This gives a total thickness of almost 20 radiation lengths. The initial particles discussed in this thesis have a maximum energy of 5.8 GeV. Equation (15) shows us that at that energy, the longitudinal length at which 95% of the particles is contained, is equal to approximately 21 radiation lengths.

The silicon sensors which are glued to the absorber are 2 chips with an area of 30×15 mm^2 each, spanning a total area of 30×30 mm^2. The Molière radius of tungsten is 9.3 mm

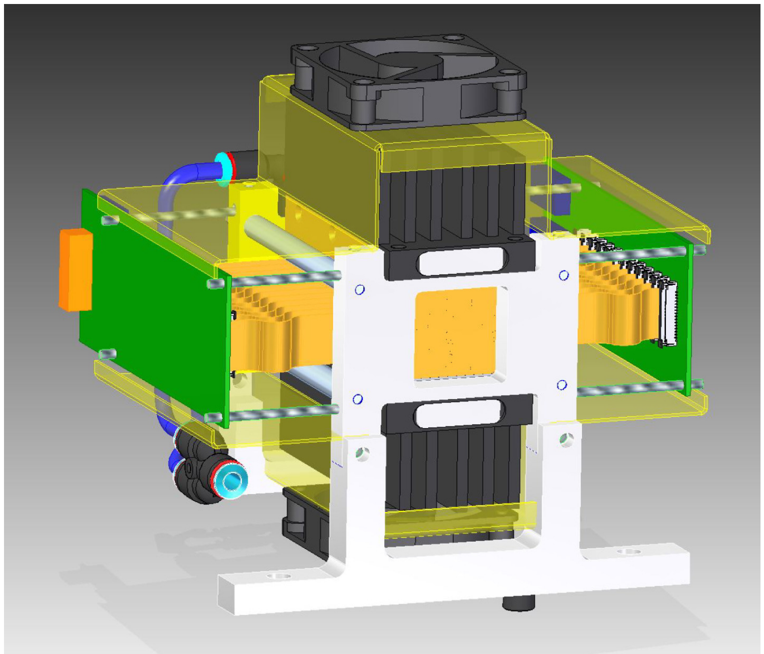


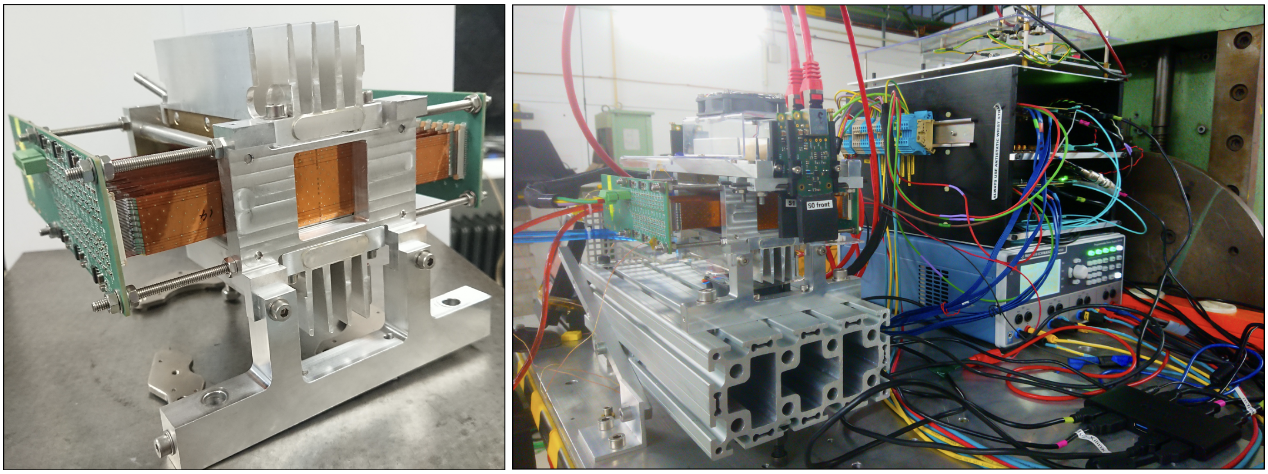
Figure 9: The design of the full mTower [3].

[22], so the width and length are about 3 Molière radii. The silicon sensors and the cabling slightly increase the Molière radius of the detector.

The design of a single mTower layer is depicted in figure 8. The chips are glued on top of the tungsten absorbers. On top of that the chip cables are applied to be able to read out the pixels. These consists of a layer of approximately 20 μm polyimide and a layer of approximately 30 μm aluminium. These chip cables are connected with the readout units via an aluminium-polyimide flex cable, with a total thickness of about 180 μm. The total thickness of chips, cabling and glue is less then 0.5 mm. Figure 8 shows that the tungsten absorber is a bit wider then the sensors, chip cable and flex on two sides. At these locations, next to the sensors, two tungsten spacers of 0.5 mm thin are placed to protect the sensors and cabling when the layers are stacked. The full mTower design is depicted in figure 9. A photo of the mTower is shown in figure 10 (a).

In front of the prototype two plastic *scintillators* of the same area are attached, as can be seen in figure 10 (b). When a particle passes through a scintillator, a flash of light is produced, which is captured by a silicon *photomultiplier* (SiPM), that converts the light into an electric signal. The signals are send to an electrical AND port. That means that if the signals from both of the SiPMs are above a certain threshold at the same time, a signal is send to the readout boards to start gathering data. This signal is called a trigger. The data gathered after a trigger is called an *event*. Ideally an event contains the response of the detector to a single electron.

The scintillators and SiPMs are covered in black tape to prevent external light sources sending light to the SiPM, resulting in a false trigger. Despite that, it is possible that a light



(a) A photo of only the mTower (b) The setup for the test measurements at DESY

Figure 10: Photos of the mTower prototype

signal is detected by a SiPM when no particles passes through the scintillator. The main source of this noise is the dark count rate (DCR). These dark counts originate from thermally generated electrons. The scintillator does not distinguish between thermally generated electrons and high energy electrons/positrons from the electron/positron beam [23]. The dark count rates of the used SiPMs are not bigger then 1 Hz for the used settings. A false trigger will only appear if both SiPMs have a dark count at approximately the same time. Because of this, the rate of false triggers is about 10^{-7}, which is extremely low.

3.2 ALPIDE chips

The silicon sensors are ALPIDE chips [24]. This are CMOS monolithic active pixel sensors (MAPS). Every chip has 512×1024 pixel cells, with an area of $29.24 \times 26.88\ \mu\text{m}^2$ each. The pixels are fabricated on a substrate with a high-resistivity epitaxial layer on top of it. When a particle enters the chip, charge carriers are liberated inside of this epitaxial layer due to ionization. A visualisation of this is shown in figure 11. This charge diffuses through the epitaxial layer and is collected by small collection diodes, there is one diode per pixel. When the data readout starts after a trigger, the amount of collected charge in a diode will be measured for a short amount of time. The length of this time interval is called the *strobe length*. If the amount of collected charge passes a certain threshold, the pixel will send a signal to the readout unit. This is called a *hit*. The fact that a pixel can have just two states, hit or no hit, makes this a digital sensor. Because the epitaxial layer is not separated between pixels it is possible that the liberated charge will create hits in several pixels. This is called *charge sharing*. Particles with a higher energy will generally liberate more charge carriers, with the result that more diodes collect charge over threshold, i.e. several neighboring pixels will measure a hit. Such a group of adjacent hits is called a *cluster*. The energy of an incoming particle is estimated by counting the total number of hits in the detector. The amount of hits in the detector is related to the total track length, since the ALPIDE chips could be thought of as very thin detection layers where energy is deposited. As stated in

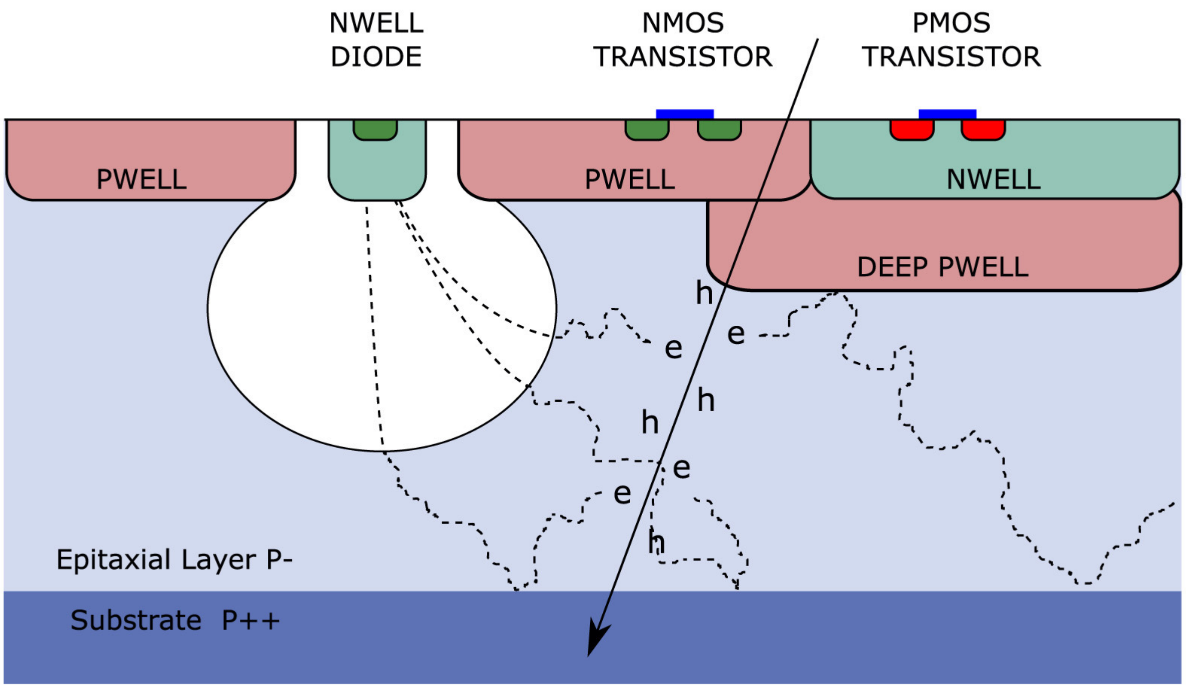


Figure 11: A schematic overview of the charge collection in the ALPIDE chips [25].

section 2.3.1, the relationship between the total track length and the energy of the incoming particle is expected to be linear. This linear relationship also holds for the number of hits in the detector verses the energy of the incoming particle.
The ALPIDE chips are being kept at a steady temperature of about 20°C by using water cooling.

3.3 mTower test beam setup

In February 2020, test measurements have been performed with the mTower at the DESY test beam facility in Hamburg. The information in this section about the DESY test beam facility is based on [26, 27].

3.3.1 Test beam generation

Free electrons are generated at a cathode, which are accelerated towards an anode. They arrive at the LINAC II (Linear Accelerator II), where they are accelerated towards an energy of 0.45 GeV. Every 160 ms about 10^{10} electrons or positrons are injected from the LINAC II via a small storage ring (PIA, Positron Intensity Accumulator), into a circular particle accelerator, the DESY II. This accelerator has a circumference of almost 300 m, and accelerates the electrons from initial energy 0.45 GeV to 6.3 GeV, or even as high as 7 GeV. One group of injected electrons is called a *bunch*. During those 160 ms, it travels around the accelerator twice, after which the new injection takes place. The electrons in the ring come into collision with a fiber target. This happens at three distinct locations, to create particle beams in three test areas. These collisions create Bremsstrahlung photons, which collide with a

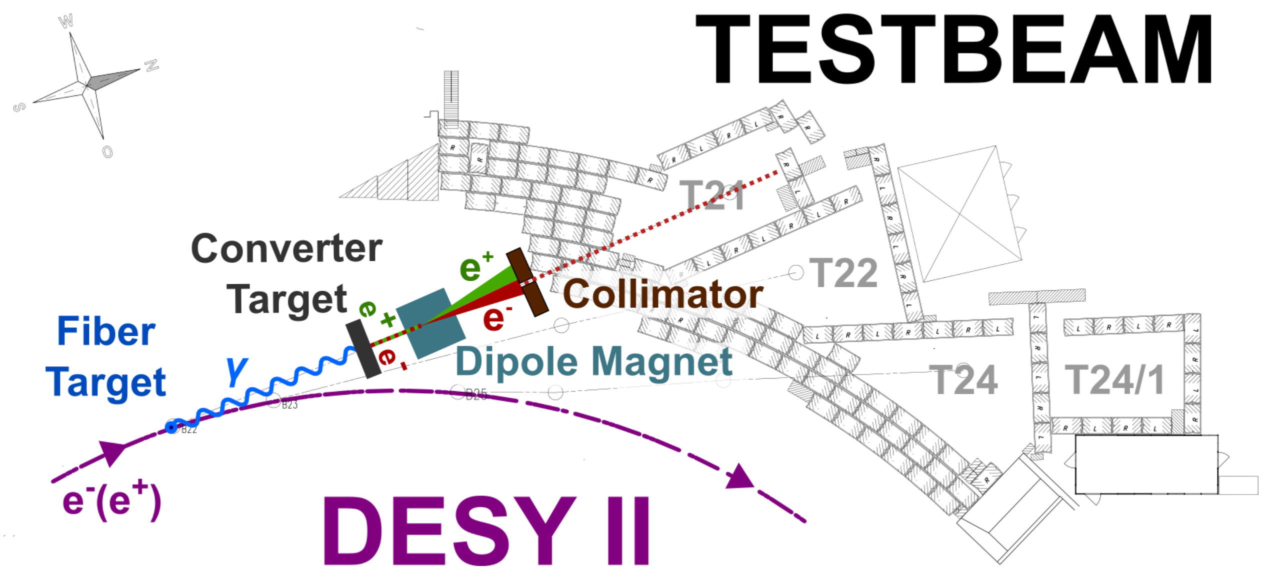


Figure 12: A schematic overview of the test beam generation at DESY for beam line TB21 [27].

secondary (converter) target, creating electron-positron pairs. These electron-positron pairs are separated by a dipole magnet. The polarity and strength of the magnetic field induced by this magnet can be varied, to select electrons or positrons with different energies. After the secondary target, the particles travel through a vacuum towards a collimator. A *collimator* is a heavy metal object with a rectangular hole in the center. It is used to narrow the beam. The selected particles go through the *beam shutter*. The beam shutter is a device which can put a 40 cm long lead block with a diameter of 20 cm in front of the beam when people are inside of the test area. After going through the beam shutter, the beam can again be collimated by a second collimator, after which the test beam can be used to do measurements. A schematic overview of the described test beam generation can be found in figure 12.

