UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Characterization of The FoCal-H Prototype Calorimeter

Sebastian Søderlund Wierød

Master Thesis in Computational Physics at the Niels Bohr Institute

Supervisor: Ian Bearden
Faculty of Science

February 15. 2022

Formal information

fkg212
CPR 250894-****
Sebastian Søderlund Wierød
Kabbelejevej 18A 1. tv.
2700 Brønshøj
Faculty of Science
Niels Bohr Institute
Hand in date: 15.02.2022

Acknowledgments

I would like to thank my supervisor Ian Bearden, Professor, NBI for guidance and support. A special thanks to Christian Holm Christensen for supporting with programming and software challenged I encountered. I am also grateful for the cooperation with my fellow graduate students.

Abstract

The ALICE Collaboration has considered installing a high-granularity forward calorimeter (FoCal) as an upgrade to the ALICE detector for Run 4 in 2026-2028. This thesis presents a thorough characterization of a proof of concept of the first prototype of the hadronic part of FoCal (FoCal-H). Moreover, the ongoing simulations of Focal-H were migrated from a local machine to a server to utilize the greater computing power of Computerome2.0 (C2).

The FoCal-H prototype is a sampling calorimeter which uses copper as the passive media and scintillating fiber as the active media. The testing of FoCal-H was done in collaboration with Focal-E at the SPS test-beam north area. Being the first prototype, many unknown obstacles occurred during the test run. The main issue was that the two Front-End CAEN boards had problems with synchronizing and matching events in-between themselves. For this reason, focus was shifted to only the inner board (B1), and an energy resolution was calculated on this basis. The energy resolution varied across the different experimental setups. The estimated energy resolution most similar to the simulations - of approx. 0.17 - resulted from the runs with a beam energy of 60 GeV. A rough translation was done from The Analog to Digital Conversion (ADC) values recorded in the CAEN boards to their corresponding beam energy (in GeV) to ensure that the energies were distinguishable in Focal-H.

The simulation framework was migrated to the C2 servers provided by The "Danish National Supercomputer for Life Sciences" (Computerome). The simulation framework is built in ROOT using the GEANT4 simulation tool kit, and a model similar to the first prototype was constructed. Using the same setup as the best-performing test runs, the compensation ratio was calculated to be ranging from 0.26 to 0.834, which is not sufficient to fulfill the compensation condition $\frac{e}{h} = 1$.
Hopefully, the findings in this thesis will contribute to optimizing the upcoming second prototype.

Contents

List of Figures

List of Tables

1 Introduction

1.1 ALICE and FoCal

ALICE, A Large Ion Collider Experiment, is an ongoing experiment at CERN first started in the early '90s. One of the main goals at ALICE is to characterize the physical properties of the Quark-Gluon Plasma (QGP). This state is believed to have existed just after the Big Bang and only for a millionth of a second [5]. It is no easy matter to achieve this

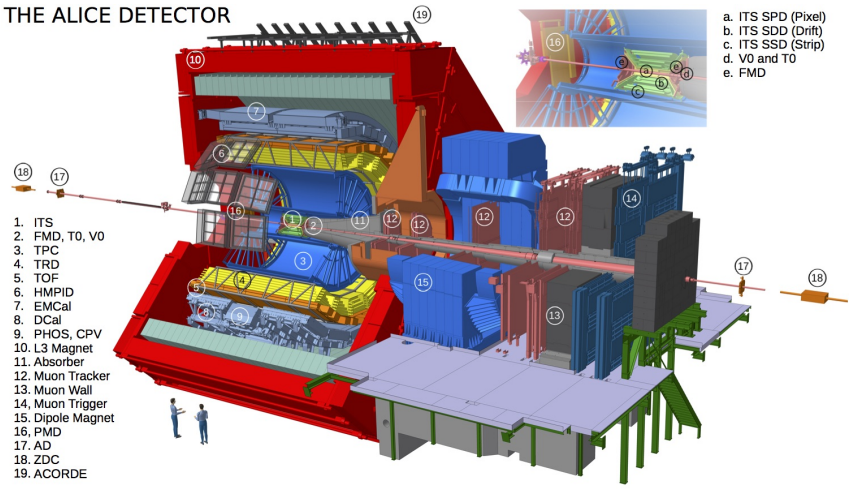


Figure 1: The ALICE Detector [7].

state, hence the need for the Large Hadron Collider (LHC). The LHC is the worlds largest particle accelerator, and it consists of a 27 kilometer long ring of super conducting magnets [6]. The LHC accelerates beams of large nucleus to velocities near the speed of light. The state is achieved by the collision of the beams of nuclei which creates extremely high temperatures similar to those of the early universe [5]. To measure and analyze the particles in this state, several different detectors are needed. An illustration of the sheer size of the ALICE detector can be seen in figure 1. Though already rather large, there is still room for another detector at ALICE, and a Forward Calorimeter (FoCal) could be the new addition to the fold of detectors. FoCal will be placed in front of the ALICE detector and the ALICE compensator magnet, as indicated in figure 2 and will be around seven and half meters away from the interaction point closest to ALICE [13].

1.1.1 FoCal

In the Letter of Intent for FoCal, several main goals of the FoCal physics program are mentioned, specifically [13]:

1. To measure the gluon density in protons and lead nuclei and to quantify its nuclear modification at small x and Q

2. To explore the physical origin of shadowing effects

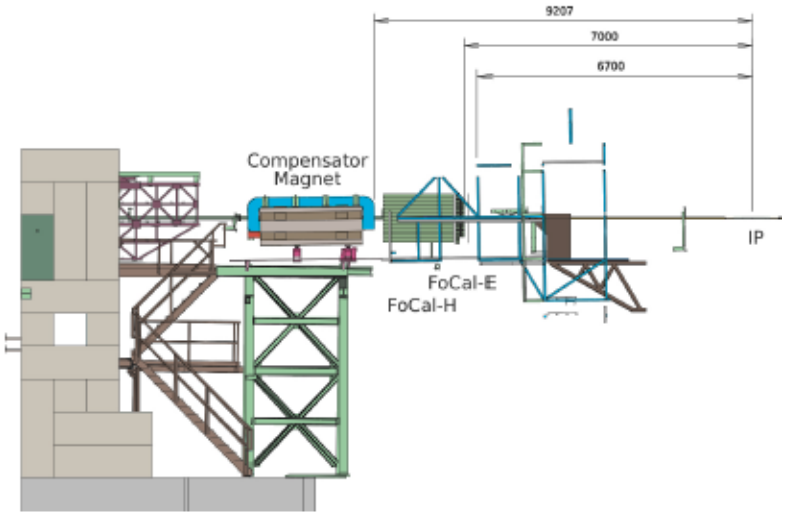


Figure 2: Installation of the FoCal at the 7m location with FoCal-E and FoCal-H detectors. [13]

3. To investigate the origin of long-range flow-like correlations in pp and p–Pb collisions

4. To explore jet quenching at forward rapidity in Pb–Pb collisions

But before any of the above can be achieved, several stages of testing the FoCal are needed. In this thesis, a proof of concept prototype for the hadronic part of FoCal (FoCal-H) will be characterized. The hardware and software used for the simulation of FoCal-H will be migrated to a server for faster processing, both now and for all future prototypes of FoCal-H.

1.2 The Standard Model

The Standard model of particle physics describes the elementary particles and three out of four fundamental forces, these forces being the electromagnetic force, the weak, and the strong force, only missing the gravity as the fourth and last force. In figure 3 the elementary particles derived into three families are shown, one being the gauge bosons marked by red, which act as force carriers and are thus responsible for interactions between the elementary particles. All gauge bosons have spin 1, leaving only the Higgs boson with a spin 0. The W and Z bosons are the force carriers of the weak force, and the gluon is the force carrier of the strong force [12].

The particles marked by green and purple are called fermions. They all have spin $\frac{1}{2}$ and are divided into two families, namely Quarks and Leptons. The main difference between the two families is that quarks interact with the strong force, while the leptons do not. The leptons carry electromagnetic charge, not including the neutrino, so they interact with both the electromagnetic and weak force. The Quarks carry both color and electromagnetic charge, hence they interact with all the three forces in The Standard Model [12].

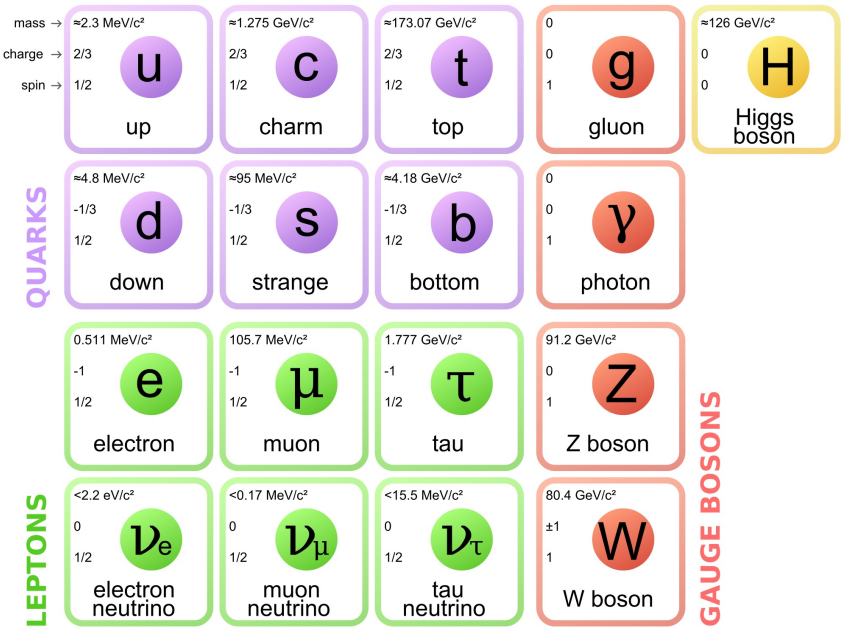


Figure 3: The elementary particles of The Standard Model [9].

The strong force is what binds the quarks together. A quark has never been observed as a free particle, rather they are always observed as composite particles called hadrons. This is called color confinement and is one of the phenomena described in Quantum Chromo Dynamics (QCD). It affects all particles with color charge and this is because the force between the two particles carrying color charge remain constant as they are separated. The energy then grows until a quark–antiquark pair is spontaneously produced, initiating a hadron jet. The most commonly known hadron is the proton consisting of two up quarks and one down quark. This is but one of properties of QCD which is the theory that describes the strong interactions between quarks and gluons [10].

1.2.1 Strong Interaction in matter

At the LHC, the high energy hadron collides at different collision points located around the accelerator. It is in these collisions that the properties of QCD take effect and the remains of these interactions are the ones measured at the different detector sites.
When these high energy hadrons transfer through a block of matter, they will ionize the atoms of the traversed medium and lose roughly 1-2MeV$g^{-1}cm^2$ of each their kinetic energy doing this process [2].
If the hadron travelling through the medium carries enough kinetic energy, it will at some given length interact strongly with an atomic nucleus and initiate a shower, in turn creating new hadrons which are primarily pions. The struck nucleus is left in a highly excited state and decays into multiple nucleons, followed by the emission of γ rays.
In figure 4, an illustration of a hadron travelling through a given material and the com-

ponents of the hadron shower is shown. The strong interaction with the medium will not

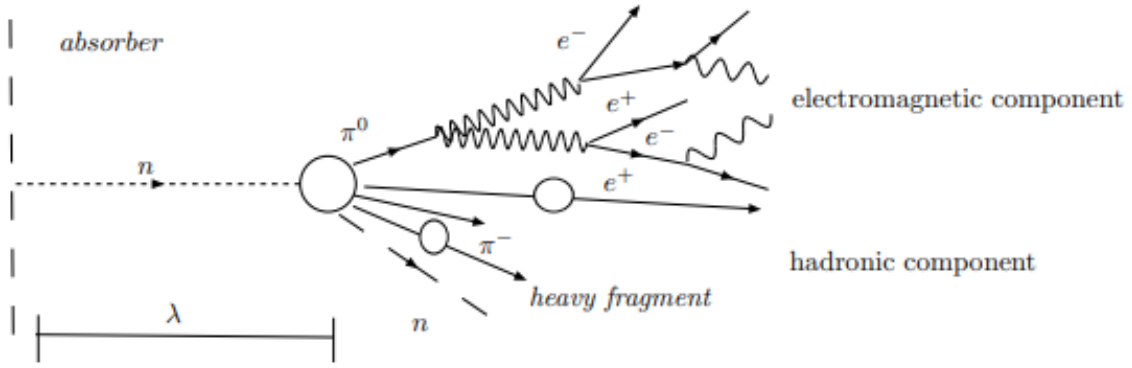


Figure 4: A schematic depiction of the hadronic shower development and its subprocesses [2].

only consist of a hadronic part but also an electromagnetic part, because some of the particles produced in the hadron shower decay electromagnetically. In particular, the π^0s and ηs decay through the electromagnetic interaction to: $\pi^0, \eta- > \gamma\gamma$. The π^0s decay into two gammas in 98.8% of the cases, and into an electron-positron pair plus an additional gamma in 1.2% of the cases. The fraction of the initial hadron energy that is converted into π^0s and ηs varies strongly from event to event. The average fraction (f_{em}) can be approximated as a function of the incoming hadron energy E [2]:

$$f_{em} \approx f_{\pi^0} \approx (0.11 - 0.12) \cdot ln(E/GeV) \tag{1}$$

The hadronic fraction would then be $f_{had} = f_{em} - 1$. The size of this electromagnetic component is largely determined by the production of π^0s and ηs in the first interaction, hence it fluctuates heavily from event to event. The majority of the electromagnetic induced shower comes from π^0. Only a small contribution to f_{em} is provided by gamma radiation of the excited nucleus mentioned previously.

The mean length of which the hadron will interact strongly with a nucleus is called the Nuclear Interaction Length (λ_{int}) which is approximately [2]:

$$\lambda_{int} \approx \frac{35A^{1/3}}{\rho} \tag{2}$$

Where λ_{int} is the Nuclear Interaction Length, A is the mass number of the material, and ρ is the density. Compared to protons and neutrons, mesons such as pions or kaons are smaller and, therefore, less likely to encounter a nucleus, resulting in roughly a 25% increase in distance traveled and thus losing 25% more energy. The probability of an inelastic hadron-nucleus interaction P_1 after a distance x is given by [2]:

$$P_1 = 1 - e^{-x/\lambda_{int}} \tag{3}$$

All this must be considered when characterizing and analyzing the properties of a hadronic shower, and a hadron calorimeter could be an appropriate solution.

2 Calorimeters In Particle Physics

Calorimeters were first used in thermodynamics, and the word "calorie" is a unit of energy which is defined as the energy needed to increase the temperature of one gram of water. Though a calorimeter in particle physics still measures energy, it is on a much smaller scale. It measures the total energy of any particle interacting with electromagnetic or strong force. So it make sense to distinguish between an electromagnetic calorimeter and a hadronic calorimeter. Calorimeters are either homogeneous or have a sampling structure. For the homogeneous calorimeter, the entire volume is active. The homogeneous calorimeters are rarely used as hadronic calorimeters, but they are typically used for neutrino experiments with the production of Cherenkov light. Cherenkov light, or cherenkov radiation, is electromagnetic radiation that is emitted when a charged particle travels through a dielectric medium at a speed greater than light in that given medium. [1]

However, the sampling calorimeters is widely used as hadronic calorimeters. The sampling calorimeter consists of two media, an active one and a passive one, for signal generation and particle absorption, respectively. The passive media or the absorber material is where the particle interacts strongly and initiates the hadron shower, creating charged particles which then interact with the active material [1]. Sampling calorimeters can generally be classified by the type of active medium. The different classes of sampling calorimeters are: scintillation calorimeters, gas calorimeters, solid-state calorimeters, and liquid calorimeters. In scintillation calorimeters, the signal is collected in the form of light. In the three other classes, the signal is collected in the form of electric charge. For the scintillation calorimeters, the active medium is often a plastic (organic) scintillator in the form of fibers or plates. Combined with an easy accessible absorber material, the calorimeter can be relative cheap. They are also easy to assemble into a variety of geometries. They have a fast response and a good light yield. Furthermore, they can be made compensating by properly tuning the ratio between the amounts of absorber and scintillator [3]. This will be elaborated upon in the next section.
When a charged particle enters the active medium, it loses energy through the electromagnetic interaction with the Coulomb fields of the electrons. Some of the energy will bring some atoms into an excited state. This excited state is unstable, and the atom will return to its ground state, usually by emitting one or more photons. This is called scintillation, i.e. signal generation by the emission of light. Since a single photon only carries a small of amount of energy, an amplifier such as silicon photomultiplier or a photomultiplier tube would be an ideal addition to the setup. Photomultipliers converts a single photon into electric signals multiplying the signal to a measurable scale. However, they can suffer from aging and radiation damage over time [3].

There are several good candidates for an absorber material for a hadron calorimeter. As one knows from equation 3, the probability of a particle interacting strongly is highly dependent on both the particle itself and the media it traverse. In table 1, various different absorber materials and their respective interaction length are shown. So when building a calorimeter, the properties of the absorber material are crucial for the amount

Absorber	Z	A ($g \cdot mol^{-1}$)	$\lambda_{int}(mm)$	$\lambda_{int\_\pi}(mm)$
Aluminum	13	26.98	397.0	506.4
Iron	26	55.85	167.7	204.2
Copper	29	63.55	153.2	185.1
Tungsten	74	193.84	99.46	113.3
Lead	82	207.2	175.9	199.3
Uranium	92	238.0	110.3	124.2

Table 1: A list of various absorber materials nuclear interaction length and pion interaction length [14]

of deposited energy in the calorimeter.

2.1 Particle Measurements with Calorimeters - A simplified experimental setup

A simplified experimental setup with the focus of measuring particles with calorimeters could be a tracking chamber, a electromagnetic calorimeter followed by the hadronic calorimeter.
The Tracking Chamber tracks the particle going into the detector. One method would be to use a magnetic field bending the path of the charged particles. It is then possible to both derive the momentum and the sign of the charge of the particles given the curvature of the track in the magnetic field. Thus, the magnet controls what kind of particles that enters the calorimeters. Particles without a charge will be unaffected by the magnetic field. Thus, when detecting energy in the calorimeters, and no particles were "tracked" in the magnet, one can determine that this particle is either a muon or a neutral particle like photons and neutrons. [11]
From the tracking chamber, the particles go through an electromagnetic calorimeter. This is where the electron/positrons and photons interact with absorber material via the electromagnetic force and are absorbed within the calorimeter. The energy is then calculated from the size of the electromagnetic shower detected and/or from how far the particles traveled in the calorimeter, if the detectors design allowed it.
Like the electromagnetic calorimeter, the hadronic calorimeter does the same but with hadronic matter. Though, as one knows from section 1, a hadronic shower contains both a hadronic and an electromagnetic part. The ratio between the two is[2]:

$$\frac{e}{h} \equiv \frac{\epsilon_{vis}(e)}{\epsilon_{vis}(h)} \tag{4}$$

Not all energy is detectable why it makes sense to talk about visible/detectable energy. From equation 1, the approximation of the f_{em} from π_0 was done. Hence the detectable energy from a pure pion signal is given by [2]:

$$\epsilon_{vis}(\pi) = f_{em}\epsilon_{vis}(e) + (1 - f_{em})\epsilon_{vis}(h) \tag{5}$$

Therefore the e/π ratio is:

$$\frac{e}{\pi} \equiv \frac{\epsilon_{vis}(e)}{\epsilon_{vis}(\pi)} = \frac{\epsilon_{vis}(e)}{f_{em}\epsilon_{vis}(e) + (1 - f_{em})\epsilon_{vis}(h)} = \frac{e/h}{1 - f_{em}(1 - e/h)} \tag{6}$$

Then using equation 6 the e/h ratio can then be written as:

$$\frac{e}{h} = \frac{e/\pi \cdot (1 - f_{em})}{1 - f_{em} \cdot (e/\pi)} \tag{7}$$

In the specific case where $e/h = e/\pi = 1$, the energy dependence of f_{em} cancels out, and the ratio e/π becomes energy-independent. It then follows that:

$$1 = \frac{e}{h} = \frac{\epsilon_{vis}(e)}{\epsilon_{vis}(h)} \rightarrow \epsilon_{vis}(e) = \epsilon_{vis}(h) = \epsilon_{vis}(\pi) \tag{8}$$

This means that the electromagnetic signal of the hadronic shower does not clutter the hadronic signal in the detector because it lies within the same visible range. The condition where $e/h = e/\pi$=1 is called the compensating condition. If the condition is false, the relation between incoming energy and signal becomes non-linear because of the energy dependence of f_{em} [2]. The condition depends on both active and passive material in the detector and the ratio between these two. To ensure this condition, it is necessary to find the optimal materials and ratio of these, and this could ideally be done in computer simulations.
The calorimeter's performance is highly dependent on containing the entire shower within it. So, not only are the properties of absorber and scintillator an important consideration when building a calorimeter, but so is the ratio between the two and the size of the calorimeter. The performance or energy resolution of a sampling calorimeter can be calculated by statistical processes [2]:

$$\frac{\sigma(E)}{E} \propto \frac{\sigma(\epsilon_{vis})}{\epsilon_{vis}} \tag{9}$$

Where E is the energy and $\sigma(E)$ is the variance/spread of the energy measured in the detector.

3 Simulation Setup

Before starting to acquire any materials needed to build a detector, simulating and testing different geometries and materials should ideally be done by computer simulations. To simulate the physics in a calorimeter, one needs a high amount of computing power to gain a statistically significant amount of data within a reasonable amount of time.

It is no easy feat to construct a simulation framework, especially not one involving the complexity of the physics in a hadronic calorimeter. And since the final product of FoCal

does not only consist of a hadronic calorimeter but also the electromagnetic calorimeter, the electromagnetic part will be simulated as a block of tungsten, to save both time and computing power. While this is not an ideal simulation, it does provide a reasonably realistic indication that is useful for preliminary testing purposes. In the future, the full FoCal should be simulated for optimal results.