Every few minutes all of the particles from the accelerator are injected into another particle accelerator, PETRA III. That means that every few minutes there are no particles entering the test beam area.

3.3.2 Test setup

The measurements used in this thesis have been done in test beam line TB22. A schematic overview of the setup during these measurements is depicted in figure 13. The first collimator (as depicted in figure 12) was a tungsten collimator of 14×14 mm^2, the second collimator (as depicted in figure 13) was a lead collimator of 12×12 mm^2. A photo of the setup is shown in figure 10 (b).

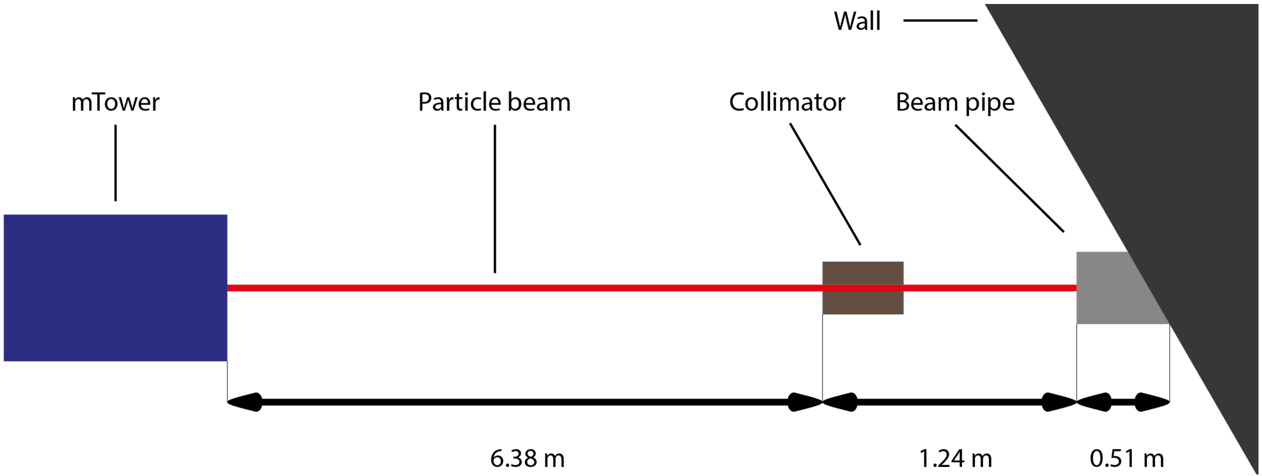


Figure 13: A schematic overview of the mTower test beam setup in February 2020 as seen from above. This picture is not to scale.

4 Event selection

In February 2020, the mTower is tested at the DESY test beam facility in Hamburg, by firing an electron beam at the prototype. The goal of this test was to find out if the prototype gives results as expected, to draw up points of improvement on the prototype and to do analysis on electromagnetic shower behaviour, such as the shape of the produced electromagnetic shower.

As stated in section 3.1, an *event* is the data gathered after a trigger originating from the scintillators. If two or more electrons enter the prototype in a short amount of time, it is possible that they will be measured during the same event. This is called *pileup*. Events with more than one electron are not useful for many analysis purposes. This thesis describes the composing of selection criteria to select events which are useful for data analysis. These criteria are based on the location of the hits in the detector. The selection criteria are implemented using C++ and the ROOT framework.

4.1 Test beam data

The measurements at DESY were done at several energies: 1.0 GeV, 2.0 GeV, 3.0 GeV, 4.0 GeV, 5.0 GeV and 5.8 GeV. Per energy between 1.2 and 3.5 million events were recorded using electrons under standard conditions. Besides that some extra tests were done, for example measurements with the mTower put under certain vertical angles, with the mTower reversed, with the beam pointed at different locations of the layers and with different temperatures at the ALPIDE chips. Measurements with positrons instead of electrons were also done. The data used in this thesis only uses the measurements at standard conditions.

The measurements were done in different *runs*. A run consists of *cycles*, which in turn consist of *events*. During one cycle, between 1500 and 3000 events are measured and temporary stored in the readout board. If the cycle is done, the data is transferred from the readout board to the storage device. This takes a few seconds. For that reason it is optimal to have as many events in one cycle as possible. If too many events are measured in one cycle however, the maximal storage capacity of the readout board is reached, and any events measured after that cannot be stored anymore. Events with lower energy incoming particles generally contain less hits, so when dealing with lower energy particles, the amount of events per cycle can be higher compared to when dealing with higher energy particles. A run is a series of cycles recorded immediately after each other. All events in a run are be measured under the same conditions. The runs are labeled by a unique run number, to make referring to specific runs easy.

When a pixel is activated during an event this is called a *hit*. *Noise hits* are random hits, measured in a pixel that is not correlated to the incoming particle. If a pixel consistently keeps measuring hits, even if there are no particles being fired at the detector, that malfunctioning pixel is called a *hot pixel*. These unreliable pixels are found by doing a *pedestal run*. During such a measurement run, no particles are fired at the detector, i.e. the beam shutter is closed. The pixels which still measure many hits are turned of. This process is called *pixel masking*. Such a pedestal run has been done every few hours.

During the test beam, one chip was disabled, because the readout of this chip did not work properly. This chip is situated in the 22$^{\text{nd}}$ layer, so it has a very small influence on the data analysis.

4.2 Developed selection criteria

In this section the different criteria will be explained. The effect of these criteria on the set of events will be discussed in section 5.

4.2.1 Criterion 1 - Only central events

The focus of the first criterion is making sure the particles entered in the central part of the detector. If a particle enters the detector close to the edge, part of the produced particles will leave the detector, the shower will be just partially contained. If this happens, less hits will be measured than expected. To make sure there are as few of these events as possible, only events with no hits outside of the central 124×125 pixels (that is about 3.5×3.5 mm^2) of the detector in the first layer will be accepted. This selection criterion has two purposes: It picks events where the shower is well contained within the calorimeter, and it also rejects pileup with beam particles hitting the periphery of the detector.

4.2.2 Criterion 2 - Only events with 1 cluster in first layer

This criterion looks at the number of clusters in the first layer. A cluster is a group of one or more horizontally, vertically or diagonally adjacent hits. When an electron enters the detector, it is expected to create exactly one cluster in the first layer. Because of this, counting the number of clusters in the first layer is a sensible criterion for sorting events into groups of different number of electrons. Since this selection is pointed at finding events with one electron entering the detector, only events with one cluster in the first layer are accepted.

4.2.3 Refinement of criterion 2 - Search for hit behind clusters

This criterion gives a good first approximation of the number of particles entering the detector, but refinement is necessary to make the event selection better. A potential problem with this crude criterion are noise hits in the first layer. The pixel mask helps a lot with 'cleaning up' the data, but it does not help against random noise hits, caused by for example thermal fluctuations.

When a particle causes hits in the first layer, there will be hits behind it in the second layer as well. If the hit in the first layer is a noise hit however, there are probably no hits behind it in the second layer. To make sure a cluster that is found in the first layer is not the result of random noise, the area behind every cluster is searched, to see if there are hits behind it as well. If there are hits, this cluster is accepted. If there are no hits behind it, this cluster will be assumed to be a result of noise, and the cluster is ignored.

The exact procedure of this search for hits is as follows. First the coordinate of the cluster in the first layer is found (the mean coordinate of the hits in the cluster). After that, every hit in the second layer is checked. The distance between the coordinate of the hit in the second layer and the coordinate of the cluster in the first layer is calculated. Note that this distance is calculated in the plane of the layers, so the distance between the layers is not included in this distance calculation. If this distance is smaller than 10 pixels for any hit in the second layer, the cluster is accepted. If this is not the case, the cluster is ignored. After this refinement the number of electrons in an event is estimated as the number of accepted clusters in the first layer. Only events with exactly one accepted cluster in the first layer are accepted.

A cluster is only accepted if there is a hit in the second layer closer than 10 pixels. An alternative way of writing this down, is considering a *search area* in the second layer, centered around the mean coordinate of the cluster in the first layer. If there is a hit in this search area, the cluster is accepted.

4.2.4 Criterion 3 - Reject events with non negligible ignored clusters

If a cluster in the first layer is ignored in the previous step, but the cluster has more than one hit, it might be the case that it is still caused by a particle, but that for some reason there are no hits found behind the initial cluster. This can for example be the case if the particle (or the particles it created) travels further than 10 pixels away, or if it goes through the second layer without activating a pixel. For this reason every event with ignored clusters containing more than one hit, is rejected so it will not be used for analysis.

4.2.5 Criterion 4 - Reject events with 'loose' hits

The expectation is that there will only be hits behind the accepted cluster. If there are hits found in the second layer, which are further than 120 pixels away from the accepted cluster in the first layer, the event is rejected. This search is very similar to the search described in section 4.2.3. First the coordinate of the accepted cluster in the first layer is found. After that, every hit in the second layer is checked. The distance between the coordinate of the hit in the second layer and the coordinate of the cluster in the first layer is calculated. If this distance is larger than 120 pixels (approximately 3.5 mm) for any hit in the second layer, the event is rejected. If all hits are closer than 120 pixels, the event is accepted.

4.2.6 Criterion 5 - Reject events with many hits on the border

This criterion extents the idea of the fourth criterion to later layers. Because showers get broader, it is dangerous to allow there to be no hits outside of a certain area in later layers. Instead of allowing no hits outside of a circle, the hits inside of a circle are compared with hits very far from the circle. Two areas are defined. The first area is a circle centered around the accepted cluster in the first layer. This circle has a radius of 80 pixels. The second area is the area close to the edge of the layer, every pixel within 170 pixels from that border. The total number of hits in the third layer until the tenth layer within those two areas are counted. The fraction f is defined as the number of hits on the border divided by the number of hits in the circle,

$$f = \frac{N_{\text{hits on border}}}{N_{\text{hits in center}}}. \tag{22}$$

We expect that there will be many more hits around the shower center, so in the circle, which corresponds to a low value of f. For that reason, this fraction should not be larger than a certain critical fraction f_c,

$$f = \frac{N_{\text{hits on border}}}{N_{\text{hits in center}}} \leq f_c. \tag{23}$$

This critical fraction f_c is set to be 0.15. If the fraction f is larger than f_c, the event will be rejected. The first two layers are not used for this analysis, since criterion 1 makes sure there are no hits in the borders in the first layer and criterion 2 makes sure there are no hits in the borders in the second layer, so the number of hits in those layers would always be zero.

5 Results and discussion

In this section the effectiveness of the criteria to select single electron events will be shown and discussed. An important aspect to keep in mind is the possibility of a bias entering the selection. Every time an extra criterion is made, there is a chance of a bias occurring. This is because every time some unwanted events are rejected, inevitably some good events will be rejected as well. This is not a very large problem, as long as the rejected part of good events is a good representation of all of the good events. When the good events which are rejected all possess a relevant property, the selection of events that remains will not be a good representation of all of the good events, i.e. the selection is biased. For every criterion the possibility of a bias is discussed.

The figures that are shown in this section are from a 5 GeV measurement run, because the effects of the criteria can be best seen in a run with that energy. The result of a single 5 GeV measurement run will be shown instead of a combination of all of the 5 GeV runs, because it has not been determined yet if all of the measurement runs of the same energy are equivalent. The run with the most events, run 1413 with 1,103,740 events, has been chosen.