3.1 The simulations pre prototype

The early simulations were done on a local machine with limited amount of computing power. Each simulation runs on a single core and could take up to several hours to run, depending on the amount of particles and energy.
The software used for these simulations was Geant321 which is a collection of Geant3 and VMC into a single package (Virtual Monte Carlo). Geant3 was the base of the simulation geometric. The early simulations were used for testing and discussing the first prototype design. Though due to the lack of computing power, it was impractical to generate sufficient data to attain statistical significance. Nonetheless, it did give an approximation of what one might expect.

The idea was to build a spaghetti calorimeter with copper tubes as the absorber material and optical fiber as the active material. As seen in table 1, there are several good candidates for the absorber material. Several of them were simulated and tested. In the end, the end copper proved to both be cost efficient and show promising results. One of the main things that were tested was the significance of the shape of the tubes as well as the ratio between the absorber and the scintillator [24].

The simulated detector consisted of 25 by 25 copper tubes working as the absorber. Each tube was filled with a optical fiber. The fiber had a diameter of 1 mm, and the tubes measured an outer diameter of 2.5 mm, an inner diameter of 1.2 mm, and a length of 1.1 m. The tubes were stacked to form a square geometry. At first, two designs of the copper tubes were considered, - One with circular tubes, and the other with hexagon tubes. One could expect that since air is a poor absorber, the air gaps might have had an impact on the results. While it did have an influence on energy deposition, it did not make an impactful difference in energy resolution. It was discussed whether by filling the air gaps between the circular copper tubes with thin copper thread , it would become on par with the hexagon shaped tubes. While this is worth consideration, it was not done for the first prototype, both due to lack of time and the consequent increase in cost of the detector[24].

The simulations was simulated with a beam energy of 500 GeV. The ratio for the setup with circular copper tubes between the absorber and the scintillator was three to one. The estimated a compensation ratio of this setup was: $\frac{e}{h} = 1.02$. Thus, the calorimeter in this setup would be considered compensated. The calculated energy resolution from that setup was 0.08 [24].

3.2 Upgrades for Hardware and Software

Since each simulation was time-intensive, an upgrade was clearly needed if threshold for the significant amount of data should be reached within an acceptable time frame. One solution to this is to migrate the simulation software from the local machine to a server. Fortunately, the University of Copenhagen Research Support section provides such services.

3.2.1 Computerome 2.0

The "Danish National Supercomputer for Life Sciences" (Computerome) is installed at the Technical University of Denmark National Lifescience Center.[18] Computerome is a Danish High-Performance Computing system (HPC) co-owned by The University of Copenhagen (UCPH) and the Technical University of Denmark. Computerome 1.0 (C1) was opened in November 2014 with the intent to address the need for a secure and modern HPC infrastructure, with a focus on being able to process sensitive health data and supporting world-class life science research. The current running setup, Computerome 2.0 (C2), was opened in 2019. There are two kinds of accessibility on C2. The first one is the C2-spc (Secure Private Cloud) which can be viewed as a secure private super computer, based on Linux servers. The second is the "Sandbox" which is accessible for free for a limited amount of time for project development, and test or educational purposes [19].

On C2, there are four types of machines:

- Login nodes
- Fat nodes
- Thin nodes
- GPU nodes

Where its total computing resources consist of around 31760 CPU cores with 210 TB of memory, connected to 17 PB of High-performance storage. The specification for the number of available cores and nodes for each category nodes can be seen in figure 5.
The login nodes are the default nodes a user connects to, upon login in. They contain three nodes in total, each with 40 cores. These nodes serve to navigate between projects, one's personal folder, etc.. Users are generally not allowed to run their own programs on this node, but except for testing small programs. Upon violation of this rule, C2 administrators will contact the user and stop the running jobs. The fat and thin nodes are the versatile nodes, which are used for running all regular programs. The GPU nodes are, used in part for high-resolution images and graphical dependent programs, though GPUs can perform parallel operations on multiple sets of data, hence they can be used for non-graphical tasks as well.
Through the support from either C2 directly or through the UCPH IT department, any employee/student from UCPH can order a "project" on the C2 servers. A single user can have up to multiple projects simultaneously, and multiple users can have access to the same given project as well.
For every project, five default folders are generated: "apps", "archive", "data", "people"

Category	Nodes#	Type	CPU	CPU#	Cores/CPU	Cores/node	Total cores#	Memory	Total Memory
Login node	3	HPE ProLiant DL380 Gen 10 Server	Intel Xeon Gold 6230 @2.1 GHz	2	20	40	120	192GB, PC4-2933 @2933 MHz	576 GB
Thin compute node	696	HPE Apollo 2000 System	Intel Xeon Gold 6230 @2.1 GHz	2	20	40	27.840	192GB, PC4-2933 @2933 MHz	133.632 GB
Fat compute node	55	HPE ProLiant DL360 Gen 10 Server	Intel Xeon Gold 6230 @2.1 GHz	2	20	40	2.200	1536GB, PC4-2933 @2933 MHz	84.480 GB
GPU compute node	40	HPE Apollo 2000 System + NVIDIA TESLA V100	Intel Xeon Gold 6230 @2.1 GHz	2	20	40	1.600	192GB, PC4-2933 @2933 MHz	7.680 GB
Total	**794**						**31.760**		**221 TB**

Figure 5: The available hardware on the C2 servers [8].

and "scratch". The only folder restricted by C2 is the "people" folder where each individual connected to the project has his or her own personal folder. The other folders are somewhat up to each user to define, but the general purpose for each folder is as the name states.
When installing software for a given project, the user has two options. One is to use C2 service for pre-installed modules by using the command "load module," and if the software isn't installed on the C2 serves, one can submit a ticket and fill in the software request. C2 will sometimes decline due lack of documentation for the software or for various other reasons. In general, the C2 supporters are very helpful, and only a few cases get declined. C2 already has an impressive software library, and they strongly encourage the user to take advantage of the C2 modules as they give better control of the software environment in the project.
If C2 is somehow unable to create a module for the required software, it leaves the user with the second option; to install and build your own software, preferably in the folder "apps". Here is permitted any software or program, not in violation of C2 terms and conditions.[20]
When all necessary software for the users programs and scripts have been installed on the servers, the users can submit a "job" on the C2 servers asking to "lend" the nodes and cores for a specific amount of time. If the user does not ask for a specific node, the job might be submitted to a GPU node which can cause problems and incur unwanted costs. The user can submit jobs and select the required nodes by using the "qsub" or "msub" command. For testing purposes that exceed what is considered acceptable on the login nodes, the user can run an interactive job by running the "iqsub" command, proceed through the guiding C2 questions regarding GPU, cores and ram, and then load the module etc. The "iqsub" command only allows for a single node to be occupied at a time. Thus, while using this command, the user only has access to 40 cores in total. The prices for using the C2 nodes and storage can be seen in figure 6.

At the time of writing, the use of the C2 servers is free of charge for all people associated with UCPH SCIENCE faculty, and the executive management at SCIENCE has determined that the use of C2 for researchers at SCIENCE will continue to be funded by the faculty until the end of 2024. This, coupled with the next-to unlimited computing

CPU/GPU node	3.9000 DKK/Node/Hour
Storage	0.0917 DKK/GB/Month

Figure 6: C2 and C2-spc prices for nodes and storage [21].

power to a standard local machine, makes it a perfect opportunity to use the C2 facilities for the base for future simulations of the FoCal-H.

3.2.2 Software - Geant3 to Geant4

The Software needed to run the simulations on C2 are:

- Gcc v5.4.0 - GNU Compiler Collections used for C, C++
- Root v6.24.06 - Data Analysis Framework. [30]
- Geant4 v10.7.2 - A simulation tool kit. [31]
- VGM - Virtual Geometry Model. [33]
- Geant4VMC - Virtual Monte Carlo for Geant4. [32]
- Root-framework - A generic framework using ROOT. [34]
- Root-simulation - A generic Simulation framework using ROOT. [35]

The C2 supporters were unable to build the complete software environment needed for the simulations. It was only Geant4 there was used as a C2 module. This is because, even though C2 had several versions of Root installed, the installed versions were incompatible with the gcc compiler used in Geant4, and, unfortunately, they declined creating a module with this compatibility. Since all the other software were dependent on Root, they and Root itself needed to be built and installed from scratch.

As mentioned above, the early simulation used Geant3 as the simulation tool kit. Therefore, it was necessary to either rebuild the geometry with Geant4 or use the build-in tool in VMC to translate Geant3 to Geant4 geometry. VMC is fully integrated with the Root geometry package TGeo and, as shown in figure 7, the VMC package is compatible with both Geant3 and Geant4. So, in theory, it should be easy to translate Geant3 geometry directly to Geant4 geometry with Geant4VMC. Despite making a few changes to the program building the geometry of the simulation, it was done. It was mainly the base of the geometry, since Geant4 cannot use the "Assembly" function when building the geometry. Instead, a redefinition fitting Geant4 syntax was used.

3.3 Running The Simulation On C2

When all the software had been installed into the project on C2 and the setup scripts from the early simulation were uploaded, the system was ready to run the simulations.

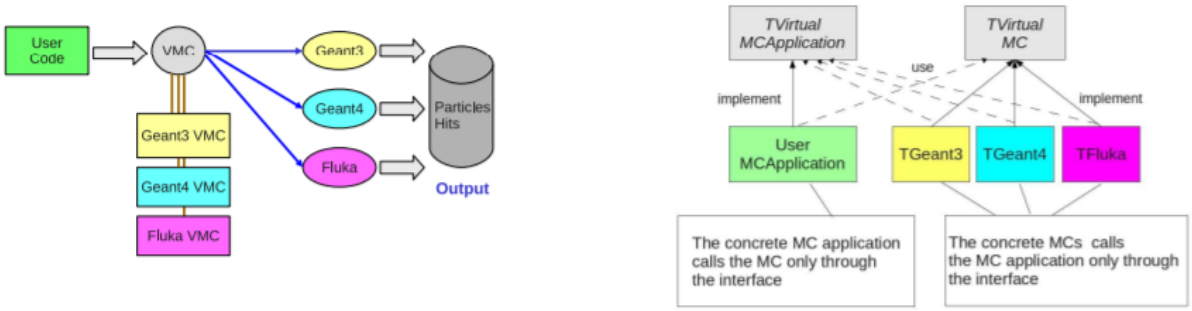


Figure 7: The VMC concept and design [4].

As mentioned, one can access the immense amount of computing power by typing the command "qsub" followed by the number of nodes, cores etc.. To avoid the repetitive work of loading the entirety of the software for every simulation, a shell script was created to automatize the process.
To run the simulation on C2 using the shell script:
Navigate to the project and the folder where the configurations scripts are located: /home/projects/ku_00098/TestSim/Focal_G4_Sim.
Use the "qsub" with the shell script, by typing in:
qsub -t 1-10%5 run_qsub.sh -F "1 40 500 0.2 2.5 1.2 0.5 40 275 1 3000 275 50 0"
Where the -t is the command that allows it to run parallel jobs. In this case, a total of ten jobs is ordered and five jobs are allowed to run at the same time. This should be matched with the number of cores ordered from the C2 servers to avoid occupying cores which stay unused. In this specific case, five cores would be enough. -F simply allows for inputting numbers after the shell script. The numbers following -F are the variables for the simulation geometry.
From left to right, the variables are the following: The number of times the program runs (this is not needed when one is running parallel jobs), the energy of particles, the number of particles, the random incident angle, the outer diameter of the tubes, the diameter of scintillator, the fraction (unused), the number of tubes, the half-length of the detector in mm, the outer material in this case copper, the placement of beam to the left of x-axis, placement of beam to the right of x-axis, the length of tungsten if enabled, enable/disable tungsten

More use full commands can be seen in appendix A. The description/guide on how to start the simulation is also available on C2 in the folder */home/projects/ku_00098/TestSim* called "Args_for_qsub".
Currently the shell script does the following:

- Loads in all necessary modules and software installed in "apps"
- Orders 1 node, 40 cores and 100 GB memory from the fat/thin nodes
- Starts the simulation with a random seed and the written variables
- creates three output files, the console output, the errors generated (if any) and the data from the simulations.

As mentioned, one might not need all 40 cores every run. So when running simulation, either adjust the number of cores to the amount of runs or create new shell scripts fitting ones needs if running multiple test runs. This is a good habit, both to not waste C2 resources but also because the C2 service might not be free beyond 2024, thus requiring payment for every node/hour.
Since multiple data files will be created when running parallel jobs, it is necessary to add the data files together to get the proper representation and statistically significant amount of data. At the moment of writing this it is done on a local machine. This should also be migrated to the server, to further optimize the process.

At the moment, three people have been added to the project "KU_00098". It is recommended to keep the project alive, not only to avoid the hassle of installing and building the software again, but also because it might be useful for other future projects where a local machines power is lackluster.

3.3.1 On Going Simulations And What Is Next

Using the basis of the simulation setup and analysis of the early simulation [24], both the first prototype and first concepts of the second prototype will be simulated on C2. The geometry of the first prototype will be elaborated in the following section. Currently the idea for the second prototype is to load several modules like that of the first prototype together to construct one large detector. The end geometry has not yet been decided, but the length will be 1100mm like in the early simulations.

4 SPS Test Beam Area and Experimental Setup

4.1 SPS Test beam Area

The Super Proton Synchrotron (SPS) is the second-largest accelerator at CERN, measuring nearly seven kilometers in circumference. It accelerates particles from Proton Synchorotron and is the last link in the accelerator chain that feeds particle beams to the LHC.
It also provides beams for a range of non-LHC experiments like the NA61/SHINE and NA62 experiments, the COMPASS experiment, and the north and south testing areas [17]. The SPS testing areas is, therefore, an ideal place for testing the FoCal-H. The first prototype was tested at the north area.
In figure 8, an illustration of the north area is shown. The beam going into the north area is split into three primary targets; T2, T4 and T6, which then provide the beam to the different beam lines; H2, H4, H6, H8, M2 and K12.
The north area also has three large magnets, the "CMS M1 magnet", "Goliath" and "Morpurgo" in beamline H2, H4 and H8 respectively. The CMS M1 magnet is a superconducting dipole with a magnetic field of 3.0T, the Goliath is a classic dipole magnet with a magnetic field of 1.5T and the Morpurgo magnet is also a superconducting dipole with a magnetic field of 1.5T [16].

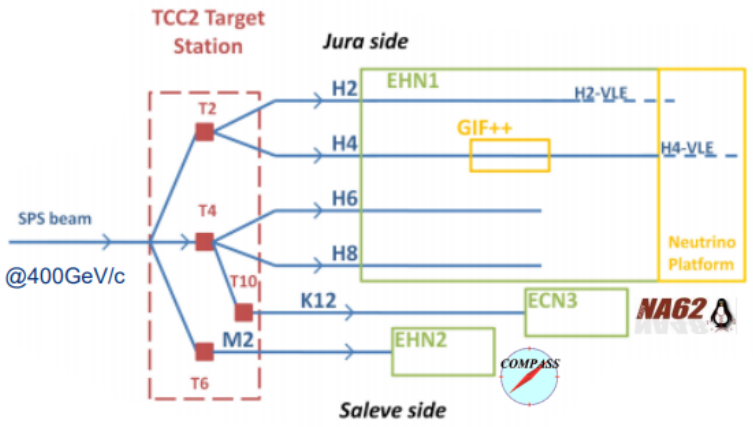


Figure 8: SPS North Area [16].

4.2 Experimental setup

The characteristics of the beam from T2 and T4 are shown on figure 9. Here, one can see the maximum momentum, acceptance, intensity of the spill for hadrons and electrons, and what kind of particles types the beam could contain . The FoCal-H testing was done in H4, and it was possible to regulate the beam energy locally through an on-site computer, between 20 GeV, 30 GeV, 40 GeV, 60 GeV, 80 GeV, and 120 GeV. For most of the testing period, the beam contained p^+ and e^+.

Parameters	T2		T4	
Beam Line	H2	H4	H6	H8
Maximum Momentum [GeV/c]	400 / 380	400 /380	- / 205	400 / 360
Maximum Acceptance [uSr]	1.5	1.5	2	2.5
Maximum Δp/p [%]	± 2.0%	± 1.4 %	±1.5%	±1.5%
Maximum Intensity / spill * (Hadrons / Electrons)	$10^7/10^5$	$10^7/10^7$	$10^{7\,**}/10^5$	$10^{7\,**}/10^5$
Available Particle Types	Primary protons*** OR electrons OR muons OR mixed hadrons (pions, protons, kaons)			
Other / Special requests	sba-physicists@cern.ch & sps.coordinator@cern.ch			

Figure 9: SPS beam characteristics [16].

The spill rate of the beam could be monitored live, and figure 10 illustrates the the beam intensity building up. In this example, there are two spills in each spill cycle, which happened in 40-seconds intervals for 4-5 seconds each. Throughout the majority of the testing period, there was only one spill each cycle. The spill length and repetition frequency depend on the physics program of all the facilities served by SPS. Therefore, we had no influence in this matter, and the same goes for the charge of the beam. It is possible to send in beam requests in advance to the SPS coordinator if needed. During

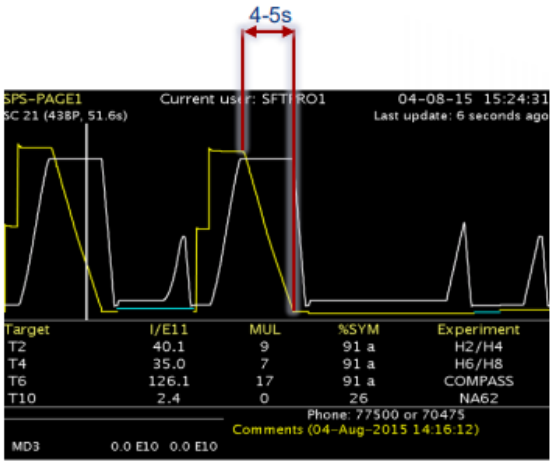


Figure 10: Example of the monitored SPS Beam spill rate [16].

the testing period, the beam charge was changed from positive to negative, but in theory it should have no effect for the FoCal-H testing, since it just changed the charge of the pions from p^+ to p^-, which has the same properties.

Three other experiments from FoCal were running along with the test of FoCal-H prototype, EPICAL, FoCal-E-PIXEL (PIXEL) and FoCal-E-pads (PADS). This made it possible to create a setup resembling the real end-setup for Focal with both the electromagnetic part and the hadronic part of the calorimeter. Multiple different setups were tested both with multiple "obstacles" and with none at all. Obstacles - meaning one of more of the three other detectors - were in front of the FoCal-H. This created around 118 analyzable data files in total during the test period.
Each run was recorded in an E-log, along with some of the other experiments. Because the E-log did not provide specifics regarding which FoCal-E experiment as the obstacle, thus PADS and PIXEL will both be regarded as Focal-E going forward. The setups which will be focused on in this thesis is:

Focal-E Setups: Setups with Focal-E as the obstacle.

Setup 1: Without tungsten and FoCal-H backend tilted by an 11.5mm spacer

Setup 2: Without tungsten

Setup 3: With 70 mm tungsten

EPICAL Setups: Setups with EPICAL as the obstacle

Setup 4: With 72 mm tungsten

Setup 5: With 72 mm tungsten and FoCal-H backend tilted by an 11.5mm spacer.

Figure 11: Photograph of the different steps in assembling the main module [22].

The setups where FoCal-H was transverse to the beam direction are not included in this thesis.

4.3 Prototype Design

4.3.1 Prototype main module

The main module of the FoCal-H prototype consists of a copper box measuring 90mm, 90mm, 550 mm. The copper box was filled with 1440 copper tubes which had an outer diameter of 2.5mm and an inner of 1.2 mm. Like in the simulations the tubes were stacked on top of each other forming a square with 36x40 tubes filling out the copper box. Each tube was filled with a scintillating fiber with a diameter of 1 mm, thus creating a 90x90x550 mm spaghetti hadron calorimeter. The fibers were gathered in bunches of 30 with an O-ring holding them together/in place, creating 48 bunches of scintillating fiber. The first steps of assembling the prototype can be seen in figure 11. The main module was then connected to a frame holding eight boards each with six Silicon Photo Multiplier (SiPM). In this way, each bunch had its own SiPM. The frame was carefully placed as close as possible to the fiber bunches. In figure 12a, the layout of the map of the 48 SiPM is shown.

1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8

(a) The matrix map of the 48 SiPM

0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1

(b) The CAEN Board Channel representation

Figure 12: The separation of channel from SiPM to the two CAEN boards.

The SiPM were connected to one of two CAEN board, B0 and B1. The CAEN boards

each had 32 channels, and, at first, the 48 SiPM were equally distributed between the two CAEN boards. This can be seen in figure 12b where the numbers represent the CAEN boards.
Due to one of the boards out-performing the other (B1) at the early stages of testing, it was decided to rearrange the channels such that B1 had all the 32 inner SiPM, and B0 had the remaining 16 outer SiPM, thinking that this would yield the best possible results.

0	0	1	1	1	1	0	0
0	1	1	1	1	1	1	0
0	1	1	1	1	1	1	0
0	1	1	1	1	1	1	0
0	1	1	1	1	1	1	0
0	0	1	1	1	1	0	0

(a) Final channel separation.

10	8	38	32	26	29	6	0
11	44	39	33	27	21	16	1
12	45	40	34	28	22	17	2
13	46	41	35	29	23	18	3
14	47	42	36	30	24	19	4
15	9	43	37	31	25	7	5

(b) The channel numbers.

Figure 13: The left figure illustrates the final separation of channels between each board and the right figure is the channels with their referred channel number.

In figure 13a, an illustration of the final configurations of "channel management" is shown. The CAEN Boards were then connected to a computer where one could access the data from the test runs.

4.3.2 Absorber and scintillator

Copper was chosen as the absorber material for the first prototype of FoCal-H. It had promising results from the early simulations and it is frequently used as an absorbed material, even though there are other materials with shorter interaction length. Note that in table 1, pions, which are the primary particle in a hadronic shower, have an interaction length of 185.1mm in copper, which is indeed roughly 25% longer than the nuclear interaction length of 153.2mm in copper. It is clear that the length of the detector is not nearly sufficient, so undoubtedly there will be some leak of energy through the back-end of the detector. The interaction length of copper is roughly 185.1mm, and the probability of a pion interacting with the copper in the detector can be calculated by using equation 3:

$$P_1 = 1 - e^{-550mm/185.1mm} \approx 0.949 \tag{10}$$

Compared to the length of the early simulation of 1100m:

$$P_1 = 1 - e^{-1100mm/185.1mm} \approx 0.997 \tag{11}$$

Which is clearly a sufficient length. Though note that the gain in probability of interacting with the detector by increasing the length has highly diminishing returns. The active

material was the scintillating fiber "BCF-10" with blue emission light, an emission peak of 432 nm, a decay time of 2.7 ns, and roughly 8000 photons per MeV [25].

4.3.3 The Silicon Photomultiplier

The Silicon Photomultiplier (SiPM) as the Photomultiplier Tube (PMT) is a sensor that excels in quantifying low-light signals. It also provides the benefits of a solid-state sensor. It is unaffected by magnet fields and can operate at low voltages. Therefore, the low-light signal coming from the scintillating fiber makes the SiPM a crucial element of getting any form of signal from the detector [26].
The SiPM model used in this prototype was the six by six mm MICROFC-60035-SMT-TR by the manufacturer Onsemi. It was primarily chosen because of its accessibility in the short time frame and its relatively low cost [27].

One could consider whether it would be easier to detect and distinguish each event/particle if each fiber had its own SiPM. This would be an ideal setup if one had the resources and equipment to handle the high amount of SiMP. For the first prototype, it was considered cost inefficient to do and simply not possible with the design. Furthermore, reading the signal from that amount of SiMP requires multiple CAEN boards, and as one will see later in thesis, two boards posed plenty challenges.

4.3.4 CAEN Boards

The CAEN Boards or the Silicon Photomultipliers Readout Front-End Board model A1702 made by the manufacturer CAEN Electronic Instrumentation was used to process the signal generated from the SiPM [28].

As mentioned, this specific model had 32 readout channels, and it was possible to change each channel sensitivity individually by increasing/decreasing the voltage/gain on the boards. As one will see later in the thesis, this could have had a large impact on the data collected.

The CAEN board transform an analog voltage to a binary number - Analog to Digital Conversion (ADC), which in this case is an arbitrary energy unit. The maximum ADC a channel could produces was 4089 ADC. A translation from ADC to GeV can be done to gain a proper understanding of the signal generated from the CAEN boards.

4.4 Merging data sets

As mentioned in the previous section, one of the boards did not trigger as much as the other, and this would create an unequal amount of events in the two boards. So It was necessary to run a separate program to merge the data set from each board together. The code from a shared github[15] was used for the merger of the data sets between the boards and the program "Collate.C" select and save events as follows:

First step; It opens the Ttree's from each board. Then it scans the files from each board, starting from last entry, until it finds a time smaller than the previous one. This is a synchronization event and this time is stored as an entry number.
Second step; For every time stored in each board from the first scan, it checks if the current event board time is larger or equal to that time. If it is, and either of the two following conditions are met, it adds that board to the event to be merged:

1. The time is within the specified tolerance.

2. If time is not within the tolerance, check that the time before that is within the tolerance.

If neither of the conditions are met or no time is equal or larger to the current considered time, create a new time to be the output.

The tolerance can be controlled by the user. A higher tolerance allows more events to be merged, and low tolerance allows less. However, to high a tolerance will cause unrelated events to be merged together. The default tolerance, and the tolerance used for all of the merged data set analyzed in this thesis, is 1000. The program is set up for multiple boards, and all boards that do not participate in an event will have their output set to "0xFF".

Third step; Return to the second step and continue until there is no more input data.

This creates an output file named "merged.root" which contains a single Ttree with a branch for each Board and a header with the mean time (t) and the number of boards (n). Each branch from the ttree has the following entries: (board) the Board id, (adc) the ADC values from the 32 channels, (t) timers and (tref) the reference time (this is always zero). The program "CollateAll.C" which runs the "Collate.C" program on a set amount of data sets, can be used to merge multiple data sets. This will produce data sets with a variation of events with only one of the boards active and events with both boards active. The data sets with a high number of events with two boards (n=2) will be referred to as a high matching (HM) data set .
Due to preference, the root file was written to a csv file for processing in the program language Python.

5 Results and Data Analysis

A calorimeter detects energy, and therefore, the two of most important parts of characterizing a calorimeter are to distinguish the different beam energy deposited in it and to calculate its energy resolution. But before starting any statistical processes and analyzing data, one needs to assess how much noise clutters the data, and, in this case, what the CAEN boards considered events throughout the test runs.

5.1 Noise reduction and Event selection

5.1.1 Background noise - Pedestal subtraction

It was noted from the early tests and from the empty channels in the Focal-H setup that there was a lot of background noise. A test run with no beam was done, a fit for each channel was constructed, and the mean ADC values were subsequently calculated. The mean values were then subtracted for each channel for every event for all data sets. This will be referred to as "pedestal subtraction" from now on. This removes most of the noise

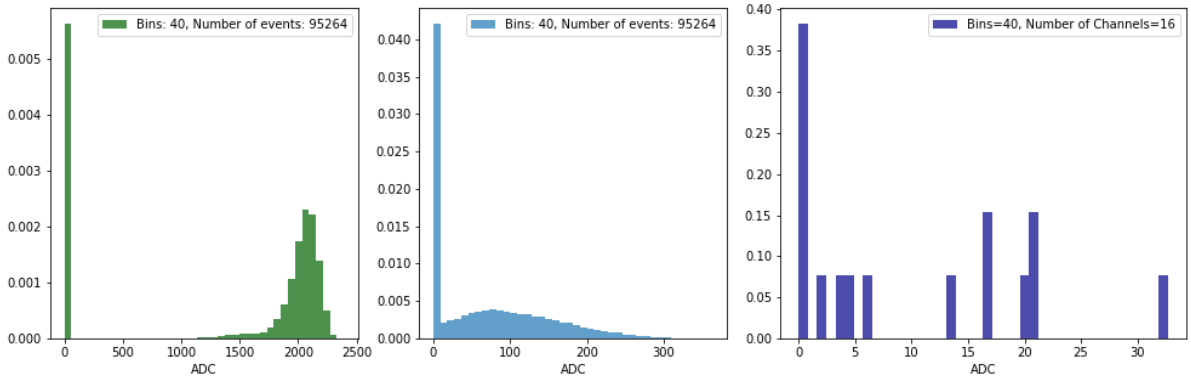


Figure 14: Three different illustration of the Empty Channels. On the left is the summed ADC value with no pedestal subtraction; in the middle, with pedestal subtraction; and on the right, ADC values from a single event with pedestal subtraction. The histograms are normalized so that they sum to one.

and can be seen in figure 14 where the ADC values from the empty channels from B0 are shown. Here, the left histogram illustrates the total ADC values with no pedestal subtraction and the middle is with pedestal subtraction. On the right is the ADC values from a single event with pedestal subtraction. As can be seen in the two histograms with pedestal subtraction, some degree of noise remains. To avoid getting negative ADC values, all values below zero were set equal 0, hence the larger peak around zero in the two histograms with pedestal subtraction. Many of the events from the empty channels did sum to zero, or close to. Though this noise is almost negligible, it can still clutter the true events. Lowering the gain might further reduce the noise.

5.1.2 Synchronization Events

The synchronization event, which was used to merge the data of the two boards, happened regularly while the experiments were running. These events where no actual shower happened - i.e. "dead events" - are saved in the data, - an event where no actual shower has happened. These should be sorted out before doing the data analysis. This process is simplified as the CAEN boards labels all synchronization events with t=0. So when either of the boards has a time equal to zero, that particular event is removed.

Figure 15 shows, an illustration of the synchronization events ADC distribution. As

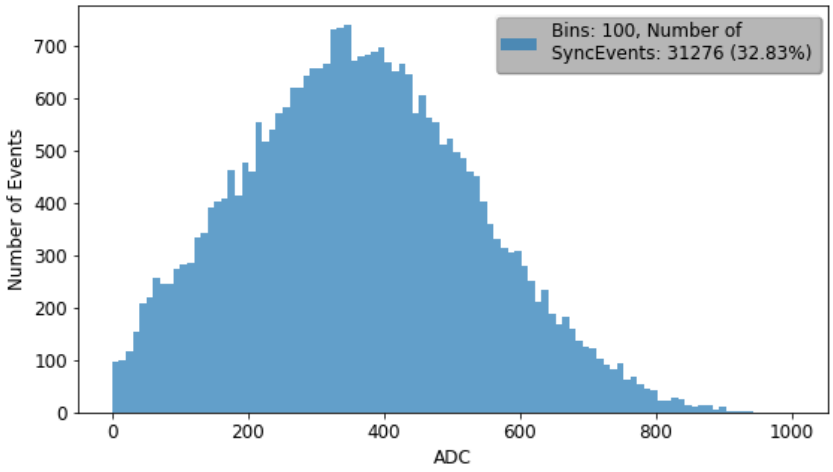


Figure 15: Synchronization Events from setup 5 with beam energy on 40 GeV.

will later become evident, these event do not lie within the true signal range. This specific data, was with setup 5 and with beam energy on 40 GeV. Roughly 30% of the data is a synchronization event. This differs significantly for each data set. It is possible in the given instance of a synchronization event, that an actual event is happening in that split second. This is, however, highly unlikely. Therefore, it was decided that all synchronization events should be removed from the data sets.

Unfortunately, many of the events where both board was active were synchronization events. This means that, for many data sets, there are next to no correlation between the events and the two boards, or very few at best. This makes it hard to see the "true" shower development in the channels where the two boards line up and effectively render B0 useless for most of the data sets since it only contains the outermost channels. This is assuming that the shower will cover more than one channel. Therefore, the data analysis will focus on the inner board (B1) and the data set with a high number of merged events.

5.1.3 Limit Events - Board overflow

As mentioned, the CAEN boards channels have a limit on the amount of information they can receive as input before they "overflow". If a given SiPM is still firing while this threshold has been reached for the respective channel, the remaining information from that SiPM would be lost, and the given event would then not be saved fully. It is highly unlikely that the boards read the energy fired from the SiPM as exactly equal to the maximum, so when encountering an event where the limit has been reached, one deduce that some information from that event have been lost.

Figure 16 shows two histograms of the sum of pure limit events from setups 1 and 3. One would expect that the amount of limit events would increase with the energies. Yet

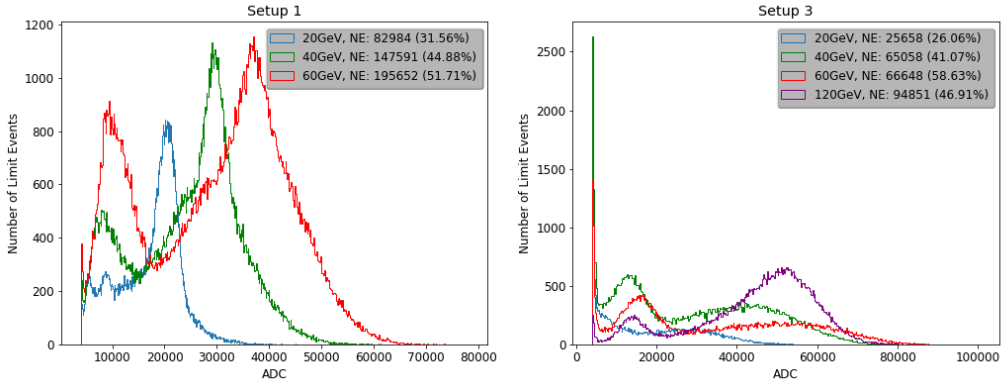


Figure 16: Illustration of ADC distribution of summed Limit Events for different energies for setup 1 (left) and setup 3 (right).

this does not necessarily seem to be the case, since the limit events from the 120 GeV plot should then made up a larger portion of the data set. However, the gain was reduced for the 120 GeV run, to avoid the boards overflowing, and this explains why there are fewer limit events in the 120 GeV run.

The large peak for 40 and 120 GeV roughly around 3920 ADC (4089 ADC minus the pedestal) on the right plot are all events where a single channel reached the limit. While all other were channels were zero. This is either a bug/feature of the boards or beam, or possibly, a particle traveled straight through the detector without interacting with the copper. The latter would then make it a muon. However they are nowhere to be seen in the left plot with setup 1. Setup 1 has the FoCal-H back-end tilted and this could be the explanation, why the sharp peak is not present.
Unexpectedly there is a high number of limit events in the lower ADC values which shows that only a few channels reached the threshold or came near it, in these particular events.

5.1.4 Empty events - Usable analysis data

It occurred that some of the events in the boards were completely empty in a few data sets, even void of the background noise left of the pedestal subtraction. This seems to be the effect of using the trigger system of one of the three other detectors from the Focal-E group. They do not have any impact on the event signal. But for good measure, if a given data set has any empty events, it should be removed.

Taking all of this into account, the actual amount of data one can analyze from each data set is:

$$AnalysisData = D \cdot M \cdot S \cdot L \tag{12}$$

Where D is the raw data before running the Collate.C program, M is the percentage of event after the merge, S is percentage of synchronization events, L is percentage of limit

events. These limit events lie within the range of what will be considered a true event, and removing them would leave the majority of the data set unusable for a statistical analysis. So it is necessary to keep them even though they do not carry the full information in those given events.

For future research design, if these are to be removed, they should be so before one subtracts the pedestals, since each channel has a different pedestal value. Notice that the pedestal subtraction does not remove any events from the data set, only raw background noise. Almost all of the reference to the data in this thesis is after the Collate.C program has run. This equation is only to illustrate how much data one loses throughout this entire process.

Another thing to consider is when looking at a specific setup, one should use the data set for the available energies within a 24 hour period, or at least a reasonable time frame. This should be done to avoid any unknown features coming from the change in setup or beam which might end up affecting the output. So in this thesis, if possible, when comparing energies within a specific experimental setup, all the data sets are with no change in setup in between the data have been taken.

5.2 The Two peaks - Data loss

5.2.1 Focal-E

After the noise reduction and the event removal, the first step was to see if Focal-H was able to distinguish the energy deposited in it.

The summed events from the beam energies 20 to 120 GeV are shown for the three Focal-E setups in figure 17, the left figure is setup 1, the middle setup 2, and the right is setup 3. Not all energies had available data sets for each setup, so the best suitable data sets for analysis was chosen. For setup 1 and 2, the energies deposited in Focal-H are clearly distinguishable. This was only possible because the gain was kept constant for each channel during all test runs, which made the different test runs comparable in ADC values. A change in gain will have an effect on the ADC values produced in each event. Such change would would render the test runs incomparable. This was the case for setup 3, where the 120 GeV had a lower gain setting than the other runs. This was done to avoid over flowing in the channels.

In the left histogram, with setup 1, there are two major peaks for every test run. For the peak at the lower ADC values, there is little variation, but for the second peak, it varies for each energy. It is evident that the higher energies produce higher ADC values. The top of the first peak of 80 Gev run is not shown here because the entries were simply too large, and the visual of the histogram would have been distorted. This, along with the fact that the second peak of 80 GeV is smaller and wider, suggests that Focal-H is reaching its limit beam energy wise.
The middle figure's setup was the same as setup 1, but without the tilt of Focal-H's back-

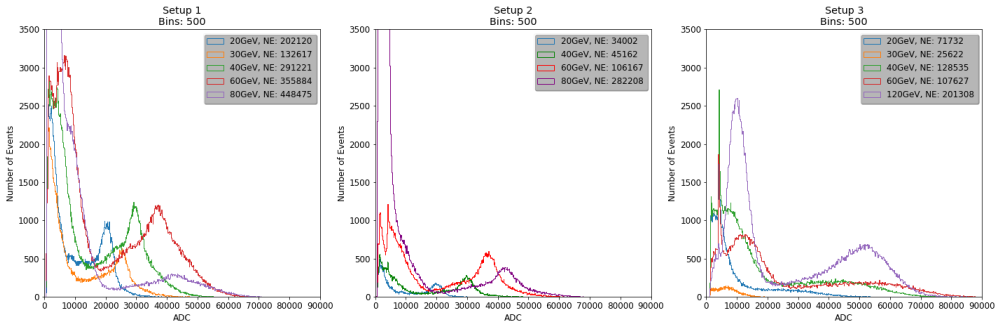


Figure 17: Illustration of difference in summed ADC output for the different energies. On the left, setup 1. In the middle, setup 2. On the right, setup 3.

end. The shapes are similar and the top half of the first peak from 80 GeV has been cut out like previously. The second peak seems to be slightly narrower than those from setup 1, though this could be due to the low amount of events. The runs from setup 2 had a low event sample, and, they look and behave similarly to those from setup 1. Henceforth in the analysis, focus will be on setup 1.
In the right-hand figure, the first peak is there, but the second peak is "flat" for all energies other than 120 GeV. Even though the 120 GeV had a lower gain and therefore produces lower ADC values, a clear second peak emerges. This illustrates perfectly the effect of interaction length of an absorber material. Also, the amount of tungsten in Focal-E was not enough to withstand 120 GeV beam. The ADC values for each energy are slightly higher than those of setup 1, which is peculiar since one would expect not only fewer events in the high ADC region but also slightly lower values due to the tungsten reducing the incoming energy. It could, however, be that the majority of electrons have been absorbed in the tungsten thus the flat second peak illustrates pure hadronic energy deposited in Focal-H. In addition, it should be noted in the figure 17 with setup 3 that, a high number of events is required before the flattened second peak starts to form. This is particularly evident from the runs for 20 Gev and 30 Gev. This suggests that, if one had enough time and resources to run these setups for a significantly longer time period, the peak would begin to form properly.

Some of the data set used for histograms with setup 1 and 3 is the same that was used in the limit event histogram in the section above. The shapes of the histograms are similar, though for setup 1, the first peak is less dominant in the limit events. This clearly indicates that the limit events lie within what would be considered true events. Even though, as mentioned before, information is lost in a limit event, it is still an approximation of the given shower. Removing them would render almost all data set insufficient for a proper data analysis. Hence, it was decided to keep all limit events for this analysis of the data from the first prototype of Focal-H.

5.2.2 EPICAL

The setups with EPICAL as an obstacle can be seen in figure 18. The left histogram with setup 4 and the right with setup 5. As evident the effect of setup 4 were similar to those of setup 3: a high peak for the lower ADC value and a flat second peak. However, there is no indication of the second peak for the runs below 60 GeV. Henceforth in the analysis, focus will be on setup 3 to represent setups with tungsten. However, the histogram on the right-hand side with setup 5 differs. It had a considerably smaller sample of events. This is due to the cross validation triggers was activated for the boards. Unfortunately, this was the case for all runs with setup 5. Cross validation should, in theory, only allow events where both boards detect something, which would of course amount to a smaller event sample, but also a peak within the true signal range. So the fact that the detector is tilted is not the cause behind the second peak.

When merging the two boards with Collate.C from a cross validation run, the number of events from the two boards should be exactly equal, hence no threshold for matching would be needed. However, this is not the case due to unknown reasons. But it suggests a lack of understanding of the CAEN boards features or settings. So, a loss in data when merging these cross validated data sets still occurs. The cross validation was used to create the few high matching (HM) rate data set, which will be discussed in the following sections.

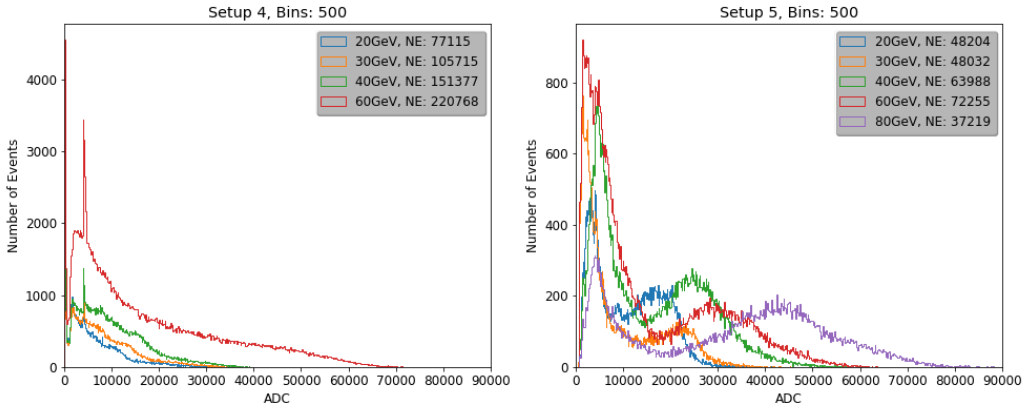


Figure 18: Histogram of the Sum of ADC Values for the different energies On the left histogram with setup 4 and the right with setup 5.

5.2.3 Board separation

When plotting the boards separately one can see that B0 solely contribute to the first peak, but there is still an unsettling amount from B1.This is illustrated in figure 19 where B0 can be seen on the left plot and B1 in the one on the right for setup 1. This further strengthens the suspicion from the section above, that if the shower is initiated in

the outermost channels, the detector loses the data from the aftereffect of matching events.

The other explanation for the lost information, which is definitely true for the higher beam energies, is simply that, due to the size of the detector, energy leaks out the the sides or out the back, because the length of the detector is to short.

The second peak from B1 are clear which means it is possible to fit and calculate the resolution of the detector from the perspective of B1. It should be considered whether it would be useful to simply cut the data from B0 and the data represented like that from B0 from B1, to avoid cluttering the left tail of the presumably Gaussian distribution of the second peak.
There is also a stark difference in numbers of events from each board, which makes sense since B1 contains all the inner channels. As the beam energy gets higher, the more events get triggered in B0. This suggests that the lower beam energies spread less through the channels than the high beam energies.