	5.8 GeV		5.0 GeV		4.0 GeV	
	Events	**%**	**Events**	**%**	**Events**	**%**
Raw data	1,516,041	100	3,483,904	100	2,176,165	100
After criterion 1	80,593	5.3	162,381	4.7	78,364	3.6
After criterion 2	45,607	3.0	95,432	2.7	41,567	1.9
After criterion 3	44,972	3.0	94,116	2.7	40,896	1.9
After criterion 4	36,771	2.4	76,083	2.2	32,660	1.5
After criterion 5	35,702	2.4	72,895	2.1	30,473	1.4

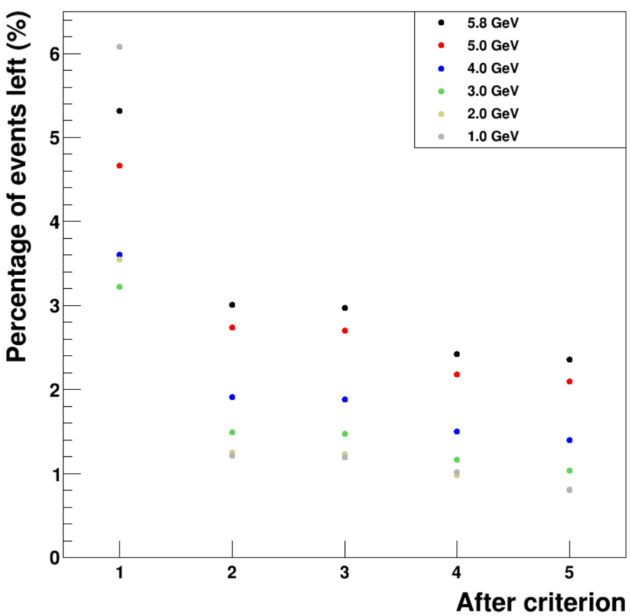
	3.0 GeV		2.0 GeV		1.0 GeV	
	Events	**%**	**Events**	**%**	**Events**	**%**
Raw data	1,288,500	100	1,432,808	100	1,860,000	100
After criterion 1	41,498	3.2	50,770	3.5	113,043	6.1
After criterion 2	19,162	1.5	17,838	1.2	22,539	1.2
After criterion 3	18,925	1.5	17,577	1.2	22,263	1.2
After criterion 4	14,992	1.2	14,005	1.0	18,834	1.0
After criterion 5	13,324	1.0	11,634	0.8	14,952	0.8

Table 1: The number and percentage of events left after applying the selection criteria, given for different beam energies. Since criterion 1 is only originally pointed at preventing partially contained showers, it can be useful to look at the percentage of events left with respect to the number of events left after applying criterion 1. This is done in table 3 in appendix C.

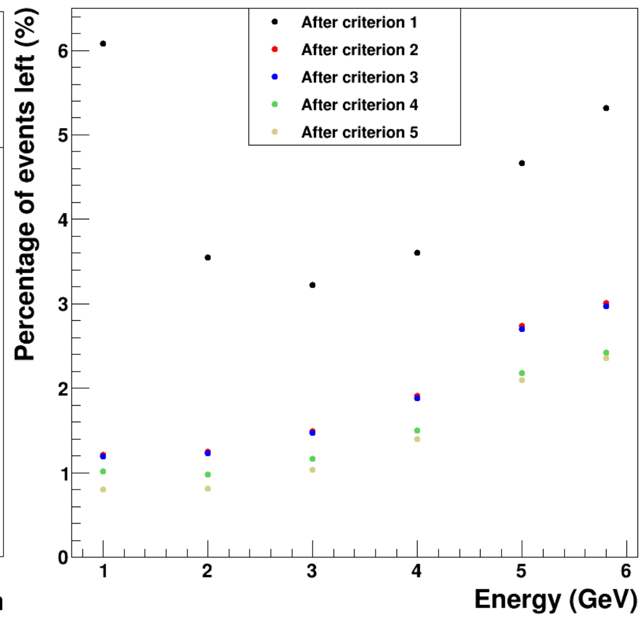
Table 1 shows the number and percentage of events left after applying the criteria. This data is represented in figure 14 in two different ways, plotting the energies separately (a) or plotting the effects of each criterion separately (b).

5.1 Distribution of the number of hits

The total number of measured hits in the detector is expected to be proportional to the energy of the incoming particle. Because of the stochastic nature of the shower development, we expect the number of hits per event containing one electron to fluctuate. Because the number of individual microscopic processes involved in every shower is large, we expect it to be normally distributed. However, an event with two incoming particles will have approximately twice as many hits, so events with two electrons will also be normally distributed, but with a mean which is approximately twice as large. That means that in the ideal case, the total distribution of hits per event will be the sum of several Gaussian curves. In figure 15 the distribution of the number of hits per event in run 1314 is shown. The energy of the incoming particles was 5 GeV. There are at least 3 peaks visible in this distribution. The first peak, around 1400 hits per event, is expected to be caused by events with one electron, the second peak, around 2700 hits per event, is expected to be caused by events with two electrons, and the third peak, around 4000 hits per event, is expected to be caused by events with three electrons. This distribution is a good measure for the number of electrons in the selected events. After every criterion, the distribution of the number of hits is checked, to



(a)



(b)

Figure 14: The percentage of events left after applying the selection criteria, plotting the energies separately (a) or plotting the effects of each criterion separately (b).

see how effective it is. Ideally, the final set of events will show only one Gaussian distribution with a mean of around 1400 hits per event.

To make these checks quantitative, the data will be compared using two different methods. Firstly, an asymmetric Gaussian curve will be fitted to the data. This fit has four free parameters, the height of the curve, the mean of the curve, the standard deviation σ_{L} on the left side of the mean and the standard deviation σ_{R} on the right side of the mean. An asymmetric Gaussian is used rather then a symmetric Gaussian, because it will show if there are more events on the right side of the peak, compared to the left side of the peak. This check tells something about the amount of events with a second shower which do not have a very high number of hits. This can happen if for example the second shower is just partially contained.

For the second check, an asymmetric Gaussian curve is fitted to the first peak of the raw data, visible in figure 15. Then the average standard deviation, so $\bar{\sigma} = \frac{\sigma_{\mathrm{L}}+\sigma_{\mathrm{R}}}{2}$ is calculated. This is equal to 158. After every criterion, the amount of events N_{E1} with hits within an interval spanning from $3\bar{\sigma}$ left of the mean to $3\bar{\sigma}$ right of the mean, $[882, 1830]$, are counted. Also the amount of events N_{E2} with hits within an interval spanning from $3\bar{\sigma}$ left of twice the mean to $3\bar{\sigma}$ right of twice the mean, $[2238, 3186]$, are counted. These intervals are indicated in figure 15. These two numbers are divided, to get

$$f_N = \frac{N_{E1}}{N_{E2}}. \tag{24}$$

This number f_N tells something about the amount of events with two electrons with a really high number of hits. In the distribution of the number of hits before any selection this num-

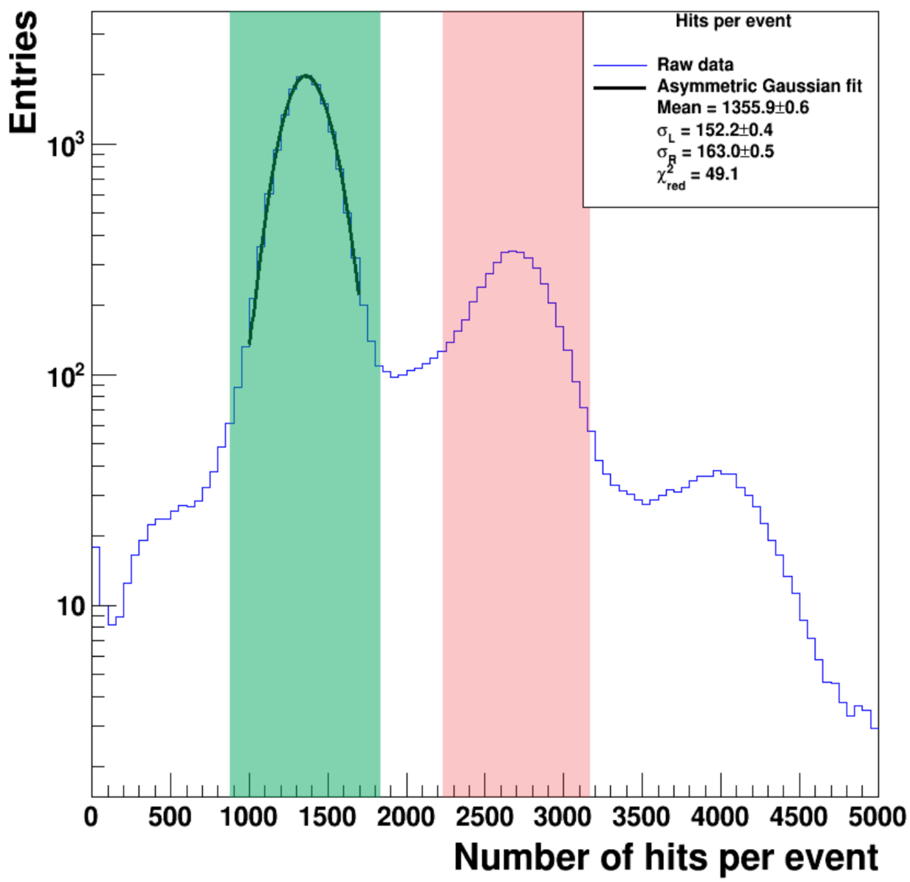


Figure 15: The distribution of the number of hits per event of run 1413 with incoming electrons with an energy of 5 GeV. An asymmetric Gaussian fit is done on part of the first peak. This is used to find the intervals covering roughly the first peak (green) and roughly the second peak (red). The amount of events in the green interval is N_{E1}, the amount of events in the red interval is N_{E2}.

ber f_N is approximately equal to 4.

The asymmetric Gaussian fit which is used to find the intervals for the second check is done on the raw data. Because of this, the mean can deviate from the 'real' mean number of hits per event with one electron. The used range covers 3σ left and 3σ right of the mean, so the effect of this deviation will not be very large. Furthermore, this is just meant to get a quantitative measure of the removal of pileup, it is not a problem if this does not completely accurately cover the peaks.
Both of these checks will only be applied on run 1413.

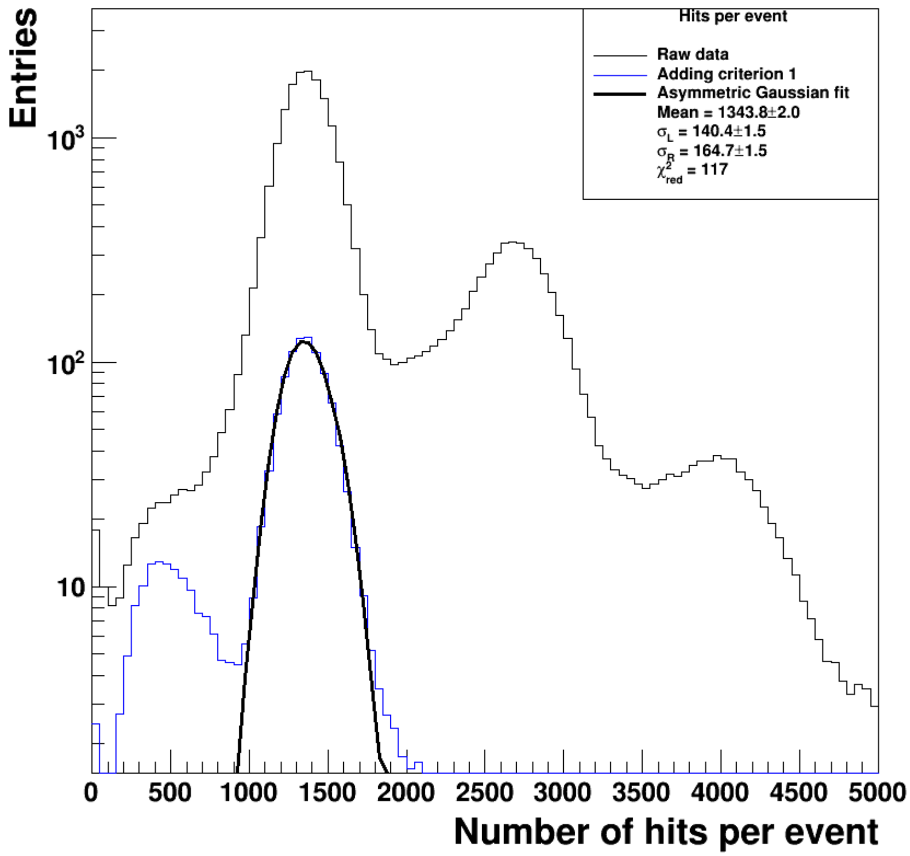


Figure 16: The distribution of the number of hits before and after applying the first criterion with run 1413 (5 GeV). An asymmetric Gaussian fit is applied.

5.2 Effectiveness of criterion 1

The first criterion only selects events with no hits outside of the central 124 × 125 pixels in the first layer. The effect of this criterion on the distribution of the number of hits is shown in figure 16. Firstly one can notice that this actually already rejects a lot of events with a high number of hits, so events with most likely more than one incoming electron. This can be seen by the number f_N, which is approximately equal to 70 after criterion 1 is applied. This is approximately 18 times as much as before. Not all of the particles are fired towards the central part of the detector, there is a spread in the beam. It turns out that this spread is quite large. That means that if there are two or more particles in the event, the probability is considerably high that at least one of them will not arrive in the central part of the detector. For that reason most events with more electrons will not be accepted by this criterion. Furthermore, if two particles do end up very close to each other, their hits will partly overlap, so the number of measured hits will be a bit smaller than expected.

The second interesting feature is the smaller peak around 500 hits per event. The events with a small number of hits are expected to be caused by partly contained showers, so this criterion should reject those events from the selected set of events. It turns out that from all of the events with less than 900 hits that are left after applying this criterion in run 1413, 98% have no hits in the first layer, these showers only start developing later. That means that the second selection criterion should remove these events from the selected set.

The asymmetric Gaussian fit is of course not really reliable due to the peak around 500 hits per event, this can be seen in the reduced χ^2 of this fit, it is much larger then 1.
In table 1 the number of events left after applying each criterion is shown. The first layer contains 1024×1024 pixels. This criterion only accepts events with hits in the central 125×124 pixels. The insensitive gap between the two chips is estimated to be 5 pixels, so the actual number of pixels that are allowed to have hits are 120×124. This means that this acceptance area contains 1.4% of the pixels of the whole layer. If the probability of an electron hitting a certain pixel would be the same for every pixel, if every electron would create exactly 1 hit in the first layer and if every event contains only 1 electron, the fraction of events left would be expected to be 1.4%. These last two assumptions are not realistic, with the result that the expected fraction of events would be even smaller then 1.4%. Table 1 shows that the actual fraction of events left after applying the first selection criterion is much higher, confirming that the beam was actually focused at the central part of the layer, and has a Gaussian-like profile. The spread of this Gaussian depends on the beam energy, as can be seen in figure 14 (b).