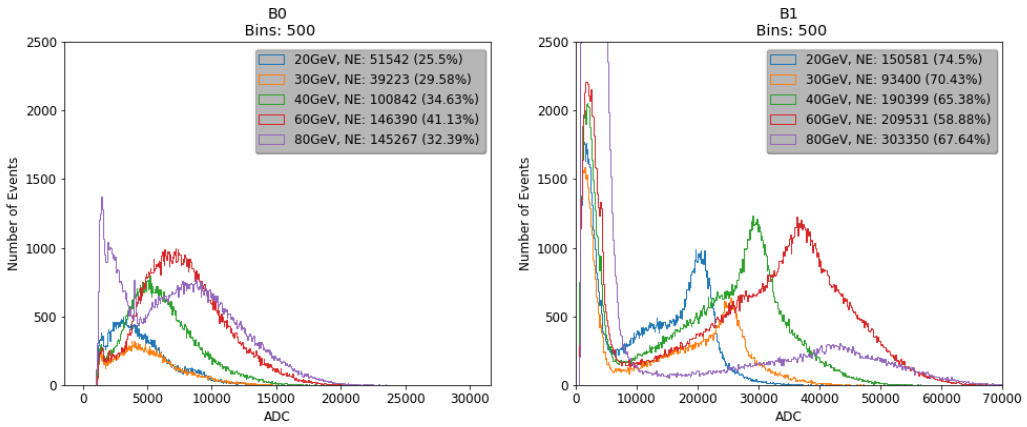


Figure 19: Histogram of the Sum of ADC Values for the different energies for each board with setup 1.

5.2.4 High matching data sets

Another way of working around this problem is to only look at events where both of the boards fired. This should, in theory, remove the first peak, as long as the shower is fully contained within the detector.

There were only two data set with over 50% of events with two boards active after the necessary events have been removed. There were other candidates, but due to most of the matched events being synchronization events, they are invalid. The best of them are shown on figure 20. This is for setup 5 with beam energy on 60 GeV. The other one has the same setup but with beam energy on 80 GeV.

For illustrative purposes, the histogram with all events is shown on the left, and on the right is the histogram where both boards were active for all events. This almost removes the first peak, and under the assumption that the second peak contains the entire shower, it begs the question whether the first peak is exclusively "unfinished" showers measured in the detector.

Note that the percentage in the right histogram shows the amount of events, with the left histogram as a reference, after the removal of synchronization events, which means that roughly 48% of the usable data was single-board events. As mentioned earlier, this should not have been the case since the runs were all cross-validation runs. Since, in theory, only events with two boards should be saved in these runs. So there is a clear flaw in the use of cross validation or in how an event is selected to be merged and matched.

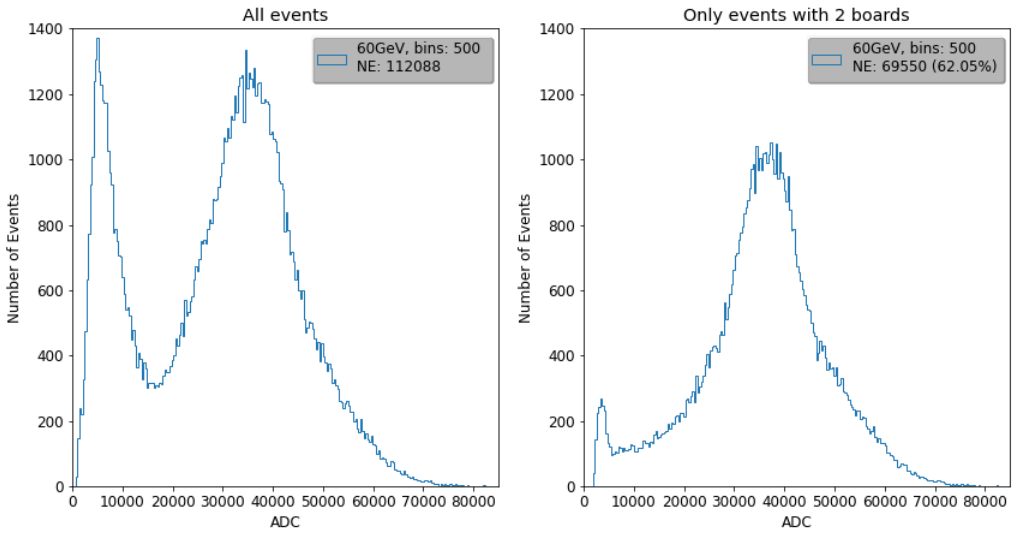


Figure 20: Histogram of the Sum of ADC Values for a HM data set with setup 5 and with beam energy on 60 Gev. Matching rate: 68.02% and of those events 35.73% was synchronization events the data. The histogram on the left shows all events from the test run and the histogram on the right is only event where both boards were active.

Since all data sets for setup 5 had cross validation on, only the HM data sets will represent setup 5 going forward.

5.2.5 Data loss

A lot of information is lost throughout the experiments and processes. The amount of data available for data analysis after the necessary conditions have been met can be calculated using equation 12. In this case, with the added rate of events with two boards and no

removal of limit events, it looks thus:

$$AnalysisData = D \cdot M \cdot B \cdot (1 - S) \tag{13}$$

$$AnalysisData = 1.00 \cdot 0.64 \cdot 0.77 \cdot (1 - 0.3573) \approx 0.32 \tag{14}$$

B refers the percentage of events with two boards fired simultaneously. Which leaves only around 32 % of the raw data ready to be analyzed. Looking back at the first example from figure 17, the data from the 60 GeV test run, here the reduction from running Collate.C, was only 4%, and if one only looks at B1, the amount of usable data is:

$$AnalysisData = 1.00 \cdot 0.96 \cdot 0.5888 \cdot (1 - 0.0595) \approx 0.53 \tag{15}$$

Here, B refers to percentage of event in B1. This is a significantly higher amount than in the other data set. Most noticeable is that there was a lot more synchronization events in the HM data set. This is because many of the matched events are synchronization event as mentioned in the section above. So the merge reduction from Collate.C varies strongly between the data sets, and there seems to be no connection between each different setup. The only thing that remains consistent is that the number of synchronization events grows with the matching percentage. So even though a data set is considered "good" because of the HM rate, it generally has a higher loss in data than those with lower matching. Therefore, when only looking at B1, one should use low matching data sets to retain as much data as possible.

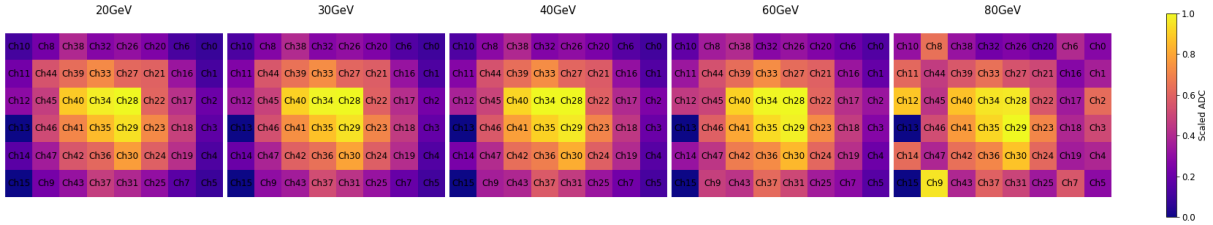
5.3 Shower development and Channel Activity

5.3.1 Channel Activity

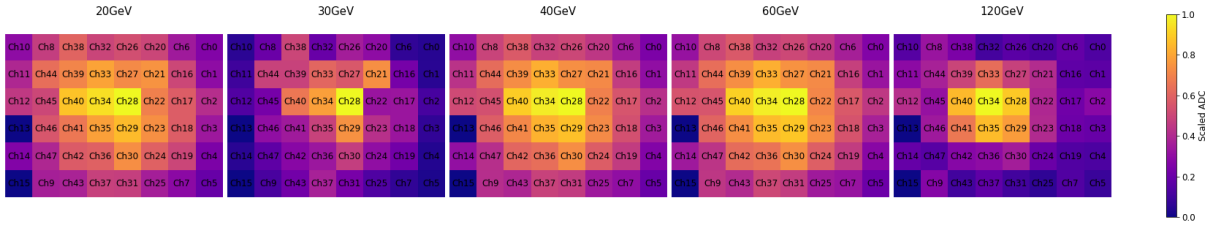
During the test run, not only the experimental setup were changed but also the beam characteristics such as the spill cycles and spill rate. The intensity map of each energy is shown in figure 21 for setup 1 and 3. The other setups' intensity maps can be seen in appendix B. These intensity maps are the sum of all events for each of the 48 channels. This is to show the behavior of the beam for the two different setups, to see how active each channel are, and to confirm the beam behaved the same or close to for all test runs for the best possible comparison.

The channel numbers can be seen in each square, B0 spans over the channels from 0 to 15, and B1 spans over the rest of the channels - 16 to 47. The intensity maps show the summed ADC value for each channel throughout the entire test run. It is scaled so the highest ADC value is equal to one, not only so they could share the same color bar, but also because the values would be ridiculous high otherwise.

At first glance, the intensity maps seems almost identical for the different energies in the detector where only 30 and 120 differ a little in figure 21b. This shows that the beam and the detector have behaved similarly for all these test runs.



(a) The intensity map for the five different energies with setup 1.



(b) The intensity map for the five different energies with setup 3.

Figure 21: The intensity map of the setups 1 and 3 of the beam in the detector. The ADC value are scaled such that the maximum ADC value equals to one. From left to right is the lowest to highest energy available from each setup.

The 30 Gev run had a significant lower number of events than the others, but it still picked up two things similar to the other energies. One is that the channel 28 had the highest ADC for all energies. In general, the four channels in the middle (ch28, ch29, ch34 and ch35) were the most active during the test runs. The other one is that channel 13 and 15 had a scaled ADC value around zero. The latter phenomenon could be caused by the channels simply being "dead" and therefore would not be recording any information. A channel can be dead either because the SiPM is not working or because the channel from the board is malfunctioning. All the SiPM's and channels were tested before the test runs, which means that the channel died after one of the changes in setup during the testing period.

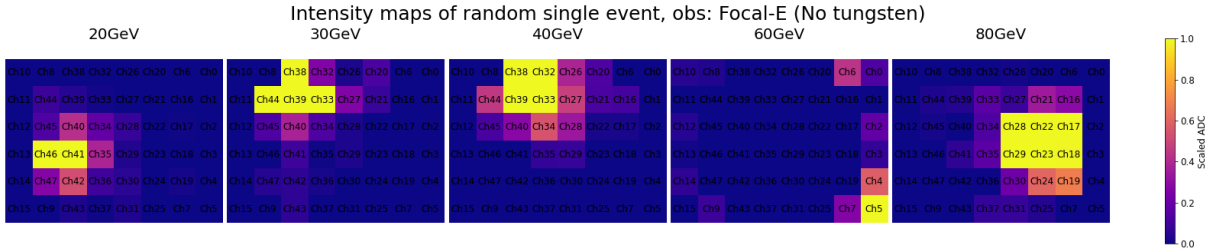


Figure 22: The intensity map of a single event for the five different energies with setup 1. The ADC value are scaled such that the maximum ADC value equals to one. The events were selected to illustrate as many channels as possible. From left to right is the lowest to highest energy available from each setup.

To see how a single event spread through the channels, the intensity maps of a single event can be seen in figure 22 for the five energies with setup 1. Notice here that no event spans over more than one board. This clearly illustrates the effect of poor matching between the boards. This is especially noticeable in the events for the 60 and 80 GeV where the edge channels and those around are active from the respective board, but no other channels from the other board are active. Under the assumption that the events should look similar to each other, it is clear that there is some sort of information lost to either leaks or the matching effect. Though the 60-GeV-one could also just be the effect of noise in B0.

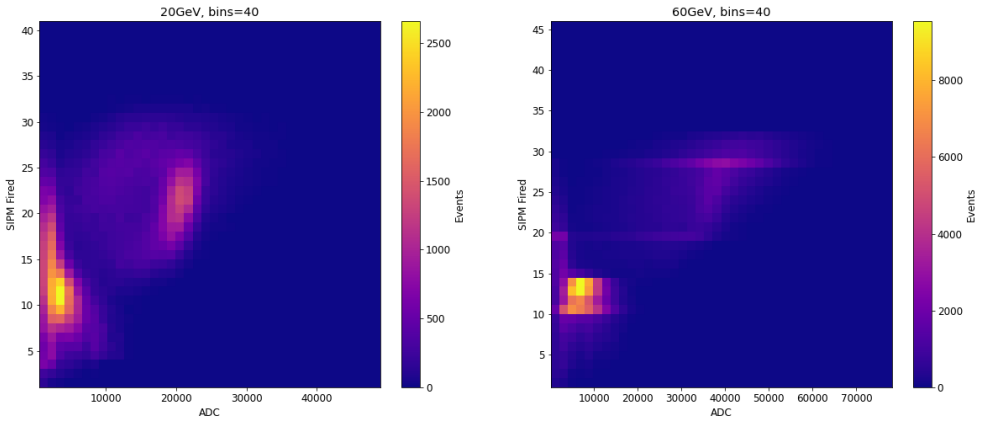


Figure 23: 2D histogram of SiPM fired vs the summed ADC value with setup 1 and beam energy on 20 and 60 GeV. On the left is the run with beam energy of 20 GeV and on the right is the run with beam energy on 60 GeV.

Another thing to look at is how many channels/SiPM fired for each event. This can be seen in figure 23, where a 2D histogram with the amount of channels/SiPM fired vs the summed ADC values for each event is shown for setup 1 with the beam energies 20 and 60 Gev. The first and second peak from figure 17 can been seen here as well in the two high-intensity areas.

For both runs, there are events with very low ADC values and still has more than 16 active channels. This must either be an event around the edge channels or it is background noise from an empty event. For the 60 GeV run, most of the higher ADC values are produced with more than 28 channels active. Since this is a low matching data set, it is most likely only B1 which is active in these events, compared to the 20 GeV run which has around 15 to 25 active channels each event. It further suggests that the width of Focal-H, or, in this case width of B1, might not be sufficient for higher beam energies and thus cannot contain the hadron showers fully. There are instances with fewer than 15 active channels for the 20 GeV run with high ADC value. This confirms that, for lower

energies, the entire shower can be contained.

As in the previous section, it could be interesting to look at each of the two boards separately, and how a HM data set would look like. In figure 24 the same plot for 60 GeV with the separation of the two boards is shown, B0 is shown on the left and B1 on the right. In B0, almost all channels are active for most of the events. This is odd since the channels are on separate sides of the detector and these ADC values lie above what would be considered background noise. Furthermore, there is almost none below eight channels, which means that almost every time B0 was active, most of its channels were active. B1, on the other hand, behaves as expected, with all channels active for most of the high ADC values.

Looking at figure 23, there is not a single event where more than 35 channels fired. Knowing from figure 24 that there were cases where all channels where active in B1, this further illustrates the effect of the poor event matching between the two boards.

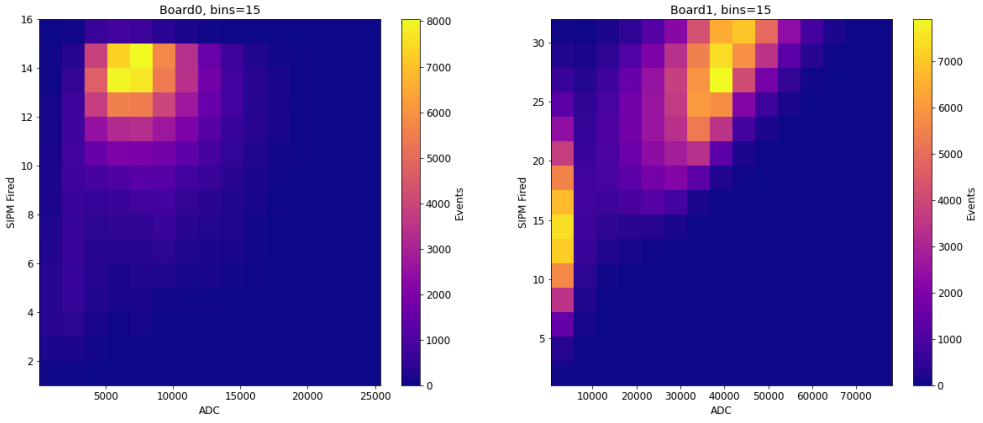


Figure 24: 2d histogram of SiPM fired vs the summed ADC value for B0 and B1. With setup 1 and beam energy on 60 GeV. On the left is B0 and on the right is B1.

When dealing with the HM data sets, it is expected that there would be more active channels in general. And looking at figure 25, one can see exactly that. On the left-hand side, the data set with all events is shown. On the right-hand side, only the events where both boards were active. There is an obvious difference in the number of active channels in these two plots compared to the others. The difference is that here, the intensity intensity map is stretched and the channel activity for the higher ADC values spans from 20 and up to 45 channels. This illustrates not only that a high matching data set provides the clear second peak but also that almost the entirety of the detector is utilized, which makes it a key necessity for future prototypes. However, do notice that there are still some cases with less than 30 active channels in the high ADC values.

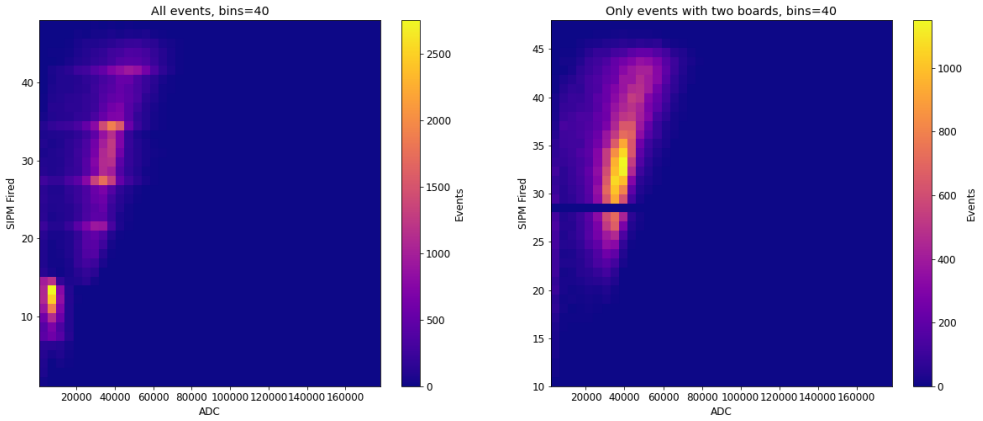


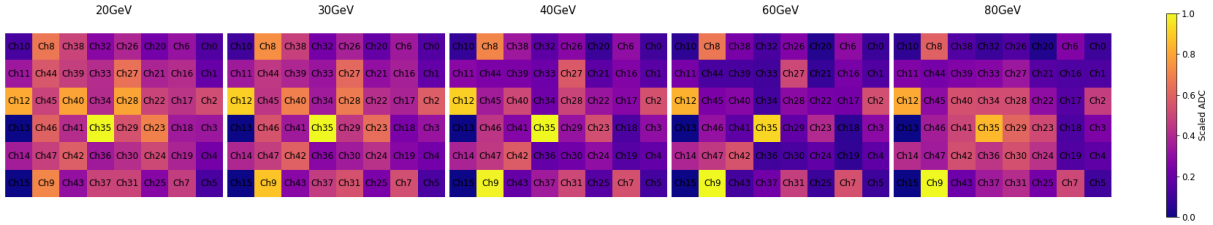
Figure 25: 2d histogram of SiPM fired vs the summed ADC value for the HM data set with setup 5 and beam energy on 60 GeV. The left plot shows all events and the right-hand plot shows only events where both boards were active.

5.3.2 Shower development in the channels

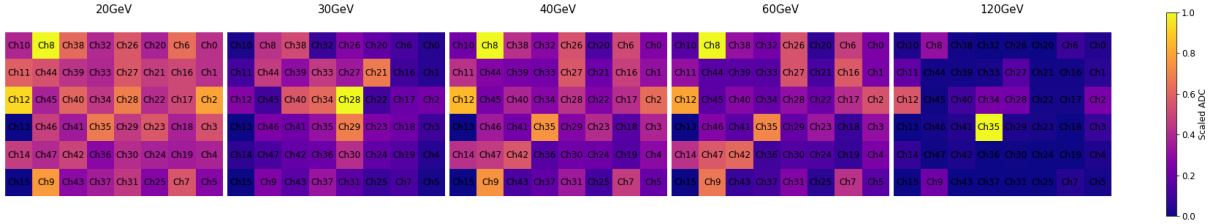
Another interesting thing to investigate is which channels initiated the most showers and how showers develop throughout the channels. Assuming that if a given channel has the maximum ADC value in a given event, then that would mean that the beam hit the given channel first, and a hadron shower was initiated in that channel.

An intensity plot for the count of every time a given channel had the maximum ADC value is shown in figure 26 for the energies from setup 1 and 3. Here, the counts have been scaled to enable comparison. At first glance, there seems to be no particular pattern other than that there is generally a lower intensity of counts in the middle on figure 26b than on 26a. This is presumably due to the effect of tungsten.

Another noticeable thing is that for most of the runs, channel 35 has the highest number of events. Since Focal-H was lined up parallel with the beam, it could mean that it was placed in such a way that channel 35 was directly in line with the beam. It could also be the effect of too high a gain in channel 35. This would make the channel produce higher ADC values than the other channels, and this would make it look like it has initiated more showers. The pedestal subtraction could also have had an influence on this, since the pedestal value varies for each channel. In that case, some channels with low pedestals values will have slightly higher ADC values than the other channels. Because the mean value was 129.79 ADC with a standard deviation of 10.31 ADC, this would only have an impact on channels with ADC values close to equal.
In figure 21, it was noted that channel 28 had the highest sum of ADC value out of all ten runs, and that the channels next to it, channel 29 and channel 34, were equally high in summed ADC values. Then it is odd that channel 35 has such a high intensity compared



(a) The intensity map of each channel with max ADC for a given event, for the five different energies with setup 1.