The good events that are rejected by this criterion are events where the shower, created by the particle, is fully contained within the detector, but where the particle enters the detector further away from the center than allowed by this criterion. Other events that are rejected unintentionally are events where there are noise hits outside of the acceptance area. Both properties (the occurrence of noise and the location of incidence) are not (relevant) properties of the electrons, so this criterion will probably not create a bias.

The area of 120×124 pixels was varied, to see what area would be the best. Areas of 181×181, 124×125, 100×101 and 76×77 pixels were tested. In the end the area of 124×125 was chosen, because with this area the criterion leaves enough events to do analysis with, but the distance between the acceptance area and the edge of the layer is large enough to allow the shower to be almost completely contained.

5.3 Effectiveness of criterion 2

The second criterion selects events based on the number of clusters in the first layer with hits behind them in the second layer. The effect of this criterion on the distribution of the number of hits is shown in figure 17. As expected almost all events from the left side of the distribution are removed, since events with no hits in the first layer are rejected now. It is also visible that there are events removed from the right side of the distribution. This is also visible in f_N, which increases from 70 to 97. That means that this criterion is effective for removing events with more than one electron. The Gaussian fit in figure 17 still describes the data not very precisely, but the χ^2_{red} already improved a lot, from 117 to 14.8.

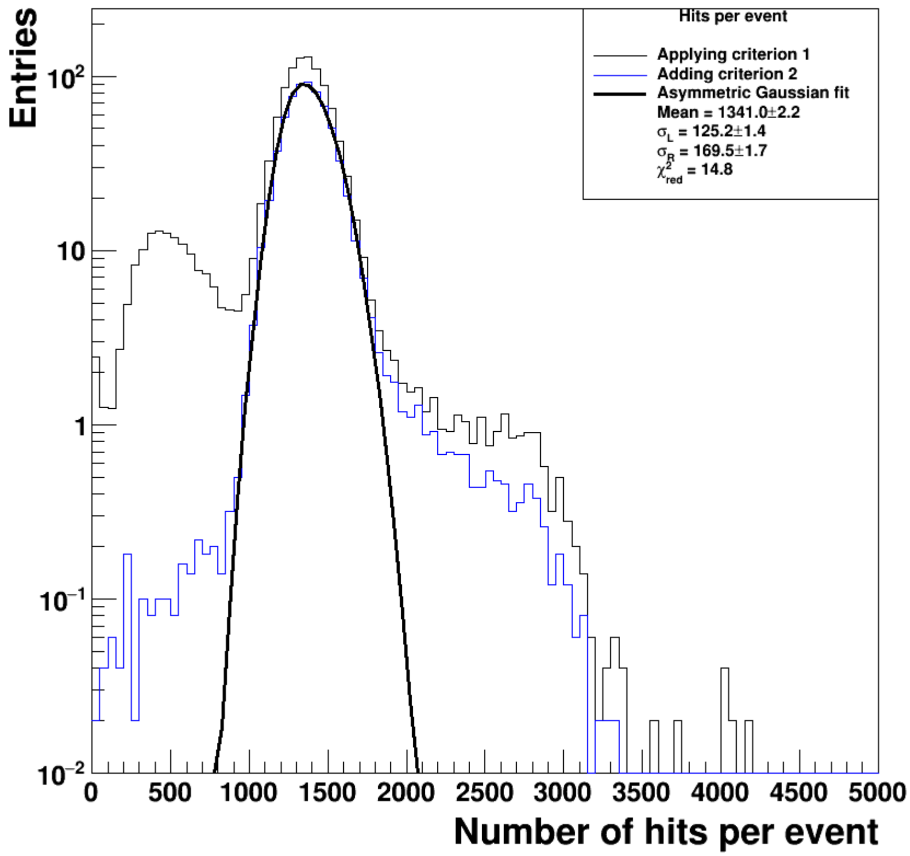


Figure 17: The distribution of the number of hits before and after applying the second criterion with run 1413 (5 GeV). An asymmetric Gaussian fit is applied.

Some events of the one electron peak are also removed. This can of course be events with multiple electrons which have less hits than expected, but this will also contain actually wanted events. That means that there could be a bias entering here. The good events which are rejected can be sorted into three different categories.

A. **A cluster in the first layer with no hits behind it in the second layer.** The occurrence of these events is very unlikely. If a particle creates a hit in the first layer, it will probably start developing a shower in the next layer of tungsten, the probability of all of the particles traveling more than 10 pixels away is very small. If the shower does not start yet, the particle will approximately continue traveling in a straight line parallel to the beam axis, which is perpendicular to the layers, so then the particle will not travel more than 10 pixels away as well. If the number of electrons which shows this behaviour is substantially, this is causing a bias, but based on theory it is save to say that these showers are very rare.

Another cause of this behaviour would be if a primary particle or all of the secondary particles go through the second layer without creating a hit. If the assumption is made that for every particle going through the silicon chips the probability of being detected is equal, this would create a bias towards early developing showers. If a shower starts developing earlier there are more particles going through the second layer, so the probability of no hit occurring there would be smaller than if there is no shower yet and just one particle goes through the second layer. For now this effect is very small, but it turns out that on another place in this criterion (C) this bias towards earlier developing showers gets even stronger. This is very important to keep in mind. It does not mean that this selection is not useful, but is is important to keep this in mind when using this selection for analysis.

B. **More than one cluster in the first layer.** One particle creating more than one cluster in the first layer is very uncommon. If this would happen, it would be due to unusual behavior of the detector itself and not due to a certain property of the particle, so this will not create a bias.

C. **No hits in the first layer.** It is possible that a particle does not create a hit in the first layer, but only starts causing hits later. Events have been observed where the shower development starts in later layers, in rare cases the shower started only in layer 5. This enforces the bias towards early developing showers, discussed before.

Several search areas were considered: square search areas of 5×5, 7×7, 9×9, 11×11, 13×13 and 15×15 pixels and circular search areas with radii of 7, 8, 9 and 10 pixels. The square search areas were first tested. The search area of 15×15 pixels gave the best selection. Afterwards the shape of the search area was changed to a circular one, because that fits the radial symmetry of the spread of particles. Out of the tested areas the one with 10 pixels gave the most similar result to the results with a 15×15 square search area. Furthermore, two definitions of a cluster were tested. In the first definition even a single separate hit is considered a cluster, in the second definition a cluster is a group of at least two horizontally, vertically or diagonally adjacent hits. The second definition failed to recognize a lot of particles, because many particles only leave one hit in the first layer. This resulted in many events with two or more electrons being categorized as a one electron event. This happened much less frequently with the first definition of a cluster, so that is the best definition to use.

5.4 Effectiveness of criterion 3

The third criterion rejects events with ignored clusters in the first layer which contain more then one hit. Figure 18 shows the distribution of the number of hits before and after applying the third criterion. This figure, as well as table 1, shows that this criterion does actually reject only a small portion of events. The good events that could get rejected are very similar to category B discussed in the last section. Combining this information with the low number of rejected events leads to the conclusion that this selection criterion does not cause a bias. The low number of rejected events has as a consequence that f_N improves just slightly, it goes from 97 to 98.

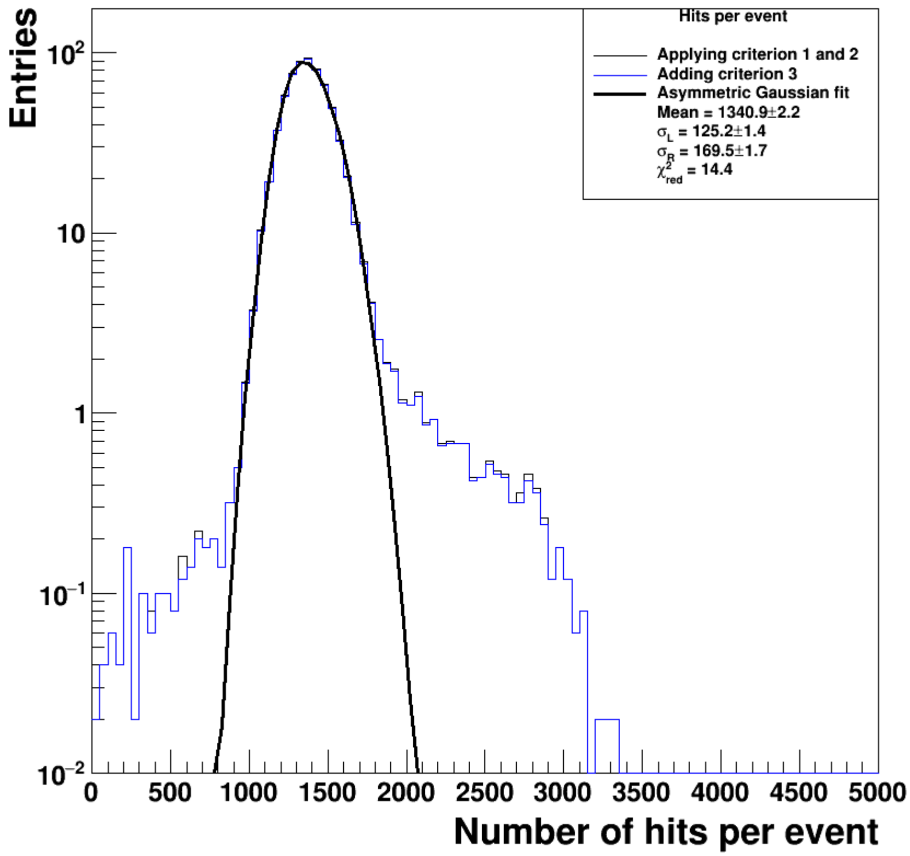


Figure 18: The distribution of the number of hits before and after applying the third criterion with run 1413 (5 GeV). An asymmetric Gaussian fit is applied.

5.5 Effectiveness of criterion 4

The fourth criterion rejects events with hits in the second layer which are further then 120 pixels away from the shower core. The effect of the fourth criterion on the distribution of the number of hits is shown in figure 19. This visibly removes a large portion of the right hand side tail of the peak. This can also very clearly be seen in f_N, which goes from 97 to 301. The χ^2_{red} of the fit also improves, it goes from 14.4 to 6.07. Good events which are rejected by this criterion have hits in the second layer which are further than 120 pixels away. This creates the possibility for getting a bias towards showers which do not expand quickly in the radial direction. For that reason the maximal distance between the location of a hit and the location of the shower center is made relatively large, 120 pixels (approximately 3.5 mm). The fact that the search area with a radius of 10 pixels for the second selection criterion is already large enough to select most one electron events, shows that the secondary particles are still relatively close to the shower center in the second layer, so 120 pixels provides a large enough margin. Still, it is good to keep in mind that such a bias can be here. When investigating the lateral shower development it might for example be useful to check the shower

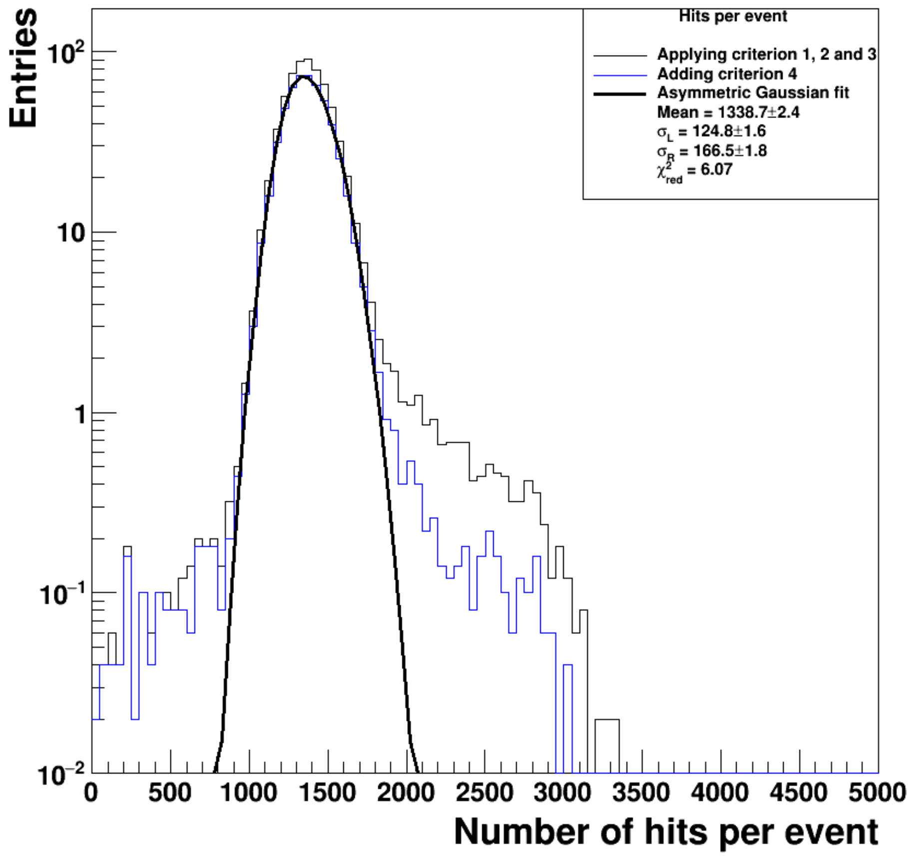


Figure 19: The distribution of the number of hits before and after applying the fourth criterion with run 1413 (5 GeV). An asymmetric Gaussian fit is applied.

profile with and without this selection applied to see if there are significant differences.