(b) The intensity map of each channel with max ADC for a given event, for the five different energies with setup 3.

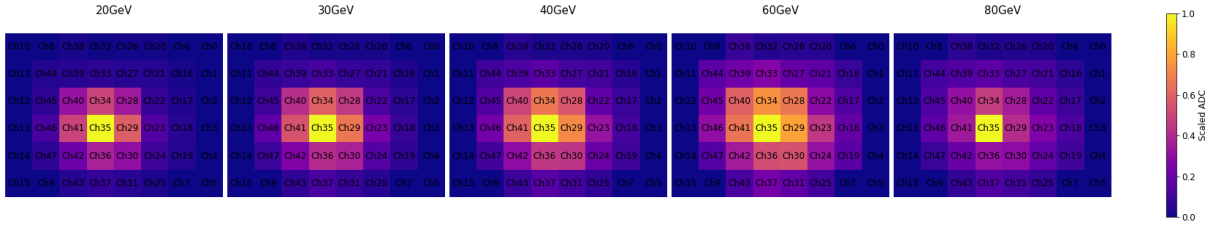
Figure 26: The two different Focal-E setup's intensity map of each channel with max ADC for a given event. The ADC value are scaled such that the maximum ADC value equals to one. From left to right is the lowest to highest energy available from each setup.

to those three in figure 26.

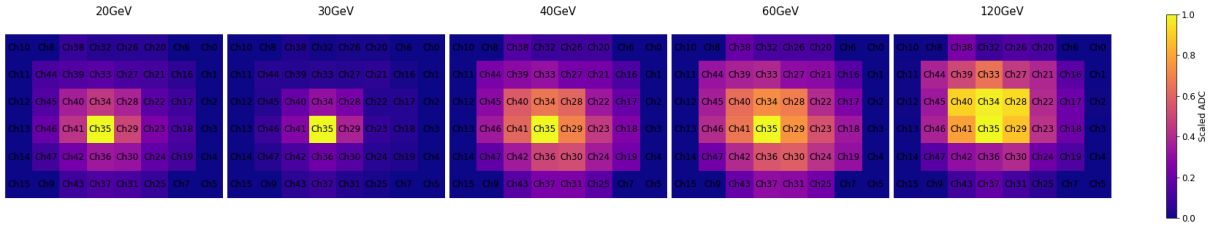
The most peculiar thing in these plots is that the 120 GeV setup, - here channel 35 - is clearly the dominating channel and almost the only channel which has the maximum ADC value in an event. Here, it might have been the effect of pedestal subtraction, since if multiple channels reached their limit every event, then the one with lowest pedestal value would be the maximum for all these events. As will be seen below, this is most likely the case.

The run from Setup 3 with beam energy on 30 GeV may belong to the same family of oddities. It had a significantly lower number of events which could be why it looks so different from the others. It is also possible that some sort of movement or relocation of the detectors happened before this specific test run, which would also explain why channel channel 28 is the most active. But this is highly unlikely since there is no entry of this in the E-log.

The edge channels from B0 tend to to have a higher number of events with the highest ADC value than B1's edge channels. A few examples include the channels 8, 9 and 12, which are surprisingly active in a majority of the plots. This could be due to noise in B0, for which reason they may not reflect the actual number of initiated showers. The mean values of the events with the respective of these channels, are: $Ch8 = 2789.34$ ADC, $Ch9 = 2676.57$ ADC and $Ch12 = 2805.616$ ADC. These examples are taken from setup



(a) Intensity maps of summed ADC values where channel 35 had the highest ADC value, for five different energies with setup 1



(b) Intensity maps of summed ADC values where channel 35 had the highest ADC value, for five different energies with setup 3.

Figure 27: The intensity map of summed ADC values for setup 1 and 3, where channel 35 had the highest ADC value in all events. The ADC value are scaled such that the maximum ADC value equals to one. From left to right is the lowest to highest energy available from each setup.

1 with beam energy on 60 GeV, but the the other data sets provided mean values within the same range. The assumption that if the channel had the highest ADC value in a given event it would be where the shower started, is actually still valid for the majority of the events from these highly active outer channels. This means the impact area of the beam covers most of the detector. These three outer channels have roughly around 6.4% of the events in that given data set, slightly lower than that of channel 35 which had 7.1% of the events.

From figure 26, the best candidate to look closer at is channel 35. The intensity map of the summed ADC values for every event where channel 35 had the maximum ADC value is shown in figure 27 for the same ten test runs. the top, figure 27a, shows setup 1 and the bottom, figure 27b shows setup 3. The energy spread throughout the detector is as expected,- the higher the energy, the wider the spread.

Again, setup 3 with beam energy on 30 GeV is slightly different than the others, which is expected due to the low sample size. The other energies from both setups all have similar appearances, other than setup 3 which seems to be slightly wider. The 120 GeV run have two other channels that are highly active, and looking back at figure 26b, this indicates even further that the reason channel 35 is almost the only one initiating showers is due to the pedestal values.

5.3.3 Effective Channels

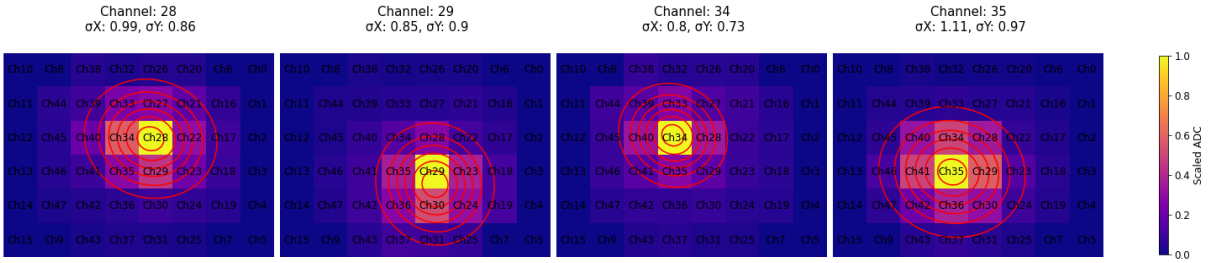
Even when focusing on B1, there is still a large first peak for most of the data, which presumably comes from the edges between B0 and B1 due to the shower spreading between the boards. As one saw from figure 27, the amount of active channels in an event increases with the energy. So there must be a different amount of "effective channels" for each energy. Where "effective channel" means that if the hadron shower starts in that given channel then the shower is contained within the detector. This is still following the assumption that if any given channel has the maximum ADC value in a event, then that is where the hadron shower was initiated in that event.

To check how many channels are effective, the width of the shower has to be approximated. From figure 21, the best bet for an effective channel would be the four channels in the middle; channel 28, 29, 34, 35. Henceforth, the effective channels and all the events they represent, will be referred to as channel "xx". Thus, "channel 35", represents all events where channel 35 had the maximum ADC value.

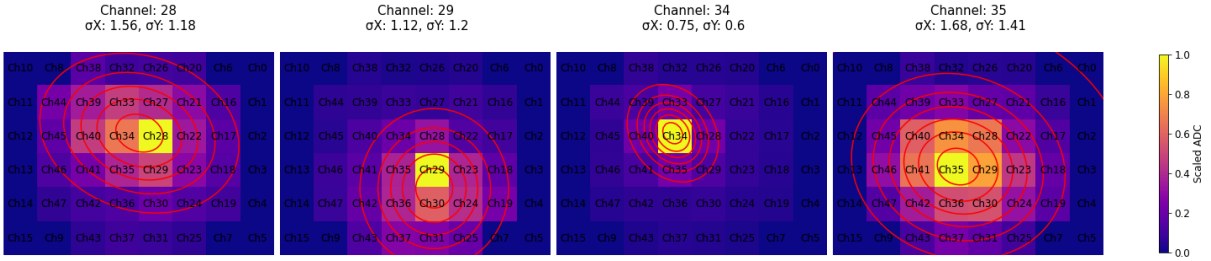
A 2d Gaussian fit is made to figure 27 and of the other presumably effective channels from B1.iThis is shown on figure 28. Here, the energies 20, 60 and 80 GeV are shown for setup 1. The data with 30 and 40 GeV, as well as the energies for the other setups can be seen in appendix C. The red rings are the 2d Gaussian. The third ring has the radius of one σ. The 2d Gaussian fits are not ideal, but they are a decent approximation. The sigma value from the setups 1 and 3 can be seen in table 2, and as one can see in figure 28, the radius of the showers spans around 2σ. The center of the 2d Gaussian fits is not directly in the middle of the channel in question. This means that the neighboring channels did have a high activity and impact on the fits. Even if assuming the radius is from the center of the given channel, then for all beam energies with a $\sigma > 1.25$ there is likely to be some sort of information lost. Ideally, calculating the mean σ of the four fits should allow for a

Setups:		Setup 1					Setup 3				
Ch/Energy (GeV)		20	30	40	60	80	20	30	40	60	120
28	σ_X	0.99	1.16	1.35	1.56	0.50	0.86	0.62	1.24	1.35	1.69
	σ_Y	0.86	0.95	1.06	1.18	0.66	0.69	0.8	0.92	1.01	1.35
29	σ_X	0.85	0.9	1.05	1.12	0.68	0.82	0.66	0.69	0.88	1.23
	σ_Y	0.9	0.97	1.11	1.2	0.45	0.62	0.78	0.87	0.74	1.64
34	σ_X	0.8	0.8	0.81	0.75	0.44	0.76	0.83	0.77	0.8	0.83
	σ_Y	0.73	0.71	0.72	0.60	0.60	0.58	0.69	0.63	0.66	0.69
35	σ_X	1.11	1.27	1.42	1.68	1.05	1.27	0.8	1.91	2.18	1.76
	σ_Y	0.97	1.09	1.20	1.41	1.18	1.06	0.75	1.62	1.86	1.52
Avg	$\overline{\sigma_X}$	0.94	1.03	1.16	1.28	0.67	0.93	0.73	1.15	1.30	1.38
	$\overline{\sigma_Y}$	0.87	0.93	1.02	1.10	0.72	0.74	0.76	1.01	1.07	1.30

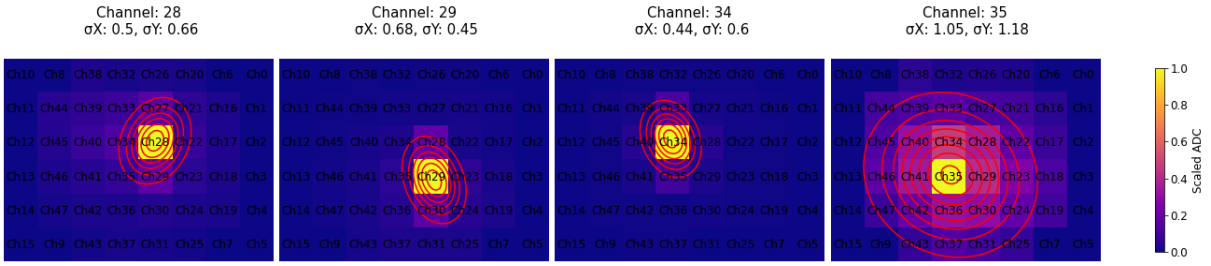
Table 2: The σ values in the X and Y dimensions for the 2d Gaussian fit of the two Focal-E setups, setup 1 and 3.



(a) Beam energy: 20 GeV.



(b) Beam energy: 60 GeV.



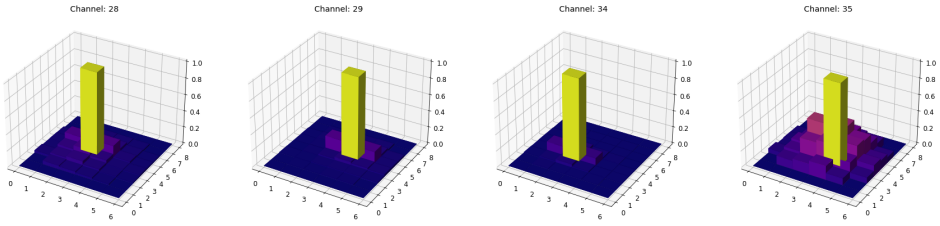
(c) Beam energy: 80 GeV.

Figure 28: The intensity map of summed ADC values where the presumably effective channels (ch28, ch29, ch34, ch35) had the highest ADC value with a 2d Gaussian fit for setup 1 for the energies 20, 60 and 80 GeV. The ADC value are scaled such that the maximum ADC value equals to one

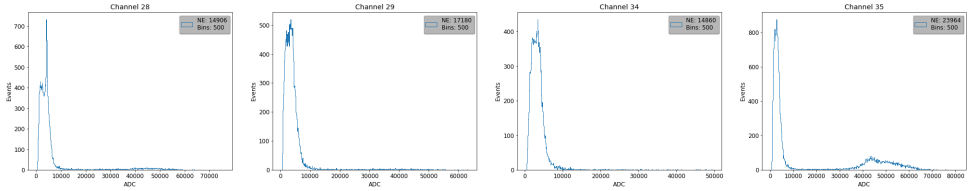
decent approximation for the shower radius for that given energy and, in theory, the channels should behave the same. Yet, channel 34 has considerably smaller σ's than the other three channels. This could mean that there is something wrong with the channel itself or that it is the effect of the pedestal value being higher than the others. The pedestal values for the four channel are: ch28=125.517, ch29=130.709, ch34=136.198, ch35=123.406, which indicates the latter. As mentioned, this would then make it less likely to have the maximum in a given event and would thus create fewer event for channel 34.
In this case, the mean value of the four σ would not yield a proper result due to the channel's inconsistent behavior, and the largest σ in each direction would be the safest approximation of the shower radius. This will be applied for all runs, but it is worth considering evaluating each channel shower radius in the detector for future prototypes.

The mean σ can, however, be used to compare the setups, since arguably all of them have the same inconsistencies. Looking at the mean for 20, 40 and 60 GeV test runs for both setups, the mean σ are within the same range. The low sample size from the 30 GeV test run from setup 3 had an effect here as well, and it is considerably lower than its peer. The 120 GeV run has the largest mean σ which is expected.

The σ vaules from the 80 GeV run is to low compared to the others. When looking at figure 28c, for the channels 28, 29, and 34 only the channel itself lights up and it seems like no showers were actually induced in those channels.
To see whether that was the case, the intensity map with a 3d projection can be seen in figure 29a. Here, it illustrates that it is almost only the channel in question that contributes to the ADC value in each event. When plotting the sum of ADC values for those



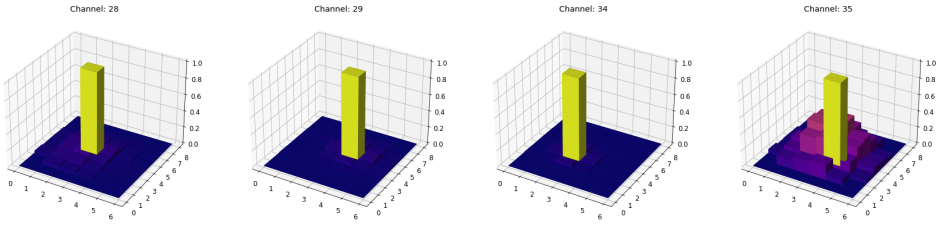
(a) 3d projection of the intensity map of the four inner channels.



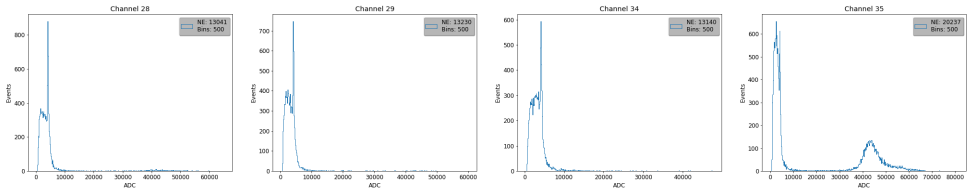
(b) Sum of ADC Values for the four inner channels.

Figure 29: The two plots are from the 80 GeV test run from setup 1. The top plot shows the 3d projection of the intensity map and the bottom one shows the sum of the ADC values for the four inner channels. From left to right is Channel 28, 29, 34, and 35.

specific channels, the ADC value lies within the range of the first peak which is concluded to not be a true event This is shown in figure 29b. The second peak is barely visible for channel 35, and it suggest that if the test run had been longer a peak suitable for fitting might have formed. Since the Setup 1 and Setup 2 80 GeV run both had a enormous first peak in figure 17. The same two plots are shown for setup 2 in figure 30 .
As can be seen, the setup 2's 80 GeV run behaves the same as setup 1's 80 GeV run. These two runs were taken on two separate days and this indicates that this might be an effect of the beam at 80 GeV, and not the three channels malfunctioning during the test run. The latter would have explained why the first peak was that much higher in figure 17 than the other energies. But at this moment the specific reason to why the 80



(a) 3d projection of the intensity map of the four inner channels.



(b) Sum of ADC Values for the four inner channels.

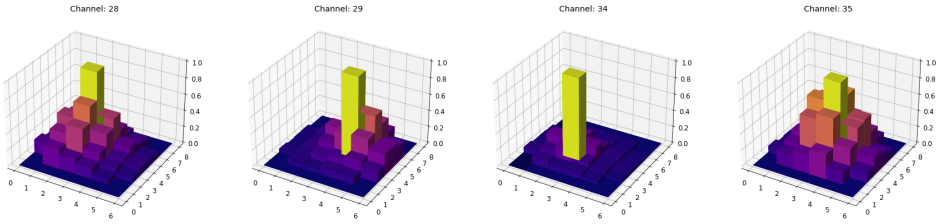
Figure 30: The two plots are from the 80 GeV test run from setup 2. The top plot shows the 3d projection of the intensity map. The bottom one shows the sum of the ADC values for the four inner channels. From left to right is channel 28, 29, 34, and 35.

GeV runs for setup 1 and 2 behaved so oddly is unknown. Channel 35 for setup 2 does, however, have a slightly larger second peak than the one from setup 1, even though it has a lower sample size.

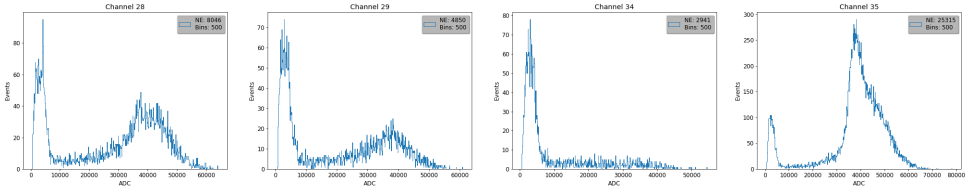
To compare the two "failed" runs with a another test run, the same plot for the 60 GeV run with setup 1 is shown in figure 31. Again, channel 34 has no second peak, but the three other channels do. - With a much more apparent second peak, it does seem like channel 35 is the candidate for the best effective channel.
The channels 28 and 29 have far less active channels, both in ADC values and in the number of channels active each event. This was also viable in figure 28b, but this plot illustrates it more clearly. This could be due to insufficient statistics compared to channel 35, but fitting a Gaussian might not yield the best result for those two channels.
Furthermore, when looking at channel 35 for figure 28b, it seems like a "bump" is forming on the right side of the second peak, which indicates the start of a third peak. It could be that something changed with time during the test run, either the beam or the detector itself. But plotting parts of the data had the same indication of the "bump". If the detector is not compensated, the visible energy from the electromagnetic lies outside the range of the hadronic signal, which could then be an explanation for the "bump". There is not yet a clear explanation of what this is. This "bump" appears in other runs as well, but mostly for channel 35, and primarily in runs where there is no tungsten in the setup.

Looking at figure 28, figure 31, and the table 2, one could argue that channel 34 should

not be considered an "effective channel" due to its inconsistency and because seemingly no showers have been induced in that channel.



(a) 3d projection of the intensity map of the four inner channels.



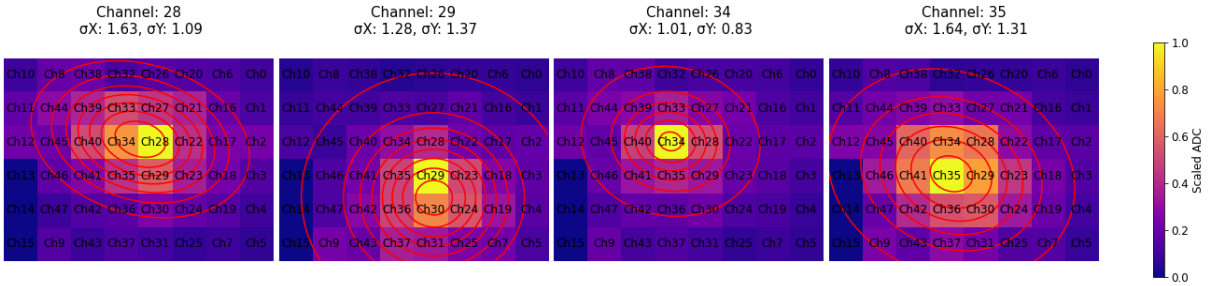
(b) Sum of ADC Values for the four inner channels.

Figure 31: The two plots are from the 60 GeV test run with setup 1. The top plot shows the 3d projection of the intensity map. The bottom one shows the sum of the ADC values for the four inner channels. From left to right is channel 28, 29, 34, and 35.

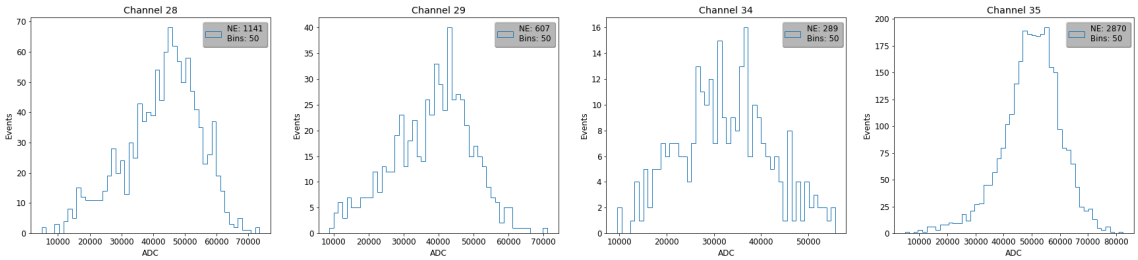
In general, channel 35 has the highest σ in both directions. Using channel 35 as a reference for the radius of the showers, no other channels outside the middle can fall under the category "effective". since this is just an approximation of the radius, then in some cases one could allow for more effective channels. And if a proper second peak forms when plotting the sum of ADC values for that given channel, it would be possible to calculate both the resolution and ADC translation.

By using the same argument, HM data set should allow for more effective channels, or at least a better approximation of the true radius of the shower in the detector.
A 2d Gaussian fit was made for the two as well. The one for the 60 GeV test run can be seen in figure 32, where both the intensity map and the sum of ADC can be seen for the four inner channels. The first thing to notice is that these intensity maps confirm the suspicion from before that the channels in the bottom left corner are dead. Unlike earlier, where one could argue that there was low activity in the channels due to low match rate, this is not the case here.

Even though the number of events from the effective channels are too low to perform proper statistical evaluation, it still provides an estimate of the radius of the shower in the detector. The σ values for both the 60 and 80 GeV test run can be seen in table



(a) The intensity map of the four inner channels with a 2d Gaussian fit.



(b) Sum of ADC Values for the four inner channels.

Figure 32: The two plots are from the 60 GeV HM data set with setup 5. The top plot shows the intensity map with a 2d Gaussian fit. The bottom one shows the sum of the ADC values for the four inner channels. From left to right is channel 28, 29, 34, and 35.

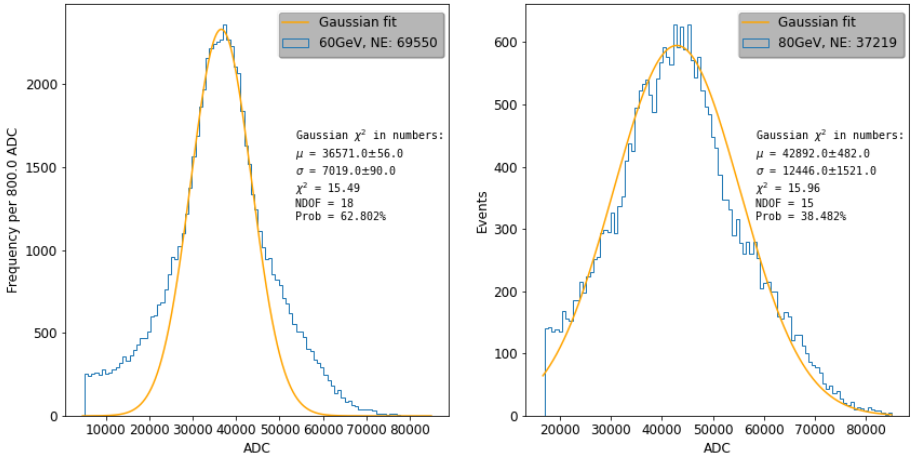
3. Again, channel 34 has lower σ which further speaks in support of its dismissal as an effective channel. The mean σ from the 60 GeV run is slightly higher than those from the Focal-E setups which shows that the shower spans over more than B1 for beam energy above 60 GeV. Figure 32b shapes are similar to figure 20 which suggests that the assumption of effective channels is true.

It could be interesting to see the energy distribution from the edge channels and compare them to the center channels. It is not possible for the HM data sets due to the low number of events in the edge channels which, when plotted, cannot provide any significant visual in a histogram.

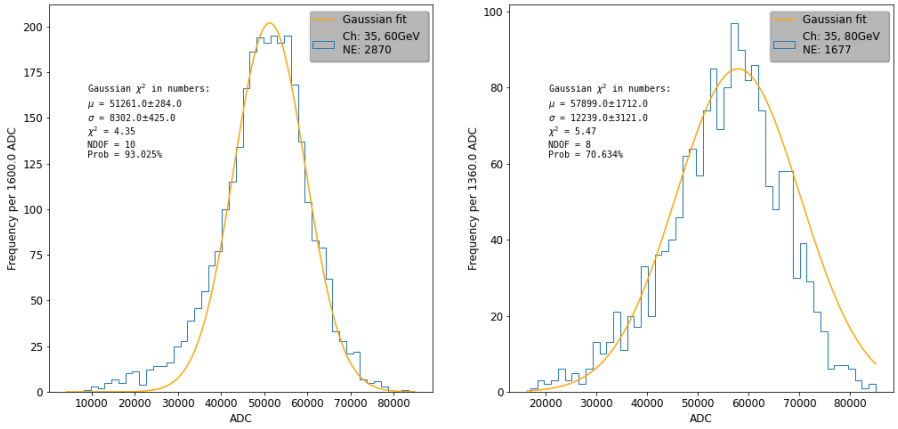
5.4 ADC Translation and Energy Resolution

To calculate the resolution of the detector and the translation from ADC to GeV, a Gaussian fit was made for the summed ADC values first for the two HM data set. To compare these and the other setups, a fit was made for the effective channels as well. In figure 33, the fit is shown for the full detector and of channel 35. The fit for the two other effective channels can be seen in appendix D.

The fitted μ value would then be an approximation of the ADC translation to GeV and the resolution can be calculated from equation 9 using the μ and σ from the fits. The



(a) Histogram of the Sum of ADC Values with a Gaussian fit. The histogram on the left shows the data for beam energy on 60 GeV. The right histogram shows the data for beam energy on 80 GeV.



(b) Histogram of the Sum of ADC Values of channel 35 with a Gaussian fit. The histogram on the left shows the data for beam energy on 60 GeV. The right histogram shows the data for beam energy on 80 GeV.

Figure 33: Histogram of the Sum of ADC Values for the two HM data set with setup 5.

Setups: Ch/Energy (GeV)		Setup 5 60	80
28	σ_X	1.63	1.78
	σ_Y	1.09	1.20
29	σ_X	1.28	1.23
	σ_Y	1.37	1.35
34	σ_X	1.01	1.00
	σ_Y	0.83	0.79
35	σ_X	1.64	1.81
	σ_Y	1.31	1.42
Avg	$\overline{\sigma_X}$	1.38	1.50
	$\overline{\sigma_Y}$	1.15	1.19

Table 3: The σ values in the X and Y dimensions for the 2d Gaussian fit of the two HM data sets with setup 5

ADC translation provides a realistic energy perspective compared to the arbitrary ADC values. Since the ADC values are arbitrary, it means that if some settings on the boards had been changed between the setups, then, for example, two different 20 GeV runs could have very a different ADC translations. Therefore, it is inappropriate to compare the translation between the different setups, and the translation is only possible if the gain has been kept constant throughout all test runs with a specific setup. It would be possible to calculate the maximum beam energy Focal-H could measure with the given gain setting.

The Gaussian fit is made to the top half of the histogram peak. This reduces the number of degrees of freedom and it narrows the Gaussian distribution so the left and right tail contributes less to the overall fit. The left tail is mostly the remains of the first peak, and that is why the histograms in figure 33a have the abrupt start. The right tail is valuable signal and in some cases it is lost as a repercussion of avoiding the left tail. A small example of this can be seen in the left histogram in figure 33a. The choice of making the fit like this might give a too great resolution in some cases, hence it is important to evaluate the fit in each case.

The goodness-of-fit or χ^2 lies within an acceptable range for the HM fits. The fit to channel 35 is exceptionally better than the fit to the entire detector for both energies. The entire peak has been shifted to the right for channel 35, which can also be seen in the difference in their μ value. This could be the effect of cutting the right tail, but it is more likely that the high ADC values for the histogram of the complete detector come from the effective channels, and this means that energy lost from the outer channels factored a lot when calculating μ. So when translating ADC values to GeV, using the effective channels as reference would yield the most precise results possible. This would also be the case for the resolution of the detector, since it is dependent on μ as well.

Setup 1						
Ch/Energy (ADC)		20 GeV	30 GeV	40 GeV	60 GeV	80 GeV
28	μ	20620.0 ± 57.0	25719.0 ± 126.0	30630.0 ± 94.0	39632.0 ± 344.0	
	Resolution	0.092 ± 0.004	0.095 ± 0.009	0.085 ± 0.006	0.201 ± 0.019	
29	μ	20699.0 ± 87.0	25165.0 ± 202.0	29745.0 ± 170.0	36700.0 ± 453.0	
	Resolution	0.072 ± 0.008	0.107 ± 0.015	0.115 ± 0.011	0.168 ± 0.027	
35	μ	21351.0 ± 38.0	25938.0 ± 53.0	30230.0 ± 42.0	37951.0 ± 85.0	45711.0 ± 315.0
	Resolution	0.072 ± 0.003	0.063 ± 0.004	0.071 ± 0.002	0.087 ± 0.005	0.20 ± 0.012
Setup 3						
Ch/Energy (ADC)		20 GeV		40 GeV	60 GeV	120 GeV
28	μ					42976.0 ± 213.0
	Resolution					0.256 ± 0.009
29	μ					30453.0 ± 1209.0
	Resolution					0.500 ± 0.098
35	μ	34212.0 ± 656.0		51365.0 ± 503.0	59420.0 ± 339.0	51576.0 ± 69.0
	Resolution	0.333 ± 0.032		0.198 ± 0.025	0.177 ± 0.010	0.150 ± 0.002
Setup 5 (HM)						
Ch/Energy (ADC)					60 GeV	80 GeV
All	μ				36571.0 ± 56.0	42892.0 ± 482.0
	Resolution				0.192 ± 0.002	0.29 ± 0.036
35	μ				51261.0 ± 284.0	57899.0 ± 1712.0
	Resolution				0.162 ± 0.008	0.211 ± 0.054

Table 4: The calculated μ and Resolution for the three setups for all possible energies and channels.

The fits not shown here can be seen in appendix D, and the different μ values and the energy resolutions from the fits with setup 1, 3, and 5 are shown in table 4. The σ values are not shown in any table, but can be seen on each fitted histogram in the appendix. The table shows the μ values (the ADC translation) and the energy resolution (in red) with a absolute error calculated by error propagation [36].

For some of the setups, there are cases where the second peak did not form properly due to insufficient numbers of events, which makes the fit fail. This was especially visible in the 80 GeV run, but it was also the case for most of the runs for setup 3. Unfortunately for setup 3, only a fit-able peak from channel 35 formed for 20, 40 and 60 GeV and none at all for 30 GeV. This was expected from figure 17, where the energy distribution was flat in the range of the second peak for all others than the 120 GeV run. This further advocates that channel 35 had the best overall performance in the test runs. Many of the effective channels had less than 10000 events and the few cases with more were always channel 35.

The two first things to notice are that the μ values all increases with the energies as expected and for setup 1 the μ values are close to equal for the different channels, and that, if the 80 GeV run is not counted, this means that the channels behaved the same and consistently throughout all these test runs. The mean value of the μ's for each energy would then be a decent approximation for the ADC translation for this setup.

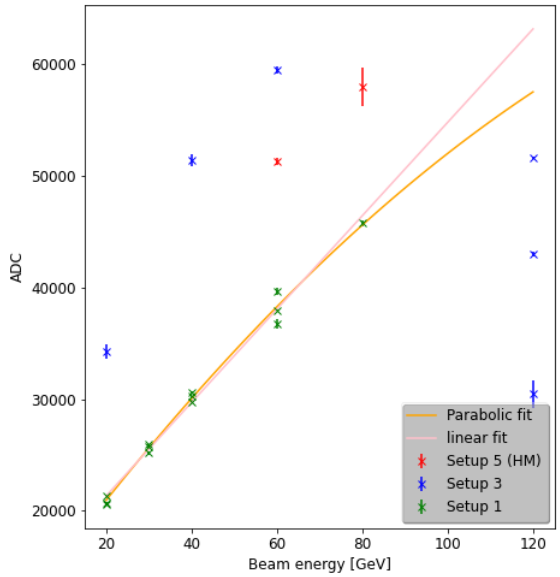


Figure 34: A scatter plot of the μ values and their respective energy for the 3 setups with a parabolic and linear fit.

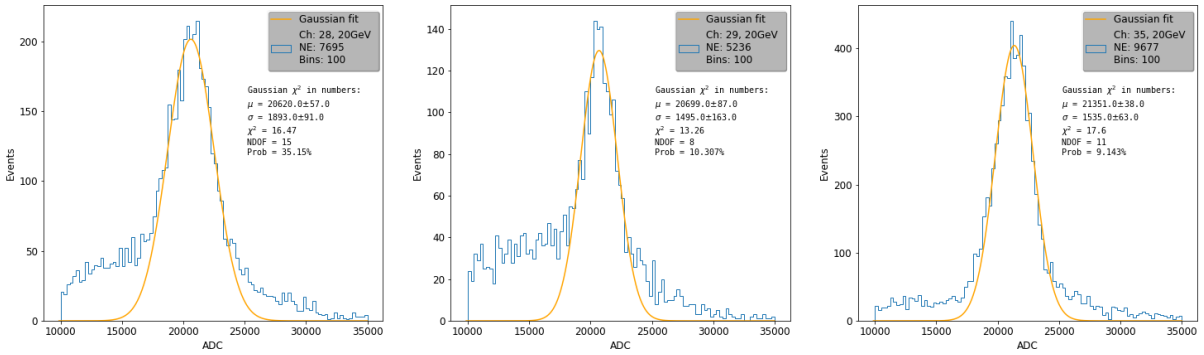
For setup 3, the μ values and their respective errors are larger by a high margin compared to setup 1, and the difference increases with the energy. It was noted from figure 17 that setup 3 had some events with much higher ADC values that setup 1. With the low number of events - even for channel 35 - these high ADC value events carry even more weight when calculating the μ. As mentioned, there was only one of the effective channels which was fit-able. So it was not possible to compare and validate the high μ values with other channels.

The 120 GeV run had all three effective channels, and the μ values are within rather large range of 30000 ADC to 50000 ADC. This shows that the channels behaved very differently compared to what was expected from the σ values from the 2d Gaussian fits on table 2. This inconsistent behavior is most likely due to the size of the detector being insufficient in length and in width, which has been mentioned several times before. It could also be that the channels did not have the same adjustment in gain, hence the wast difference in μ's. This also affects the resolution and it differs with the μ value.

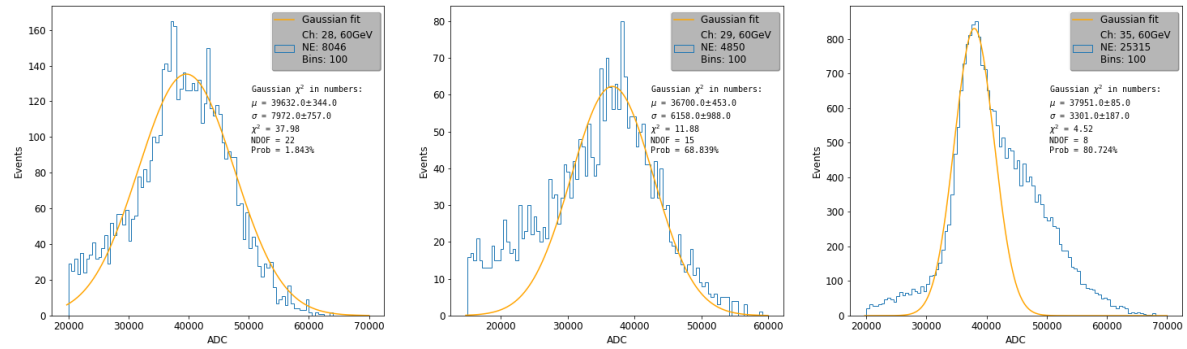
The μ distribution for the different channels and energies is shown as a scatter plot in figure 34. This scatter plot visualizes the μ values, and it shows where either the data set or the fit itself lacks. The errors are almost negligible and are barely visible in the plot. Here, it perfectly illustrates the difference in ADC value for each setup and that an global ADC translation for all setups cannot be done. A quick linear and a parabolic fit were made from the mean vales of setup 1's μ. The μ's from setup 1 was consistent

between the different channels. Arguably, the setups 3 and 5 μ's seem to have a consistent pattern as well, but unfortunately too few comparable values. So it did not not meet the requirement for a proper fit. The parabolic fits should be the better fit. This is no surprise since the channels have a limit. So at higher energies, it would be expected that the μ values would start to become constant due to all channels starting to overflow. The parameters for the parabolic fit for setup 1 are: $-1.12 \cdot x^2 + 522.67 \cdot x + 10938.85$. Where the y is the ADC value and x is the beam energy in GeV. Using the fit parameters as an approximation it is possible to calculate how high the beam energy can go before the μ values becomes constant with the current gain settings. The top of the parabola would then be the maximum ADC value the detector could measure and this would result in a maximum beam energy of: 233.46 GeV for setup 1.

The calculated resolution is significantly better in setup 1 compared to the other two



(a) Beam energy: 20GeV. From left to right is Channel 28, 29 and 35.



(b) Beam energy: 60GeV. From left to right is Channel 28, 29 and 35.

Figure 35: Histogram of the sum of ADC values for the effective channels with a Gaussian fit for beam energies of 20 and 60 GeV with setup 1.

setups, especially for channel 35. It provides a far greater energy resolution than any other channel or setup, and this suggest that the choice to fit the top half might yield an unrealistically good resolution in these cases. Looking at figure 35 where the fits for 20 and 60 GeV for setup 1 are shown. The figure shows that the loss of in information

from the right tail grows with the energy. For the 20 GeV run, the peaks are narrow, the right tail is almost negligible and the calculated energy resolutions are similar between the three effective channels. This suggests that the fit conditions for this run were sufficient. But for the 60 GeV run, there are clearly some issues for channel 35. As mentioned, it has a "bump" on the right side of the peak which creates a rather large right tail, and this "bump" is not included in the fit. This results in a too small σ for the ADC distribution, and this would yield too good a resolution. This primarily happened in setups without tungsten, and this fact should be kept in mind when comparing the results with the simulation, other setups and future results.

For setup 2, not counting the 120 GeV run, the resolution becomes better with the increase in energy, which is expected for a calorimeter. The 120 GeV had decent fits for all channels, but there is a vast difference in resolution between the channels. Again, it was channel 35 that provided the best resolution.

For setup 5, the 60 GeV run's resolution and errors are significantly better than the 80 GeV run in both cases. This was also evident from the goodness of fit in figure 33. The energy resolution from channel 35 from the setups' 3 and 5 are similar and would better estimates than the estimated energy resolution from channel 35 for setup 1.

5.5 Simulation

Simulation similar to setup 1 from the test runs of the first prototype of Focal-H was made on C2 to compare the calculated resolution. It was decided to only show the simulation of setup 1 since it had the most variety in results. In the simulations, it is possible to distinguish between the hadronic and electromagnetic part of the energy deposited in the detector, and therefore, one is able to check if the detector was compensated with current ratio of passive and active media. The compensation ratio can be calculated by e/h, and since the fluctuation in the electromagnetic part of the shower are non Gaussian a landau fit was done for that. So, the compensation ratio can be estimated by the highest peak from their respective fits [24]:

$$\frac{e}{h} \approx \frac{landau_{peak}}{Gauss_{peak}} \tag{16}$$

The energy resolution as well as the compensation ratio was estimated and can be seen in table 5. The histograms, fits, and the calculated fit values used for the calculation of the energy resolution and compensation ratio can be seen in appendix E.

The compensation ratio comes closer to 1 when increasing the energy, but the compensation ratio is still too low. This means that the visible energy of the electromagnetic part of the hadron shower is outside of the hadronic signal. Since it is lower than one, it means that the hadronic energy of the shower is dominant. As mentioned, the early simulation with a beam energy of 500 GeV and the same ratio of passive and active media had a compensation ratio estimated to be 1.02. This suggests that for higher energies, the detector will be compensated.

Simulations					
Beam energy	20 GeV	40 GeV	60 GeV	80 GeV	120 GeV
Compensation	0.31	0.562	0.772	0.700	0.834
Resolution	0.230	0.212	0.168	0.168	0.104

Table 5: The Energy resolution and Compensation ratio of the simulated runs without tungsten

The energy resolution becomes better with the increase in energy as expected, and the resolution for the simulated 120 GeV run is decent. But for the lower energies, it is considerably worse than the setup 1 energy resolutions. This further advocates that the fits from setup 1 are not ideal. However, the 60 GeV run for the simulation is very close to the calculated energy resolution from setup 3 and 5.

6 Conclusion

In this thesis, the first first prototype of Focal-H was characterized. It was tested with multiple different experimental setups, including different obstacles in front of Focal-H, changes in Focal-H's position against the beam, up to six different beam energies, and the amount of effective channels were estimated. The energy resolution varied strongly from setup to setup, going from very poor resolution to acceptable. The best calculated energy resolution was based on setup from channel 35 with an energy resolution between 0.063 ± 0.004 and 0.200 ± 0.012 for the five tested beam energies for setup 1. However, in setup 1, it was discussed that for the higher energies, the choice to leave out the right tail in the fits resulted in an unrealistically good energy resolution. So it is hardly the ideal and true resolution. Compared with the simulations which had a relatively higher energy resolution, it argues for the same. However, the energy resolution estimated from the 60 GeV run for both setup 3 and 5, were similar to the simulations energy resolution. Furthermore, the estimated compensation ratio from the simulation did not fulfill the requirement of $\frac{e}{h} = 1$ which might have affected the resolutions. It did, however, come closer to 1 as the energy increased.

The binary readout from the CAEN board - or in this thesis better known as ADC values - were translated to GeV to properly distinguish the different beam energies in the detector. The different experimental setups all provided different translations, especially the setups including tungsten. But since the settings of the CAEN board could be changed individually between setups, this is not concerning. It could, however, also be due to insufficient statistics gathered in these setups. Thus, the latter possibility should be kept in mind for the next prototype. The experimental setup without tungsten did provide a translation in ADC to GeV following the expected pattern, which was shown in figure 34.