Several distances were considered: 80, 100 and 120 pixels. These variations did not have a lot of influence, so the distance of 120 pixels was chosen, since this leaves the most events for analysis and has the least chance of creating a bias.

5.6 Effectiveness of criterion 5

Criterion 5 rejects events which have more then 0.15 times as much hits in the border region of layers 3 till 10 as in the central region in the same layers. The effect of criterion 5 on the distribution of the number of hits is shown in figure 20. It can be seen that not many events are rejected by this criterion. Many of the rejected events appear in the right hand side of the distribution however, so despite the low amount of rejection it is still a useful selection criterion. This is reflected in f_N, which increases from 301 to 379. The distribution also gets more of an (asymmetric) Gaussian shape, which can be seen by the χ^2_{red}, which increases from 6.07 to 3.94. Good events which could be rejected by this criterion are events which

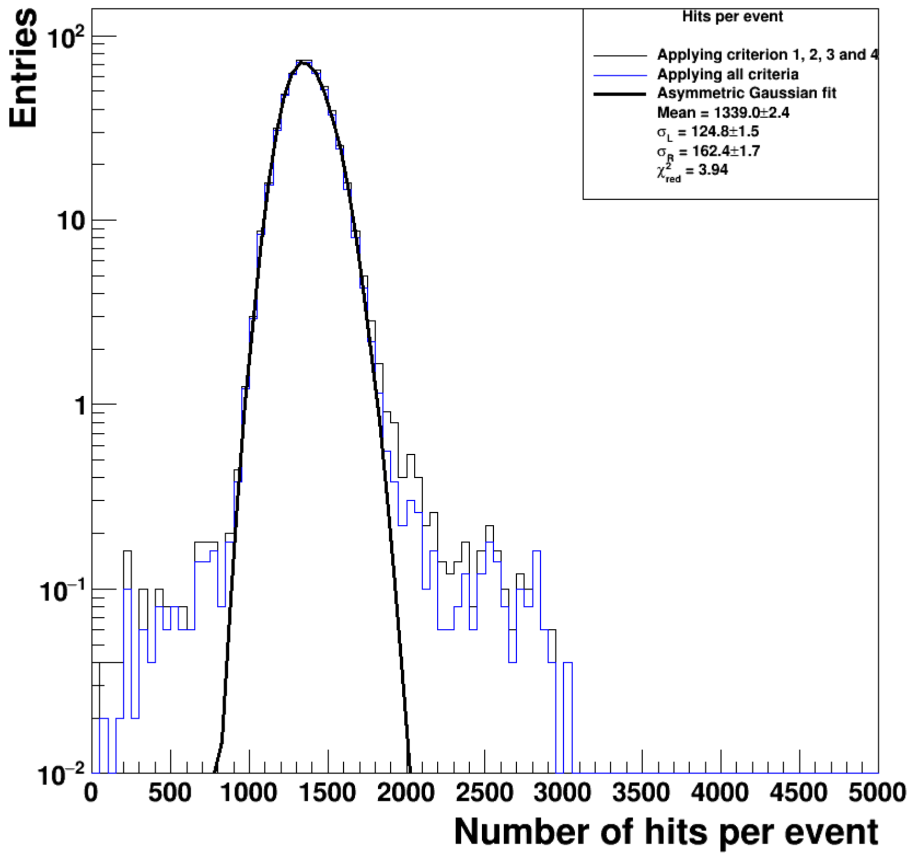
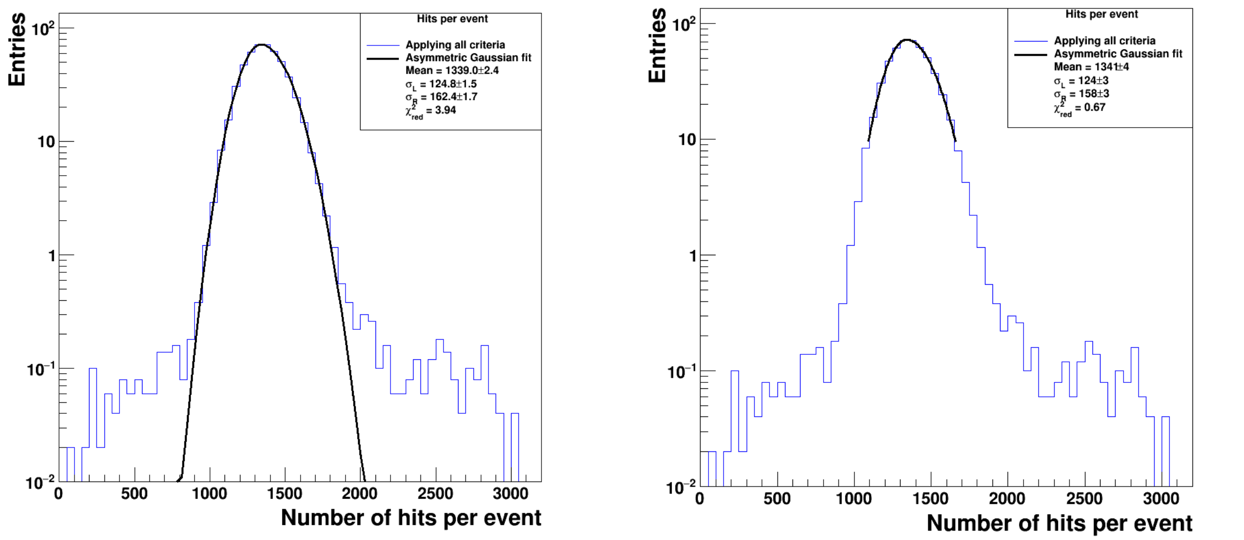


Figure 20: The distribution of the number of hits before and after applying the fifth criterion with run 1413 (5 GeV). An asymmetric Gaussian fit is applied.

spread out very far, such that they also have many hits close to the border of the layer. This would create a bias against events with a very broad shower. However, the probability of those events occurring is very small, because even if the shower is very broad, there are still many more hits in the center, so the fraction would still be small. The shower gets really broad in later layers, but only the third layer until the tenth layer is taken into account, so the layers where the shower is very broad are not used in this counting process. For this reason one can assume that this bias is not very strong.

This criterion only works if there is a clear separation between the inner circle and the border region. If the cluster in the first layer would lie on the edge of the allowed region of criterion 1, it would be 450 pixels away from the border of the layer. The circle in which hits are counted has a radius of 80 pixels, so this circle cannot come closer than 370 pixels to the edge of the chip. The border region is 170 pixels wide. That means that the distance between the two counting areas is always at least 200 pixels (approximately 6 mm), which shows that the two counting areas are always very well separated.



(a) The data is fitted with an asymmetric Gaussian over the whole distribution. The p-value of the χ^2 of this fit is almost equal to 0.

(b) The data is fitted with an asymmetric Gaussian over only $2\sigma_{\mathrm{L}}$ to the left and $2\sigma_{\mathrm{R}}$ right. The p-value of the χ^2 of this fit is 70%.

Figure 21: The distribution of the number of hits after applying all of the criteria, with incoming electrons with an energy of 5 GeV, run 1413.

Many sizes of the border and the center circular region were tested. In the end a circle with a radius of 80 pixels was chosen, and a border region which is 170 pixels wide. This choice was based on visual clues from *hit maps*. A hit map shows the location of hits in the detector. Hit maps of events in the right hand tail of the distribution of number of hits were checked, and this showed that most of the two electron events had a fraction $f > 0.15$. Because of this, the critical fraction f_c was set to be 0.15.

5.7 Quality of the final event set

In figure 21 (a) the distribution of the number of hits after applying all of the criteria is shown, together with an asymmetric Gaussian fit. The Gaussian fit shows an asymmetry in the data, σ_{L} is much smaller than σ_{R}. After fitting the whole curve, a second fit is made to only part of the data, it goes $2\sigma_{\mathrm{L}}$ to the left of the mean and $2\sigma_{\mathrm{R}}$ to the right of the mean. This is depicted in figure 21 (b). This reducing of the fit range does not change the asymmetry a lot, so the asymmetry is not just caused by the visible tail to the right, the effect is also apparent close to the maximum of the peak. The χ^2_{red} of the fit gets better however, it goes from 3.94 to 0.67. The p-value of the χ^2 of the fit over the full data is extremely small, almost 0, because that fit has 54 degrees of freedom. The fit over only part of the data has only 7 degrees of freedom. The p-value of the corresponding χ^2 is equal to 70%, so the fitted function properly describes the data. The asymmetry in the peak is also present at other energies. For every energy the run with the most data points is used to make similar fits, the results are shown in table 2. The distributions of the number of hits can be found in

	Full range			
	Mean	**Left σ**	**Right σ**	χ^2_{red}
5.8 GeV	1548 ± 3	140.3 ± 2.2	165.8 ± 2.4	2.04
5.0 GeV	1339.0 ± 2.4	124.8 ± 1.5	162.4 ± 1.7	3.94
4.0 GeV	1081 ± 3	111.8 ± 2.2	151.8 ± 2.5	2.36
3.0 GeV	814 ± 4	98.2 ± 2.1	129.6 ± 2.5	2.74
2.0 GeV	554 ± 4	80 ± 3	111 ± 3	2.68
1.0 GeV	257.5 ± 2.3	49.9 ± 1.5	87.2 ± 1.8	2.68

	2σ left and right			
	Mean	**Left σ**	**Right σ**	χ^2_{red}
5.8 GeV	1553 ± 4	141 ± 3	158 ± 3	1.18
5.0 GeV	1341 ± 4	124 ± 3	158 ± 3	0.67
4.0 GeV	1086 ± 4	115 ± 4	146 ± 4	1.02
3.0 GeV	822 ± 5	103 ± 4	121 ± 4	0.38
2.0 GeV	563 ± 7	98 ± 7	95 ± 5	1.59
1.0 GeV	261 ± 3	51.8 ± 2.5	81.8 ± 2.4	3.48

Table 2: The mean, left σ, right σ and χ^2_{red} for fits on the distributions of the number of hits after applying all of the criteria, fitting on the whole range and on just a sub range, given for different beam energies.

appendix B.

This asymmetry might mean that there is still some pileup in the events after this selection. At this point it gets very hard to discriminate between events with one electron and events with more electrons based on the location of hits in the detector. The events which still suffer from pileup have electrons which enter the detector at a small spacial distance to each other. Most times the second shower starts developing in a later layer, while the first shower is already starting to broaden. It is very hard, maybe even impossible, to get a general criterion which locates these showers, since it could also be just a single shower with more hits than average. A criterion based on a more advanced shower reconstruction could improve the results, but that would still leave some events with pileup in the selected event sets.

It could be possible that the asymmetry is not only produced by pileup but due to an asymmetry in the electron beam energy. It is assumed that the electrons entering the detector all have the same energy, but this is not completely true. If there are slightly more electrons with a higher energy than with a lower energy, this would result in an asymmetry in the distribution of the number of hits. The possibility of an asymmetry in the beam energy is being investigated.

5.8 Possible improvements

In order to improve the data quality further, the following improvements to the selection criteria and in the data taking can be considered.

Alignment In the research presented in this thesis it is assumed that the layers are stacked on top of each other perfectly. In reality this is not the case, the layers may be shifted and rotated with respect to each other. The procedure of finding out how much the layers are shifted and rotated with respect to each other is called *alignment*. This is done by measuring cosmic muons. When particles from space enter the atmosphere of the earth, they collide with atoms. As a consequence of this, the particle will decay into several particles, among which muons [28]. If these muons go through the detector they will most likely not decay and produce no shower, so they leave a straight line. By analysing these straight lines it is possible to track down how the layers are shifted and rotated. This procedure can also be used to get a better estimate of the size of the insensitive gap between the two ALPIDE chips. For this research this gap is estimated to be 5 pixels wide. When the detector is measuring cosmic muons, it is turned facing upwards, to catch as many muons as possible. There is one scintillator in front of the detector and one at the back. The data will only be used if a trigger arrives from the front as well as from the back scintillator, to make sure the muon goes through all layers of the detector. To get enough events, the measurements of muons must continue for several months due to low count rates. These measurements are ongoing. Once enough data is gathered, this alignment can be done, which will help with the improvement of the selection methods.

Pixel masking *Pixel masking* is the disabling of pixels which send a signal even if they are not hit by a particle (see section 4.2.3). A pixel mask has been applied, but this pixel mask needs some refinement. This will especially help with selection criteria 1 and 4, since one noise hit further than 125 pixels away from the central part of the first layer or 120 pixels away from the shower center in the second layer will immediately get the event rejected.

Search in later layers if no hit in first layer As discussed in section 5.3, criterion 2 creates a bias towards events with early developing showers. This bias can be reduced by extending the search to later layers if there is no hit found in the first layer. Criterion 1 can then be applied on layer 2, criterion 2 on layer 2 and 3, criterion 3 on layer 2, criterion 4 on layer 3 and criterion 5 on layer 4-11. If there is also no hit in layer 3 this can even be extended to later layers similarly. This reduces the bias towards early developing showers. If this improvement is applied, one could also label hits with the layer number relative to the first layer with hits instead of the absolute layer number. This way the longitudinal profile visualises the depth with respect to the start of the shower instead of the depth with respect to the start of the detector.