The early change of the channels positions was clearly the savior of this prototype. With-

out it, it would most likely only be the two HM data sets which had been usable for the data analysis. This was caused by the lack of event matching between the two boards, this was an issue for the majority of the data, effectively rendering B0 worthless beyond being a indication of where to cut. This might have been due to a lack of understanding of how the CAEN boards synchronized and matched events. A solution to this problem could be to only have one large CAEN board for all channels. This would avoid the boards' event matching and also the need to merge the two boards data sets in Collate.C. Because this would most likely not be possible for the final Focal-H, and this would then only be a temporary solution.

From the two HM data sets it was made clear that there were some flaws in how the merging between the data sets from each board was done. These were cross validation runs, and that would mean that there should be an exactly equal amount of events in each board. Yet, a large part of the data was lost doing this process. Thus, an obvious improvement would be a better merging system from Collate.C, to keep as much data as possible.

Another thing to consider for the next prototype is the sensitivity of the boards and how this effects the data. It was possible to change the gain and thus the sensitivity of the current CAEN boards, but this was unfortunately not utilized properly. It was observed that in a large part of the events, at least one channel hit the maximum ADC threshold. This would mean that information was lost in that event. To avoid this, the obvious answer would be to lower the gain, but too low a gain might cause problems for the low energies. One could also dynamically change and adapt the gain for the different energies, though the latter option would most likely render the ADC translation to GeV impossible. The 2d histogram of B0 showed that further noise reduction might be needed since there were cases where the entire board was active. This would also be solved by lowering the gain and this would then give a clearer look at the amount of active channels for the HM data sets. The setups which included tungsten as the obstacle had a significantly lower amount of events in the range of the second peak. Therefore, longer test runs should be applied to setups involving tungsten, especially if the gain is lowered further. The next prototype will consist of several modules which are close to the same size as the first prototype. Thus, it is likely that it will have more than one CAEN board, and it is, therefore, important to map the channels properly, optimize the gain setting, and to get a better understanding of how the boards' synchronize.

References

[1] Richard Wigmans, Michele Livan; Calorimetry for Collider Physics, an introduction.

[2] Marius Groll; Construction and commissioning of a hadronic test-beam calorimeter to validate the particle-flow concept at the ILC, may 2007.

[3] Christian W. Fabjan and Fabiola Gianotti; Calorimetry for Particle Physics, 31 October 2003.

[4] International Conference on Computing in High Energy and Nuclear Physics 2012, The Geant4 Virtual Monte Carlo.

[5] ALICE, CERN, *Alice information*, [Online], Available:
https://alice.cern/ (visited on 26/03/2021)

[6] LHC, CERN, *The Large Hadron Collider*, [Online], Available:
https://home.cern/science/accelerators/large-hadron-collider (visited on 14/02/2022)

[7] ALICE, Detector, *Alice figures*, [Online], Available:
https://cds.cern.ch/record/2282027/plots (visited on 06/04/2021)

[8] C2, Hardware [Online]. Available:
https://www.computerome.dk/display/C2W/System+Hardware
(visited on 12/09/2021)

[9] SM, The Standard Model [Online]. Available:
https://www.quantumdiaries.org/wp-content/uploads/2014/03/2000px-Standard_Model_of_Elementary_Particles.svg_.jpg (visited on 28/10/2021)

[10] QCD, Quantum chromodynamics [Online]. Available:
https://en.wikipedia.org/wiki/Quantum_chromodynamics (visited on 08/11/2021)

[11] Tracking particles [Online]. Available:
https://en.wikipedia.org/wiki/Tracking_(particle_physics) (visited on 14/02/2022)

[12] SD, Standard Model [Online]. Available:
https://www.physik.uzh.ch/en/researcharea/lhcb/outreach/StandardModel.html
(visited on 08/11/2021)

[13] ALICE Collaboration.
A Forward Calorimeter (FoCal) in the ALICE experiment. 2020

[14] Atomic and Nuclear Properties of Materials, [Online]. Available:
https://pdg.lbl.gov/2020/AtomicNuclearProperties/index.html.
(visited on 29/01/2022)

[15] Focalh_testbeam Merge Data, [Online]. Available:
https://gitlab.com/cholmcc/focalh_testbeam. (visited on 17/11/2021)

[16] North Area infomation, [Online]. Available:
https://indico.cern.ch/event/813822/contributions/3648315/attachments/1975530/3291392/BTTB_8th_Overview_over_CERN_test_beam_facilities_and_LS2.pdf.
(visited on 17/11/2021)

[17] North Area infomation, [Online]. Available:
https://home.cern/news/news/experiments/sps-experiments-are-back-action. (visited on 23/11/2021)

[18] Computerome information, [Online]. Available:
https://computerome.dtu.dk. (visited on 29/11/2021)

[19] Computerome wiki, [Online]. Available:
https://www.computerome.dk/display/C2W/Computerome+2.0.
(visited on 29/11/2021)

[20] Computerome Terms and Conditions, [Online]. Available:
https://www.computerome.dk/display/C2W/Terms+and+Conditions.
(visited on 30/11/2021)

[21] Computerome information Kunet, [Online]. Available:
https://kunet.ku.dk/work-areas/research/Research%20Infrastructure/research-it/computerome-2.0/Pages/default.aspx. (visited on 16/12/2021)

[22] Ian Bearden; Personal conversations

[23] Magnus Thøgersen; Personal conversations

[24] Sarah Andersson; FoCal-H simulations for ALICE,
A study on the design of the Forward Hadronic Calorimeter in the proposed FoCal upgrade for ALICE

[25] Scintillators, fibers [Online]. Available:
https://www.crystals.saint-gobain.com/radiation-detection-scintillators/fibers
(visited on 15/02/2022)

[26] SensL; Introduction to SiPM, Technical notes

[27] Onsemi, Silicon Photomultipliers (SiPM), Low-Noise, Blue-Sensitive, C-Series SiPM Sensors
https://www.onsemi.com/pdf/datasheet/microc-series-d.pdf
(visited on 15/02/2022)

[28] CAEN Electronic Instrumentation,
User Manual UM5833, A1702/DT5702, 32-channel Silicon Photomultipliers Readout Front-End Board, 2019

[29] Geant3, [Online]. Available:
https://github.com/vmc-project/geant3. (visited on 01/12/2021)

[30] Root, [Online]. Available:
https://root.cern/. (visited on 01/12/2021)

[31] Geant4, [Online]. Available:
https://github.com/Geant4/geant4. (visited on 01/12/2021)

[32] Geant4 VMC, [Online]. Available:
https://github.com/vmc-project/geant4_vmc. (visited on 01/12/2021)

[33] VGM, [Online]. Available:
https://github.com/vmc-project/vgm. (visited on 01/12/2021)

[34] Root-framework, [Online]. Available:
https://gitlab.cern.ch/cholm/root-framework. (visited on 01/12/2021)

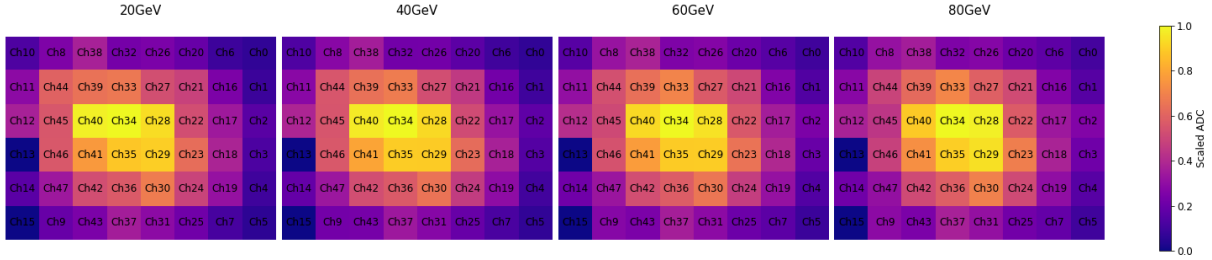
[35] Root-simulation, [Online]. Available:
https://gitlab.cern.ch/cholm/root-simulation. (visited on 01/12/2021)

[36] Error propagation, [Online]. Available:
https://www.nbi.dk/ petersen/Teaching/Stat2018/Week1/AS2018_11_20_ErrorPropagation.pdf
(visited on 07/02/2022)

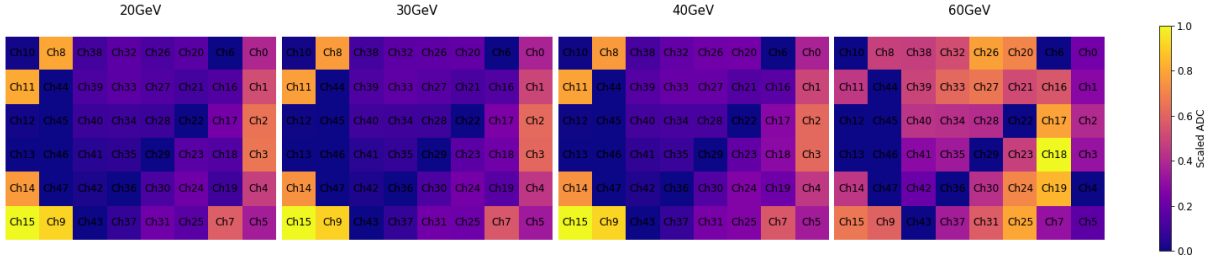
A Use full commands for C2

* screen –S "ScreenName" — Starts new screen session with the given name.
* ctrl+a+d — Disconnects from your screen session and returns to the terminal
* screen –r "ScreenName" — Connects you to the given screen
* screen –ls — Shows all your screen sessions and their status (attached/detached)
* showq — Shows ALL running jobs on C2 (From all users)
* showq -u "username" (or showq — grep) — Shows jobs for specific user
* qstat "jobid" — Checks stats on job
* qstat -u "username" — Shows information about jobs from that user
* qstat -f — Checks job historic
* qstat -t — Checks all task in parallel jobs
* qmgr -c 'p s' — Cluster information
* qstat -q — How many jobs currently running in the cluster
* qdel "jobid" — Deletes job

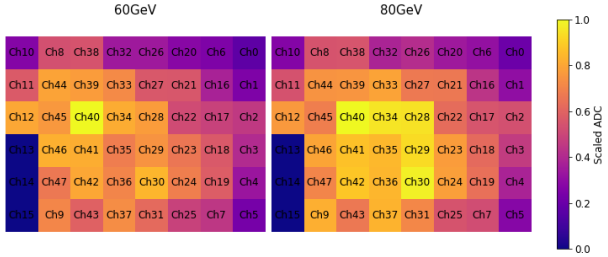
B Intensity maps of detector



(a) Setup 2.



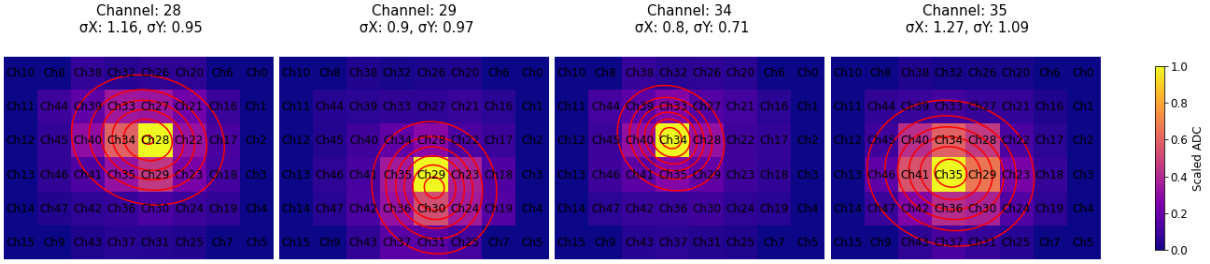
(b) Setup 4.



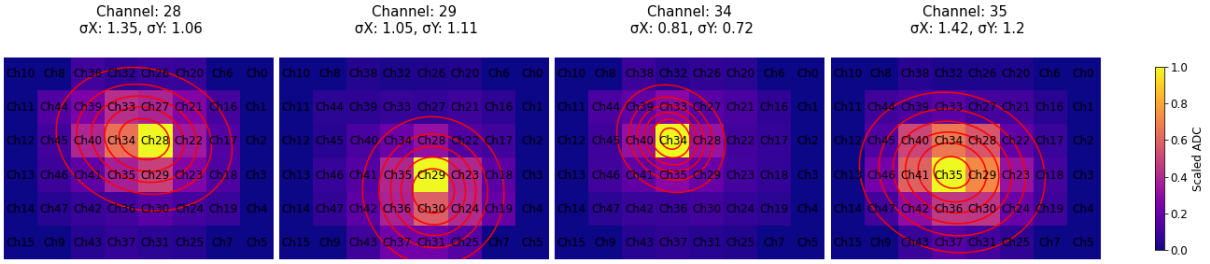
(c) High matching data set from Setup 5.

Figure 36: The intensity map's of the the different setups and available energies. The ADC value are scaled such that the maximum ADC value equals to one

C 2d Gaussian fits



(a) Beam energy: 30 GeV.



(b) Beam energy: 40 GeV.

Figure 37: Setup 1's intensity map's of the presumably effective channels (ch28, ch29, ch34, ch35) with a 2d Gaussian fit.

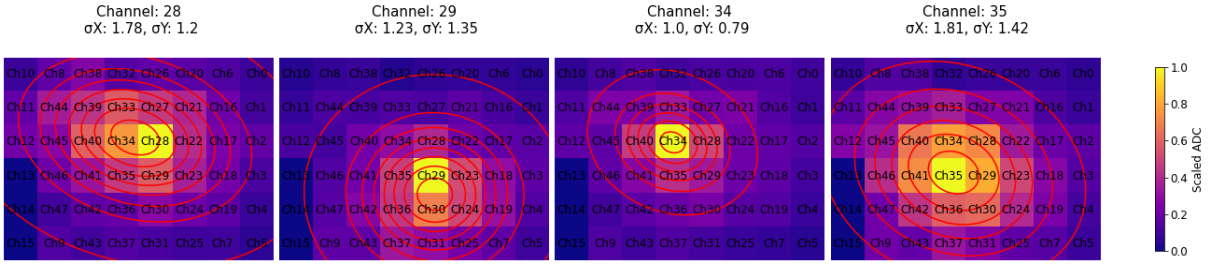
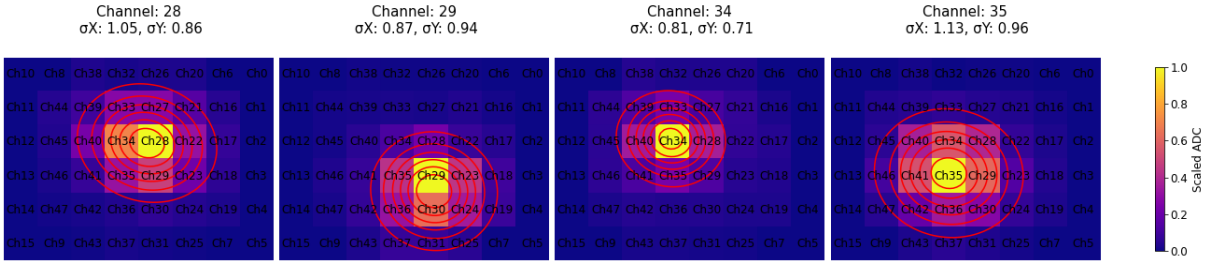
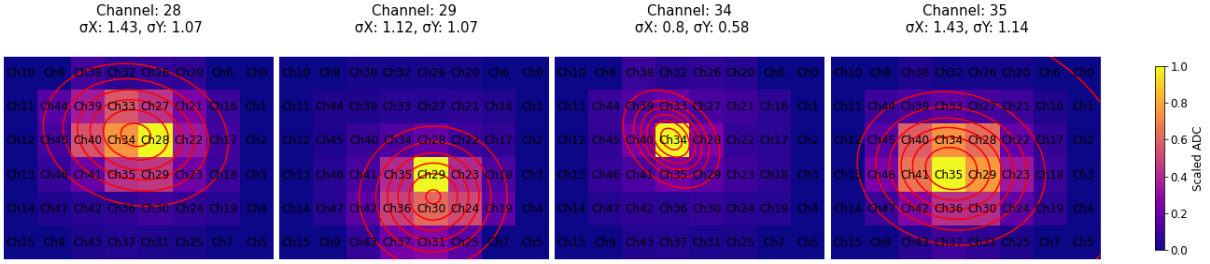


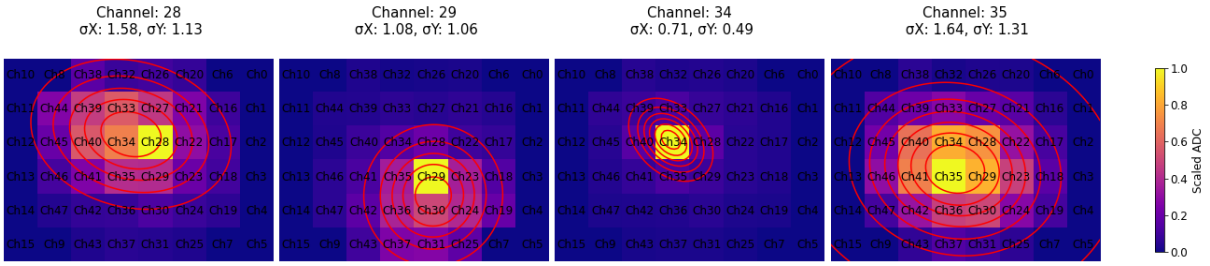
Figure 38: Setup 5's intensity map's of the presumably effective channels (ch28, ch29, ch34, ch35) with a 2d Gaussian fit.



(a) Beam energy: 20 GeV.

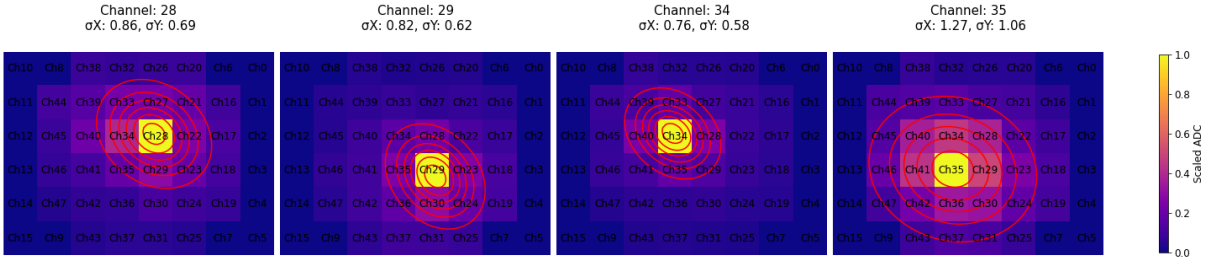


(b) Beam energy: 40 GeV.

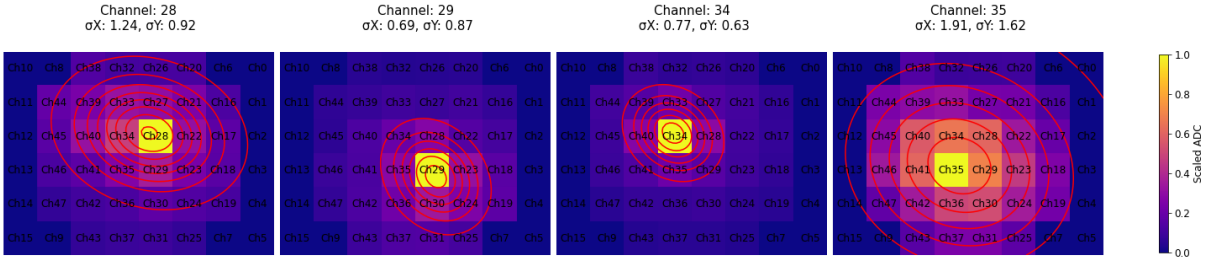


(c) Beam energy: 60 GeV.

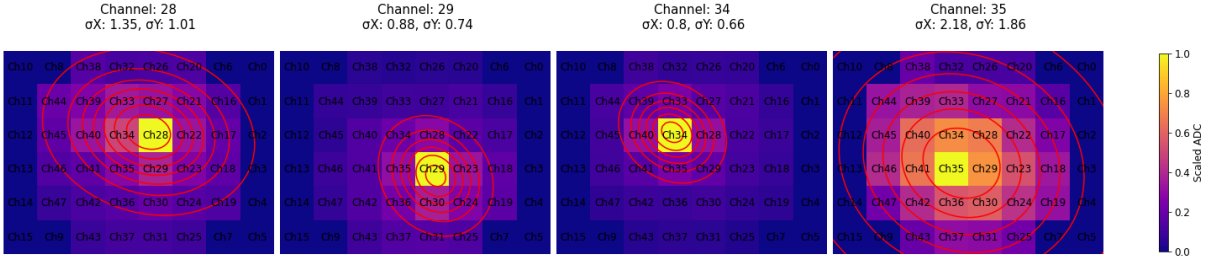
Figure 39: Setup 2's intensity map's of the presumably effective channels (ch28, ch29, ch34, ch35) with a 2d Gaussian fit.



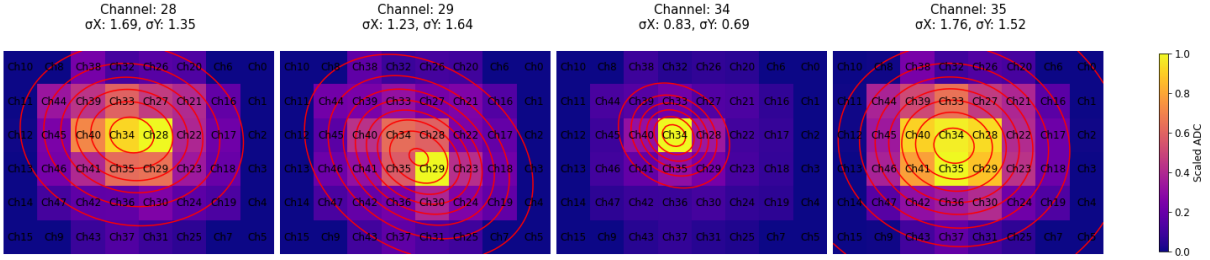
(a) Beam energy: 20 GeV.