Reducing pileup in later test measurements It would be ideal if there would not be so much pileup in the event set, that would reduce the need for very strong selection criteria. One way to achieve that is by using a smaller collimator. A *collimator* is a heavy metal object with a rectangular hole in the center. It is placed in front of the electron beam, to narrow it. During the measurements used in this thesis, a lead collimator of 12×12 mm^2 was used. If a smaller collimator is used, less electrons will remain in the beam, so the probability of pileup occurring is reduced.

Improve trigger setup At this moment, when the readout board receives a trigger from the SiPM, the readout is started and further triggers are ignored for a while. If a second electron arrives soon after that, the SiPM gives another signal. If that second signal would also be recorded by the readout board, it could cancel the readout, or give the event a label which warns that there might be pileup occurring. This way, identifying pileup does not only involve geometry, but also time. This makes it more easy to distinguish events with and without pileup.

6 Conclusion

During test measurements of the mTower electromagnetic calorimeter prototype it turned out that there was a high amount of pileup. There was no time information recorded, so another pileup removal method has been developed, based on geometry. Five criteria have been developed to select events (measurements) with only one electron entering the detector. The amount of pileup in the events is drastically reduced, but there is still an asymmetry visible in the distribution of the number of hits, as shown in table 2. This asymmetry is not only caused by a broad tail towards the right in the distributions of the number of hits, but is also present close to the mean of the Gaussian. Development of extra selection criteria might improve this result, but it is expected that new criteria would not improve the results drastically. This means that it is possible to make selection criteria based on the location of hits in the detector, but that this is likely not sufficient to remove the pileup all together. The best way to reduce the pileup even more is to lower the electron frequency of the beam or to record the triggers from the scintillators to have a second source of information to detect pileup. Since these improvements can not be applied to the current set of data, this selection of events can be used for analysis of this data set.

References

[1] CERN. *The Large Hadron Collider*. `https://home.cern/science/accelerators/large-hadron-collider`. retrieved May 8, 2020.

[2] CERN. *A new schedule for the LHC and its successor*. `https://home.cern/news/news/accelerators/new-schedule-lhc-and-its-successor`. published December 13, 2019, retrieved June 8, 2020.

[3] N. van der Kolk. "FoCal: A highly granular digital calorimeter". In: *Nucl. Instrum. Methods Phys. Res. A* 958 (Apr. 2020). DOI: `10.1016/j.nima.2019.04.013`.

[4] B. Terzić. *PHYS 652: Astrophysics Lecture 13*.

[5] C. Lippmann. *Particle identification in ALICE boosts QGP studies*. `https://cerncourier.com/a/particle-identification-in-alice-boosts-qgp-studies/`. published August 23, 2012, retrieved May 9, 2020.

[6] CERN. *ALICE - Experiment*. `http://alice.web.cern.ch/`. retrieved June 9, 2020.

[7] A. Tauro. *ALICE Schematics as during RUN3 (after upgrade)*. `https://cds.cern.ch/record/2263642`. published May 11, 2017, retrieved June 8, 2020.

[8] J. Feltesse. "Introduction to Parton Distribution Functions". In: *Scholarpedia* 5.11 (2010). revision #186761, p. 10160. DOI: 10.4249/scholarpedia.10160.

[9] R. Plačakytė. "Parton Distribution Functions". In: *XXXI PHYSICS IN COLLISION*. 2011. arXiv: 1111.5452 [hep-ph].

[10] T. Peitzmann. "Measurement of forward direct photon production in p-A at the LHC with ALICE - A probe for nuclear PDFs and saturation". In: *Proceedings of Science*. XXIV International Workshop on Deep-Inelastic Scattering and Related Subjects. 2016. arXiv: 1607.01673 [hep-ex].

[11] K.J. Eskola et al. "EPPS16: nuclear parton distributions with LHC data". In: *Eur. Phys. J. C* 77.163 (Mar. 2017). DOI: 10.1140/epjc/s10052-017-4725-9.

[12] L. McLerran and R. Venugopalan. "Gluon distribution functions for very large nuclei at small transverse momentum". In: *Phys. Rev. D* 49 (7 Apr. 1994), pp. 3352–3355. DOI: 10.1103/PhysRevD.49.3352. URL: https://link.aps.org/doi/10.1103/PhysRevD.49.3352.

[13] T. Peitzmann. "Forward photon measurements in ALICE as a probe for low-x gluons". In: *Proceedings of Science*. International Conference on Hard and Electromagnetic Probes of High-Energy Nuclear Collisions. 2018. DOI: 10.22323/1.345.0062. arXiv: 1812.07246 [hep-ex].

[14] C. Leroy and P.G. Rancoita. "Physics of cascading shower generation and propagation in matter: principles of high-energy, ultrahigh-energy and compensating calorimetry". In: *Rep. Prog. Phys.* 63.4 (Mar. 2000), pp. 505–606. DOI: 10.1088/0034-4885/63/4/202.

[15] E. Haug and W. Nakel. *The Elementary Process of Bremsstrahlung*. World Scientific Lecture Notes in Physics 73. World Scientific Publishing Company, Mar. 2004.

[16] A. Das and T. Ferbel. *Introduction to Nuclear and Particle Physics*. 2nd ed. World Scientific Publishing Company, Dec. 2003.

[17] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[18] S. Weishi Li and J.F. Beacom. *Spallation Backgrounds in Super-Kamiokande Are Made in Muon-Induced Showers*. 2015. arXiv: 1503.04823 [hep-ph].

[19] E. Longo and I. Sestili. "Monte Carlo calculation of photon-initiated electromagnetic showers in lead glass". In: *Nucl. Instrum. Methods.* 128.2 (1975), pp. 283–307. ISSN: 0029-554X. DOI: 10.1016/0029-554X(75)90679-5. URL: http://www.sciencedirect.com/science/article/pii/0029554X75906795.

[20] E. Garutti. *Calorimeters - Energy measurement*. http://www.desy.de/~garutti/LECTURES/ParticleDetectorSS12/L10_Calorimetry.pdf. retrieved June 12, 2020.

[21] C.W. Fabjan and F. Gianotti. "Calorimetry for particle physics". In: *Rev. Mod. Phys.* 75 (Oct. 2003), pp. 1243–1286. DOI: 10.1103/RevModPhys.75.1243.

[22] D. Groom. *Atomic and nuclear properties of tungsten (W)*. http://pdg.lbl.gov/2019/AtomicNuclearProperties/HTML/tungsten_W.html. Aug. 2018.

[23] SensL. *An Introduction to the Silicon Photomultiplier*. `https://www.sensl.com/downloads/ds/TN%20-%20Intro%20to%20SPM%20Tech.pdf`. published February 2017, retrieved June 12, 2020.

[24] G.A. Rinella. "The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System". In: *Nucl. Instrum. Methods Phys. Res. A* 845 (Feb. 2017), pp. 583–587. DOI: `10.1016/j.nima.2016.05.016`.

[25] B. Abelev *et al* and The ALICE Collaboration. "Technical Design Report for the Upgrade of the ALICE Inner Tracking System". In: *J. Phys. G* 41.8 (July 2014), p. 087002. DOI: `10.1088/0954-3899/41/8/087002`.

[26] DESY. *Storage Rings as Synchrotron Radiation Sources*. `https://photon-science.desy.de/research/students__teaching/primers/storage_rings__beamlines/`. retrieved June 11, 2020.

[27] R. Diener et al. "The DESY II test beam facility". In: *Nucl. Instrum. Methods Phys. Res. A* 922 (Apr. 2019), pp. 265–286. ISSN: 0168-9002. DOI: `10.1016/j.nima.2018.11.133`. URL: `http://dx.doi.org/10.1016/j.nima.2018.11.133`.

[28] G. Bonomi et al. "Applications of cosmic-ray muons". In: *Prog. Part. Nucl. Phys* 112 (May 2020). DOI: `10.1016/j.ppnp.2020.103768`.

7 Laymen summary (in Dutch)

De kleinste bouwstenen van ons universum heten *quarks*. Deze quarks verzamelen zich bijna altijd in groepjes van 2 of 3. Over de oorzaak hiervan is nog veel onduidelijk. Om hier meer over te ontdekken, laten wetenschappers hele kleine deeltjes hard op elkaar botsen, in zogenaamde *deeltjesversnellers*. De grootste deeltjesversneller van de wereld staat in Zwitserland, de *LHC*. De deeltjes botsen daar zo hard op elkaar dat die groepjes van quarks worden opgesplitst. Tijdens deze botsing ontstaan er allerlei nieuwe deeltjes, waaronder *fotonen*. Als we deze fotonen opvangen en onderzoeken kunnen we meer leren over waarom deze quarks bij elkaar blijven, en wat de structuur van hele kleine materie is.

Om deze fotonen te onderzoeken wordt er een nieuw apparaat ontwikkeld, een *calorimeter*. Er is een prototype gebouwd van dit apparaat, een kleine variant van de uiteindelijke calorimeter. Hier hebben we in februari 2020 metingen aan gedaan, ik heb geholpen met de metingen en de data-verwerking. Ik zal nu eerst uitleggen hoe deze calorimeter werkt, daarna vertel ik mijn bijdrage aan het onderzoek.

Als een foton de calorimeter binnengaat komt deze eerst door een *detectorlaag*. Dat is een laag die bestaat uit heel veel pixels, in dit kleinere prototype al meer dan 1 miljoen. Als een foton door een pixel gaat, wordt dit gemeten. Hierdoor weten we precies waar het foton het apparaat binnenkomt. Na deze detectorlaag gaat het foton door een laag van ongeveer 3 mm wolfraam, dat is een metaalsoort. Als je met het blote oog naar een metaal kijkt lijkt het een vast voorwerp, maar als je heel erg zou inzoomen ontdek je dat het eigenlijk allemaal losse *atomen* zijn die gestructureerd zijn in een soort rooster. Het foton kan in eerste instantie langs de atomen vliegen, maar vroeg of laat zal deze een keer botsen met een atoom. Als dat gebeurt splitst het foton in 2 nieuwe deeltjes, een elektron en een positron. Na dit laagje wolfraam, komt er weer een detectorlaag, waar we kunnen zien waar deze 2 deeltjes nu zijn. Daarna komen de twee deeltjes weer in een laag wolfraam, waar ze opnieuw opslitsen, er ontstaan weer fotonen. Het dan dus al 4 deeltjes. Dit gaat heel lang zo door, in totaal zijn er 24 detector- en wolfraamlagen. Elke keer als een deeltje splitst, wordt ook de *energie* van het deeltje gesplitst. Uiteindelijk is er niet genoeg energie meer over om nog verder te splitsen, dan ontstaan er geen nieuwe deeltjes meer.

Een belangrijke eigenschap van fotonen is hun energie. Hoe groter die energie is, hoe vaker het foton zich kan opslitsen. Om te ontdekken hoeveel energie dit foton heeft, hoeven we dus alleen maar te kijken naar hoeveel deeltjes er ontstaan, en hoe ver ze in de detector komen. Op deze manier kan de calorimeter dus de energie van het foton meten.

In februari 2020 hebben we testmetingen gedaan met het prototype. Tijdens dit onderzoek werden er elektronen afgeschoten in de richting van het prototype, om te kijken wat er dan gebeurt. We hebben in plaats van fotonen dus *elektronen* gebruikt, maar dat maakt voor het effect niet uit. Ik zal nu toelichten wat mijn bijdrage is aan de analyse van de data die we verzameld hebben.

Als er een elektron aankomt bij het prototype, wordt er voor een korte tijd gemeten in de eerder besproken detectorlagen. Het kan echter zo zijn dat er in die korte tijd nog een elektron in de detector binnenkomt, of zelfs nog meer. Mijn doel was het sorteren van metingen, zodat er alleen nog maar metingen met één elektron overblijven. Dit kan natuurlijk met de hand, maar dat is heel veel werk, er zijn miljoenen metingen gedaan. Daarom is dit geautomatiseerd. Hiervoor heb ik een aantal criteria opgesteld die gebruikt kunnen worden om metingen te sorteren. Alleen metingen die aan alle criteria voldoen worden gebruikt bij de verdere verwerking.

Een simpel criterium dat ik gemaakt heb is tellen hoeveel deeltjes er in de eerste detectielaag te zien zijn. Dit is echter niet voldoende. Soms wordt een deeltje in de eerste laag niet gedetecteerd en soms lijkt het ook alsof er meer deeltjes zijn, terwijl dit in werkelijkheid niet zo is. Daarom zijn er nog meer criteria nodig. Er wordt ook in de latere detectielagen gekeken, informatie uit deze lagen wordt ook gebruikt om te bepalen hoeveel deeltjes er in werkelijkheid zijn. In totaal zijn er 4 van deze criteria opgesteld, waarbij er informatie wordt gebruikt uit de eerste 10 detectielagen van de detector.

Ik heb ook een ander type criterium toegepast. We willen alleen metingen gebruiken waarbij het elektron midden in de detector aangekomen is. Als een elektron aan de rand aankomt, kan het zijn dat deze na een tijdje uit de detector ‘ontsnapt’. Als dat gebeurt mist er een deel van de data, wat we meten is dan niet meer betrouwbaar. Daarom heb ik alleen metingen toegestaan waarbij de elektronen in het midden van de detector aankomen.

Deze selectiecriteria halen heel veel metingen met meerdere elektronen eruit, maar toch blijven er altijd nog metingen over die niet worden opgepikt. Dit zijn voornamelijk metingen waarbij de twee elektronen heel dicht bij elkaar zijn. Het lijkt er dus op dat deze manier van selecteren niet voldoende is om de metingen perfect van elkaar te scheiden. Daarom is het belangrijk dat er bij latere meetsessies maatregelen worden genomen zodat de selectie makkelijker wordt. Er kan bijvoorbeeld voor worden gezorgd dat er minder elektronen aankomen bij de detector. Ook zou er een externe teller kunnen worden aangebracht die telt hoeveel elektronen er langskomen.

A Code description

This is a description of the code, written for the event selection. The code is written in the programming language C++and uses the ROOT framework. All of the necessary code can be found by logging in on the quark cluster of Utrecht University and go to the directory of user 6092594. The directory 'finalCode' contains the code. If you have no access to the quark cluster you can contact Aart van Bochove (aartvanbochove@hotmail.com) to get access to the code.

A.1 mTower classes

The data gathered during the test measurements is stored in a ROOT TTree. To make this data more accessible, N. van der Kolk wrote four classes: mTowerHit, mTowerEvent, mTowerCluster and mTowerChip. mTowerHit() can store the properties of a hit, mTowerEvent() can store the properties of an event and mTowerCluster() can store the properties of a cluster. Two functions were added to mTowerCluster as part of this analysis, to calculate the mean row and the mean column of a cluster. mTowerChip() can store the properties of a chip, such as the hits and clusters in that chip. Furthermore, this class includes an algorithm to find the clusters in the chip. This is useful since that means that the clustering does not make the main code messy.

A.2 Masking files

The pixel masks applied in this analysis are given as different .txt files for different energies. They should be in the directory './masking/' with respect to the main code.

A.3 Analyse_mTower.C

This is the main code written for this project. It imports the TTree with the data, applies the criteria and does the analysis. This code can be run by typing './Analyse_mTower.sh [runnumber]' in the terminal. To be able to run this, the file locations of the classes should be changed in Analyse_mTower.sh. The contents of the different sections of the code in Analyse_mTower.C will be explained below.

A.3.1 Conversion tables

Every chip has been assigned several unique numbers, one of which is the lane number. These conversion tables are used to change between different indices for the lanes, to find the layer of a certain chip and contain some other useful conversions. Not all of these tables are used in the code, but they are added anyways for completeness sake.

A.3.2 Initial variables

This is a list of variables to set before starting the run. First some file locations must be set, then some booleans can activate and deactivate certain parts of the code. Note that criterion

1 described in this thesis is actually called C1ES (Criterion 1 Extra Strong) in the code. After this some properties of the criteria are initialized. Then there are these two integers:

```
int nParts = 1; //In how many parts is this run done?
int part = 1; //Which part is this?
```

There is a problem, sometimes a run suddenly stops working after a while. If this happens the problem can be solved by doing the run in separate parts. nParts sets the number of parts this run is split in, and part is the part which is running now. Usually splitting the run in 2 parts is enough, but if that does not help splitting it in 5 parts should be definitely enough. Unfortunately the problem is not found yet, but when the problem is solved this part of the code is not useful anymore and can be deleted.

After these two integers some properties of the mTower are entered.

A.3.3 Automatic variables

In this part some variables are set automatically based on the previously entered information. The TString beamEnergy is used for the masking input file. If the runnumber is not recognized, no masking will be applied.

A.3.4 Alternative coordinate system

The standard coordinates given in the raw data is (row, column, lane). For the developed criteria however it is more useful to have one coordinate system for each layer instead of one coordinate system for each chip. That is why all coordinates will be translated to (x, y, layer). This translation goes as follows:

```
int x = row;
int y = column;
if (IsLeftChip(lane+laneOffset)) //coordinate is in the left chip
  {
    x = -x - 1 - nPixelsGap;
    y = columnsPerChip - 1 - y;
  }
x += rowsPerChip + nPixelsGap;
```

This way (0,0) is situated at the outer corner of one of the chips. The integers maxX and maxY are the maximum values of x and y in this coordinate system.

A.3.5 Extra masking

After this the extra pixel mask is imported. A 3D histogram is filled at the location of the hot pixels. Later, when checking the hits in the detector, for every hit in the detector it is checked if there is an entry in this histogram. If that is the case the hit will not be used in the analysis, so it will be masked.

A.3.6 Creating outputfile

All of the if statements in this section are used to create the name of the output file, such that it exactly specifies which criteria are applied, what the values of the variables of these criteria are and if a mask is applied. This was very useful when trying different criteria and varying certain variables. When a choice has been made for the criteria however, this might not be as useful anymore, and the output filename can be shortened.

A.3.7 Creating histograms

In this section various histograms are initialized.

A.3.8 Import the TTree

In this section the TTree is imported. Furthermore, one extra histogram is made. This histogram needs the number of events, so that is why it is initialized at this later stage.

A.3.9 Creating TTree

It might be useful to create a TTree which contains only the selected events. This is called a secondary TTree. In this section such a TTree is initialized. This secondary TTree has almost the same properties as the original (primary) TTree, except for two things. Firstly, the primary branch ‘eventNumber’ resets to 0 sometimes. That means that there can be different events with the same eventNumber. This makes this eventNumber a bit of a useless variable. For that reason in the code (and in this explanation) this will be called the eventIndex now, and a new eventNumber will be created which is really a unique number. So when talking about the eventNumber from now on this will mean the unique number. The branch ‘eventNumber’ in the secondary TTree is a real eventNumber. Furthermore, the secondary tree has one extra branch: eventNumberOriginal. The eventNumberOriginal contains the eventNumber of the original file (so not the eventIndex). This eventNumberOriginal is installed to be able to backtrack which event this was originally. It is not useful to use the eventIndex for this, because every eventIndex belongs to several events.

All of those different event numbers can be a bit hard to keep track off, so this is a list of the different event numbers used in the code.

Numbers from the imported TTree:

eventNumberOriginal If a secondary TTree is imported, this indicates which eventNumber this event would have had in the original file. If this is a primary TTree, this variable has no meaning.

eventIndex The eventIndex from the input file. If this is a primary TTree this is an eventIndex, but if this is a secondary TTree this does not get reset to 0, so it is actually the eventNumber. To be save it is best to ignore this number since it is mostly just confusing.

event This variable is looped over, so it goes from 0 to nEvents (or different values when testing or when splitting the run into different parts). This means that this is the real event number of the imported TTree, and much more useful than the eventIndex.

Set into the new TTree:

eventNumberNew This is the new event number that will be given to the branch eventNumber in the output TTree.

eventNumberOld This is the old event number that will be given to the branch eventNumberOriginal in the output TTree. If a primary TTree is imported this means it takes the value of 'event', and if a secondary TTree is imported it takes the value of eventNumberOriginal.

The section which creates the TTree has a bit of a weird setup, the TTree is created, even if we do not want it in the end. If we entered that we did not want to create a new TTree, this file is deleted again. This setup is needed however, because the TTree must be initialized outside of the if statement to be accessible in the rest of the code, and the TTree can not be initialized if the file is not created. To make sure no previously made file is deleted, the name of the file is changed to 'tobedeleted', assuming no useful file with that name will exist in the folder.

A.3.10 Loop over events

Everything has been set up at this point and the loop over events can start. Some counters are initialized and the portion of events which is looped over is set. The first thing in the loop is the creation of the mTowerEvent (using one of the mTower classes). After this, the first loop over the hits in the event is done, to add them to a list of hits, excluding any hits which have to be masked.

A.3.11 Read event

The real number of hits in the event is set as the variable entries. Then some counters and histograms are updated to include this event in them, for this a second loop over the hits is made. In a loop over the chips the occupancy (fraction of pixels with hits) per chip is calculated. Lastly the number of hits of this event is fed to the distribution of the number of hits of the raw data.

A.3.12 Clustering

For the second criterion the clusters in the first layer are needed, and the cluster size for the different chips is entered into a histogram. First the vector vClusters is initialized. This is a vector containing different properties of the clusters. The first element of this vector is a vector with all of the lanecodes of the clusters. The lanecode of a lane/chip is a unique number, just like the lane number and the chip id. The lanecode is closely related to the layer number: the two chips in the first layer have lanecode 0 and 1, the two chips in the second layer have lanecode 2 and 3 etc. This makes finding the lane numbers of chips in a certain layer very easy. The conversion from lanecode to lane number is initialized in the beginning of the code for the first three layers, it can be expanded to the other layers if necessary. The next two elements of the vector vClusters are vectors containing the mean x and the mean y coordinate of the clusters, an the fourth element is a vector containing the cluster size of the clusters. This vector is filled with the information of the clusters in this section of the code.

A.3.13 Event selection

In this section the different criteria are applied. These criteria are not applied starting from criterion one and ending at criterion five but in a different order, because that way the program is faster and more clear to understand. For example, criteria C1ES, C4 and C5 all use a loop over hits, so they are applied in the same section.

A.3.14 Criterion C3 part 1 and 2

In the first part a loop is implemented over all of the clusters. If there is more than one cluster with a cluster size larger than one, the event will immediately be rejected by the third criterion. The only other case where an event is rejected by the third criterion is if the accepted cluster has a cluster size of one and there is another cluster with a cluster size larger than one. This is checked after the accepted cluster is found by C2, in part 2.

A.3.15 Criteria C1 and C2

Criteria C1 and C2 both need to check every cluster, so they are combined in this loop over clusters. For every cluster, first every hit in the layer behind it is checked to see the distance between the two, see the description in section 4.2.3. If at least one hit is close enough the cluster is accepted, and the loop over hits stops. However, if there already is an accepted cluster, that means this is the second accepted cluster, so the event is rejected. After determining if the cluster is accepted by criterion C2, criterion C1 is checked. This is not the criterion C1 described in this thesis, but a more 'forgiving' one: not events with hits in the first layer outside of the area are rejected, but just events with accepted clusters in the first layer outside of the area. After checking for criterion C2 one last check has to be done: if there is no accepted cluster, the event gets rejected.

A.3.16 Criteria C1ES, C4 and C5

At this point criteria C1, C2 and C3 are completely checked, only C1ES, C4 and C5 are left. First two variables are initialized, they are used for criterion C5. C1ES needs a loop over hits in the first layer, C4 needs a loop over hits in the second layer and C5 needs a loop over hits in the third to the tenth layer. The if statements in the loop over layers makes sure that only relevant hits are looped over. In the loop over hits, first the properties of the hits are calculated. Then criterion C1ES is applied: if a hit in the first layer is outside of the central part of the detector, the event is rejected. Then criterion C4 is applied: if a hit in the second layer is far away from the accepted cluster in the first layer, the event is rejected. Then the number of hits in different areas are counted for the fifth criterion. After all of the hits are looped over, the hits in the two different areas are compared, if there are too many hits at the border of the layer, the event will be rejected.

A.3.17 Fill histograms

In this section a few histograms are filled if the event is not rejected. If a TTree is made, the not rejected event will be added to the TTree.

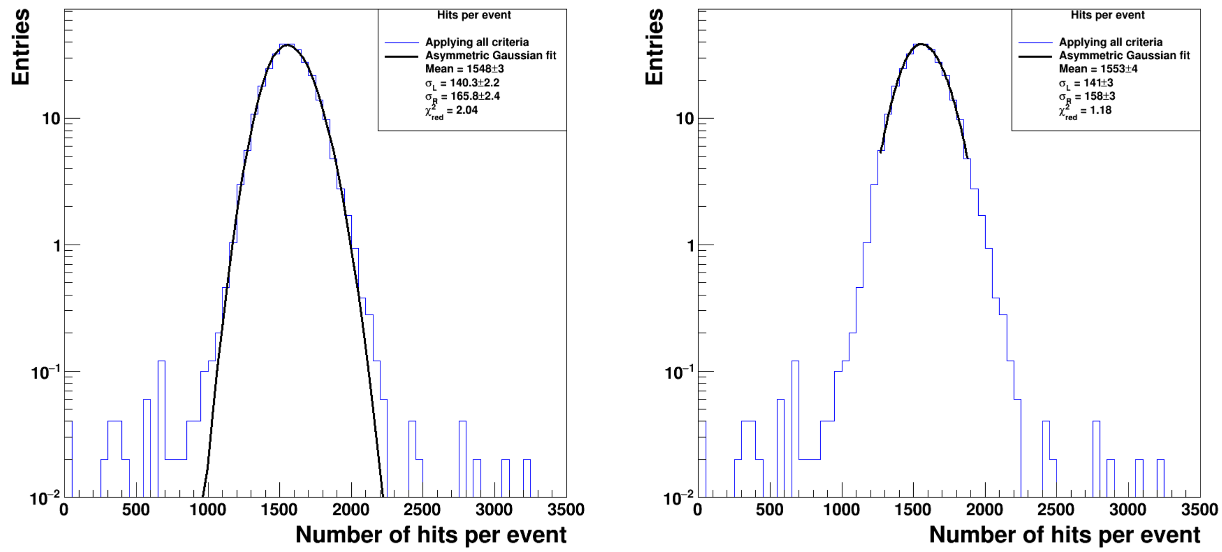
A.3.18 Closing off

Lastly, the hot pixels in this event are found, the values of a few counters are printed and the output files are written.

A.4 Final event set

In the directory 'finalSet', the root files with the event sets after selection can be found.

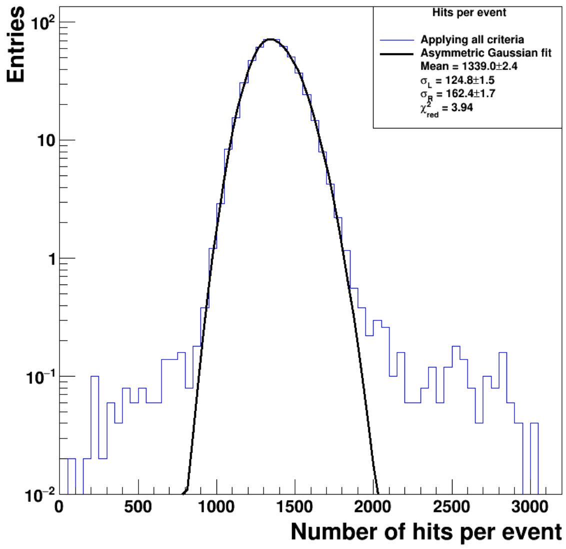
B Distributions of the number of hits for all energies



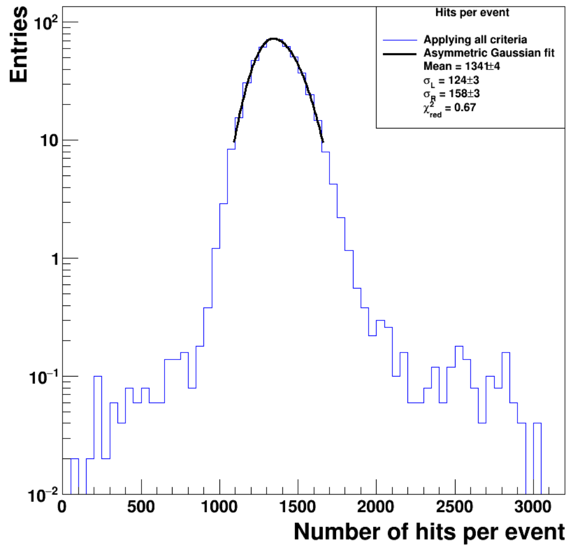
(a) The data is fitted with an asymmetric Gaussian over the whole distribution.

(b) The data is fitted with an asymmetric Gaussian over only $2\sigma_{\mathrm{L}}$ to the left and $2\sigma_{\mathrm{R}}$ right.

Figure 22: The distribution of the number of hits after applying all of the criteria, with incoming electrons with an energy of 5.8 GeV, run 1346.

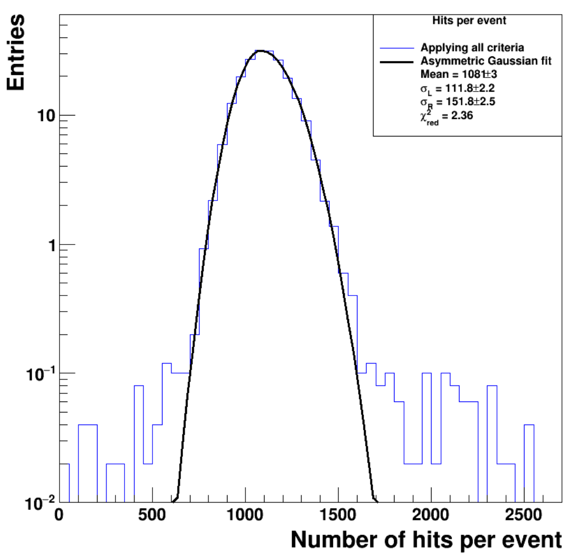


(a) The data is fitted with an asymmetric Gaussian over the whole distribution.

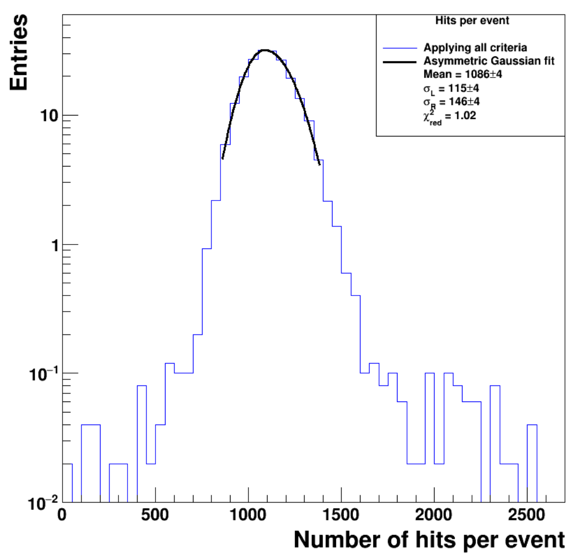


(b) The data is fitted with an asymmetric Gaussian over only $2\sigma_{\rm L}$ to the left and $2\sigma_{\rm R}$ right.

Figure 23: The distribution of the number of hits after applying all of the criteria, with incoming electrons with an energy of 5.0 GeV, run 1413.

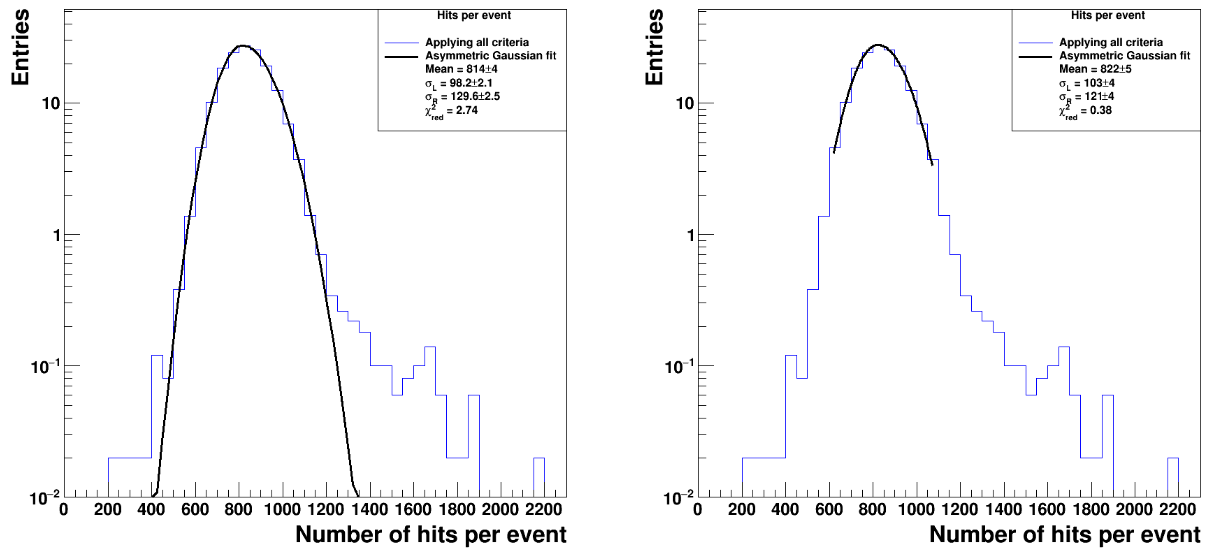


(a) The data is fitted with an asymmetric Gaussian over the whole distribution.



(b) The data is fitted with an asymmetric Gaussian over only $2\sigma_{\rm L}$ to the left and $2\sigma_{\rm R}$ right.

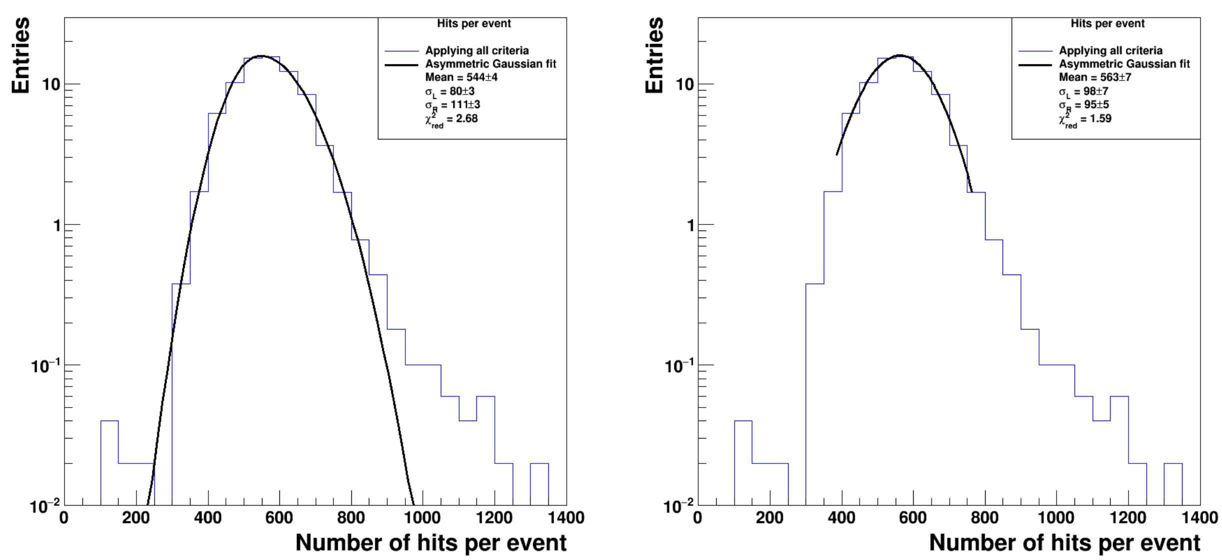
Figure 24: The distribution of the number of hits after applying all of the criteria, with incoming electrons with an energy of 4.0 GeV, run 1345.



(a) The data is fitted with an asymmetric Gaussian over the whole distribution.

(b) The data is fitted with an asymmetric Gaussian over only $2\sigma_{\mathrm{L}}$ to the left and $2\sigma_{\mathrm{R}}$ right.

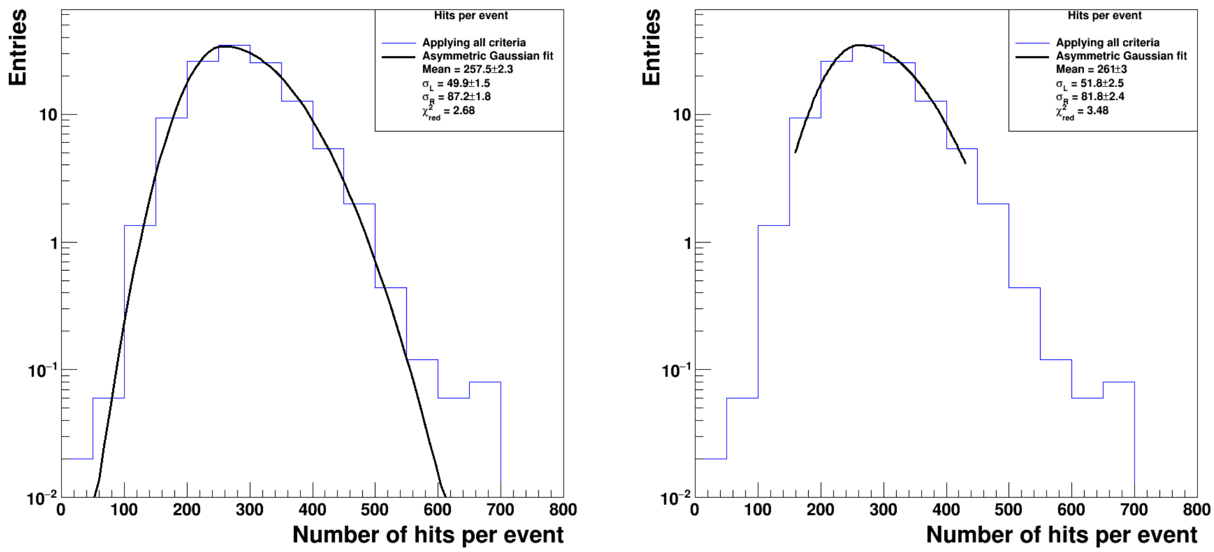
Figure 25: The distribution of the number of hits after applying all of the criteria, with incoming electrons with an energy of 3.0 GeV, run 1341.



(a) The data is fitted with an asymmetric Gaussian over the whole distribution.

(b) The data is fitted with an asymmetric Gaussian over only $2\sigma_{\mathrm{L}}$ to the left and $2\sigma_{\mathrm{R}}$ right.

Figure 26: The distribution of the number of hits after applying all of the criteria, with incoming electrons with an energy of 2.0 GeV, run 1276.



(a) The data is fitted with an asymmetric Gaussian over the whole distribution.

(b) The data is fitted with an asymmetric Gaussian over only $2\sigma_{\mathrm{L}}$ to the left and $2\sigma_{\mathrm{R}}$ right.

Figure 27: The distribution of the number of hits after applying all of the criteria, with incoming electrons with an energy of 1.0 GeV, run 1343.

C Percentage of events left after criterion 1

	5.8 GeV		5.0 GeV		4.0 GeV	
	Events	%	Events	%	Events	%
After criterion 1	80,593	100	162,381	100	78,364	100
After criterion 2	45,607	57	95,432	59	41,567	53
After criterion 3	44,972	56	94,116	58	40,896	52
After criterion 4	36,771	46	76,083	47	32,660	42
After criterion 5	35,702	44	72,895	45	30,473	39

	3.0 GeV		2.0 GeV		1.0 GeV	
	Events	%	Events	%	Events	%
After criterion 1	41,498	100	50,770	100	113,043	100
After criterion 2	19,162	46	17,838	35	22,539	20
After criterion 3	18,925	46	17,577	35	22,263	20
After criterion 4	14,992	36	14,005	28	18,834	17
After criterion 5	13,324	32	11,634	23	14,952	13

Table 3: The number and percentage of events left after applying the selection criteria, given for different beam energies. Since criterion 1 is only originally pointed at preventing partially contained showers, it can be useful to look at the percentage of events left with respect to the number of events left after applying criterion 1. This are the percentages shown in this table. The percentage of events left with respect to the raw data is shown in table 1.