(b) Beam energy: 40 GeV.

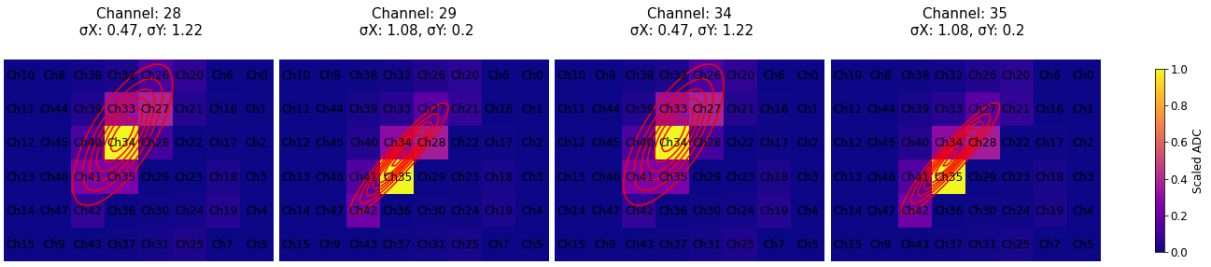


(c) Beam energy: 60 GeV.

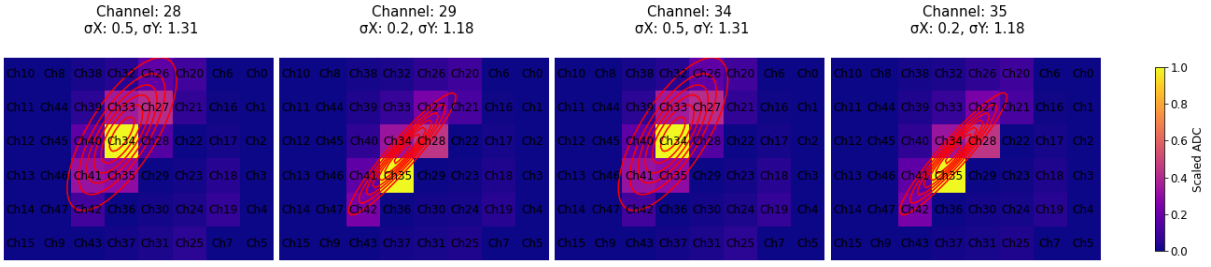


(d) Beam energy: 120 GeV.

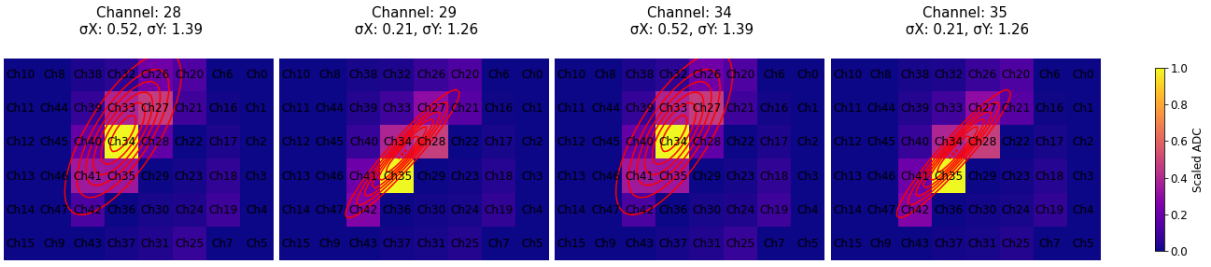
Figure 40: Setup 3's intensity map's of the presumably effective channels (ch28, ch29, ch34, ch35) with a 2d Gaussian fit.



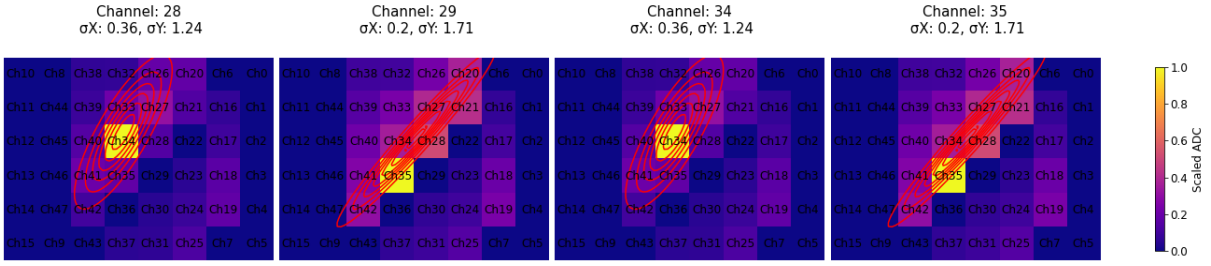
(a) Beam energy: 20 GeV.



(b) Beam energy: 30 GeV.



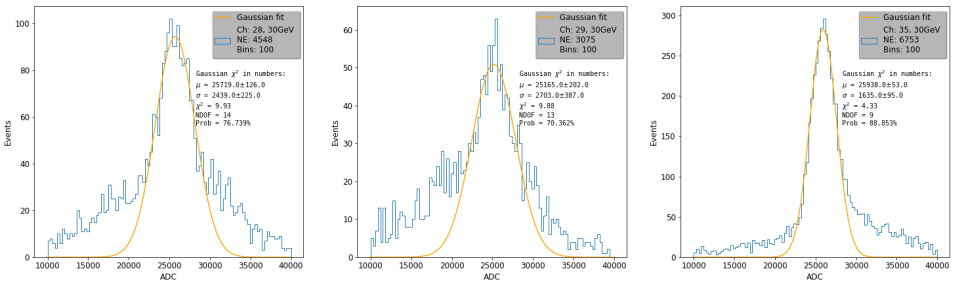
(c) Beam energy: 40 GeV.



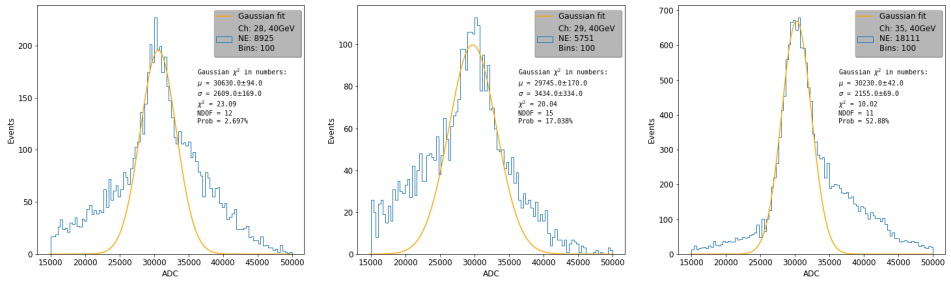
(d) Beam energy: 6 GeV.

Figure 41: Setup 4's intensity map's of the presumably effective channels (ch28, ch29, ch34, ch35) with a 2d Gaussian fit.

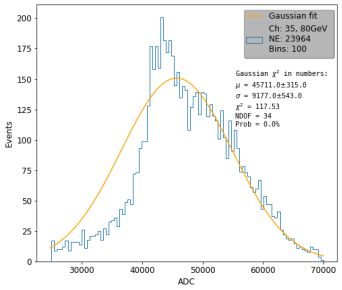
D Gaussian fits for sum of ADC



(a) Beam energy: 30GeV.

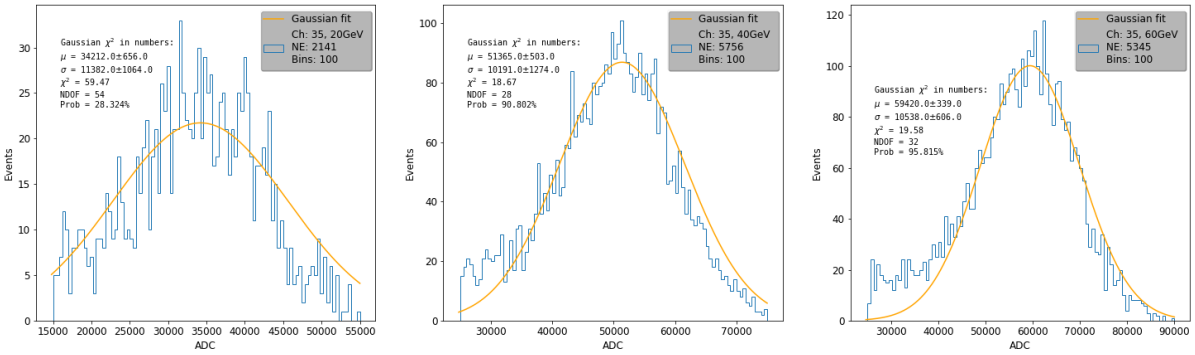


(b) Beam energy: 40GeV.

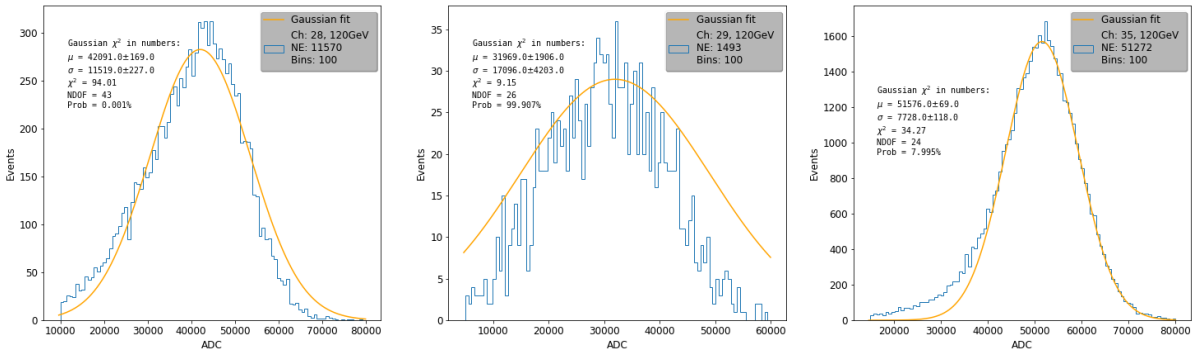


(c) Beam energy: 80GeV.

Figure 42: Histogram of the sum of ADC values for the effective channels with a Gaussian fit for beam energies 30, 40 and 80 GeV with setup 1

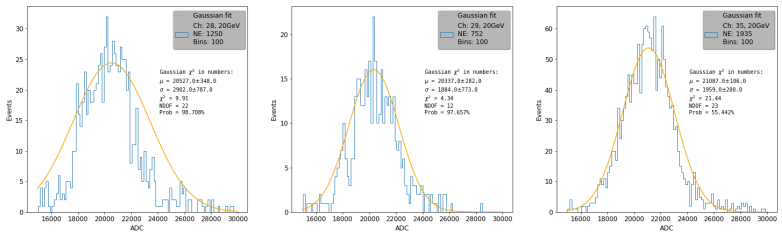


(a) Channel 35 for Beam energy: 20,40 and 60 GeV.

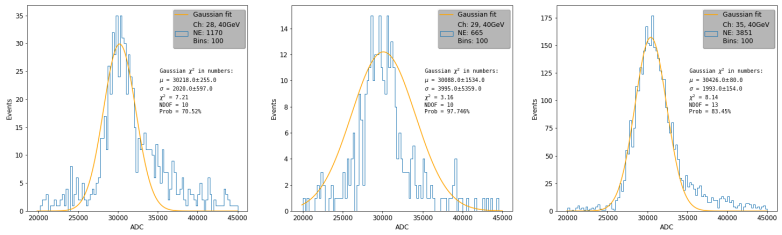


(b) Beam energy: 120 GeV. From left to right is Channel 28, 29 and 35.

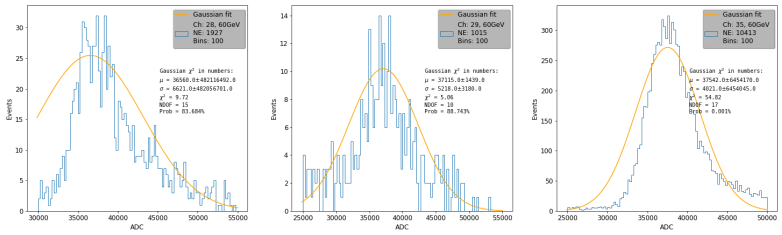
Figure 43: Histogram of the sum of ADC values for the effective channels with a Gaussian fit for Setup 3



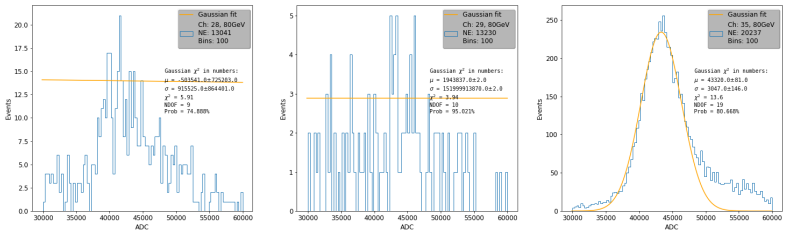
(a) Beam energy: 20GeV.



(b) Beam energy: 40GeV.

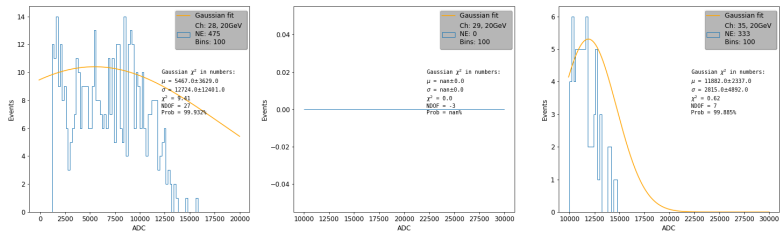


(c) Beam energy: 60GeV.

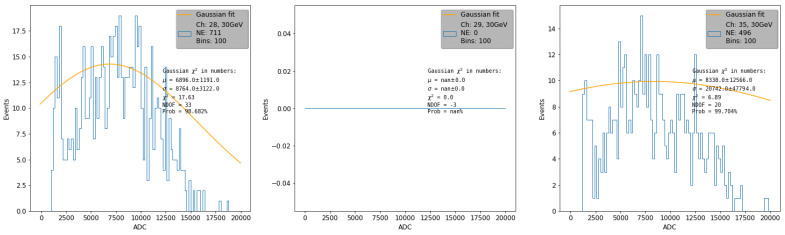


(d) Beam energy: 80GeV.

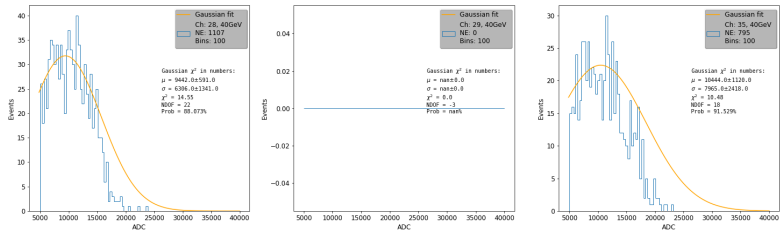
Figure 44: Histogram of the sum of ADC values for all the effective channels with a Gaussian fit for beam energies 20, 40, 60 and 80 GeV with setup 2



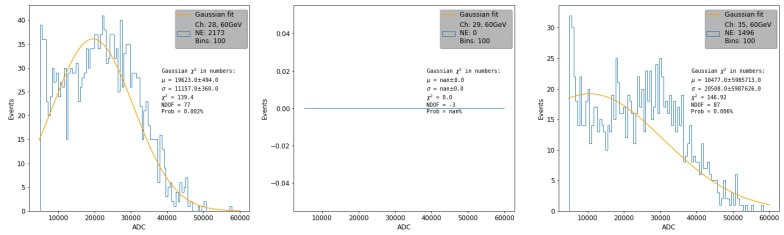
(a) Beam energy: 20GeV. From left to right is Channel 28, 29 and 35.



(b) Beam energy: 30GeV. From left to right is Channel 28, 29 and 35.



(c) Beam energy: 40GeV. From left to right is Channel 28, 29 and 35.



(d) Beam energy: 60GeV. From left to right is Channel 28, 29 and 35.

Figure 45: Histogram of the sum of ADC values for all the effective channels with a Gaussian fit for beam energies 20, 40, 60 and 80 GeV with setup 4

E Simulation plots

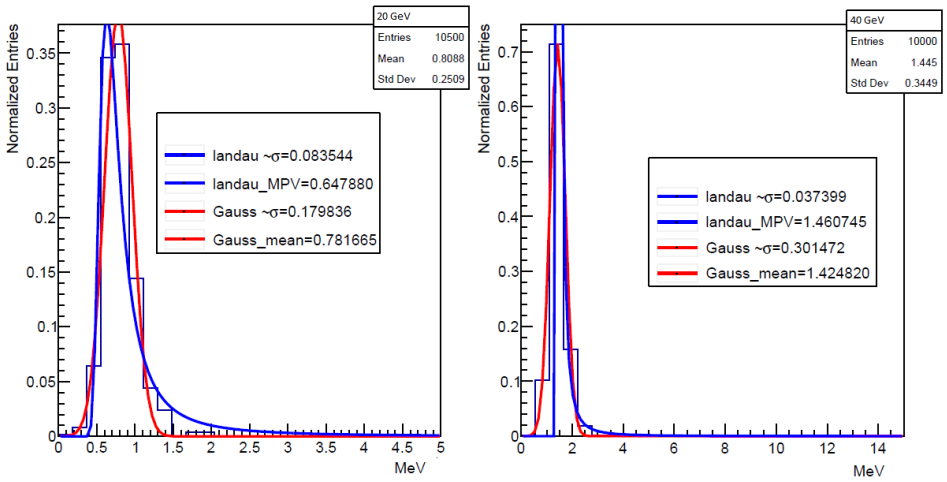


Figure 46: Energy distribution in detector with beam energy on 20 and 40 GeV.

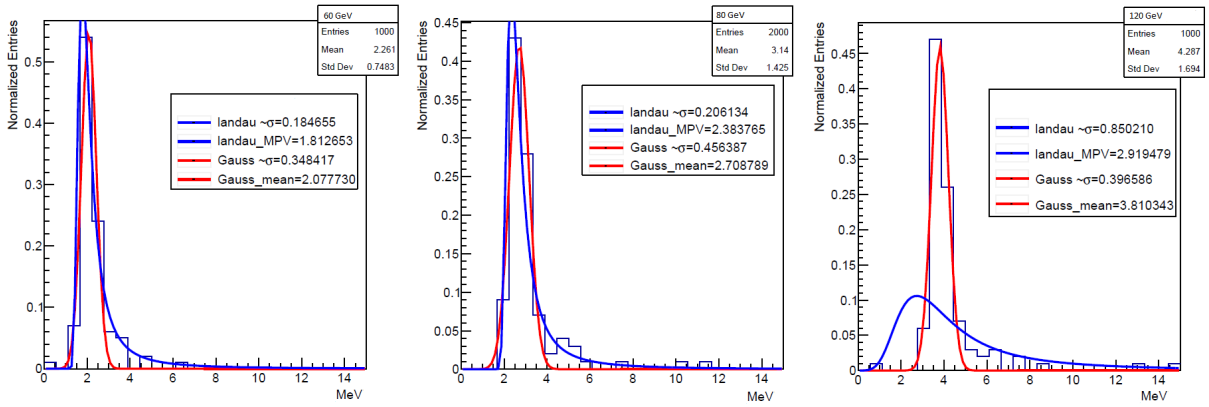
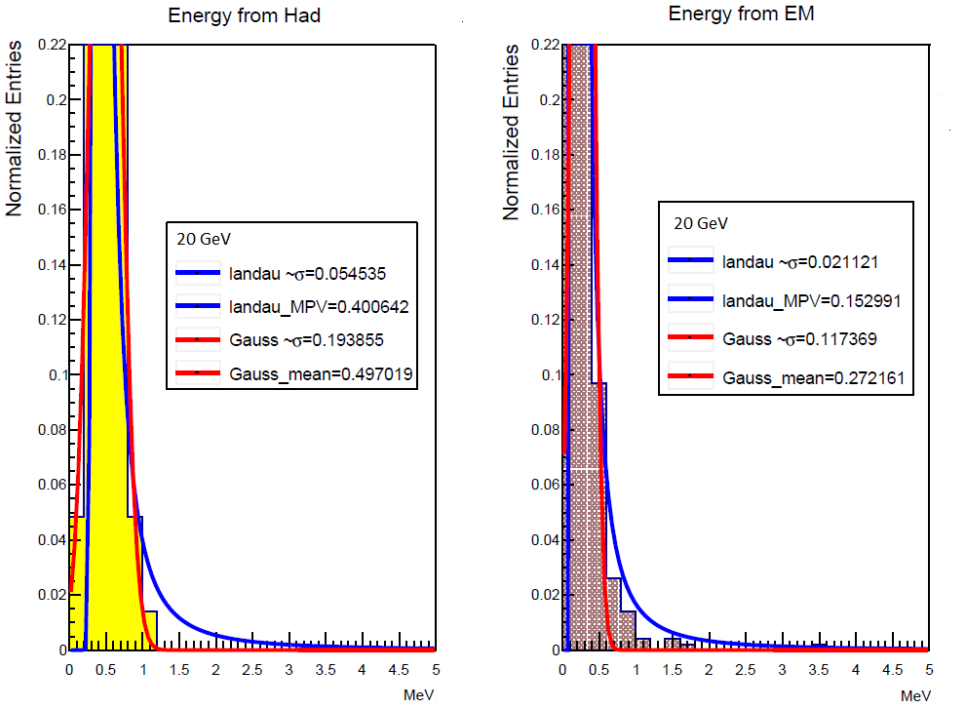
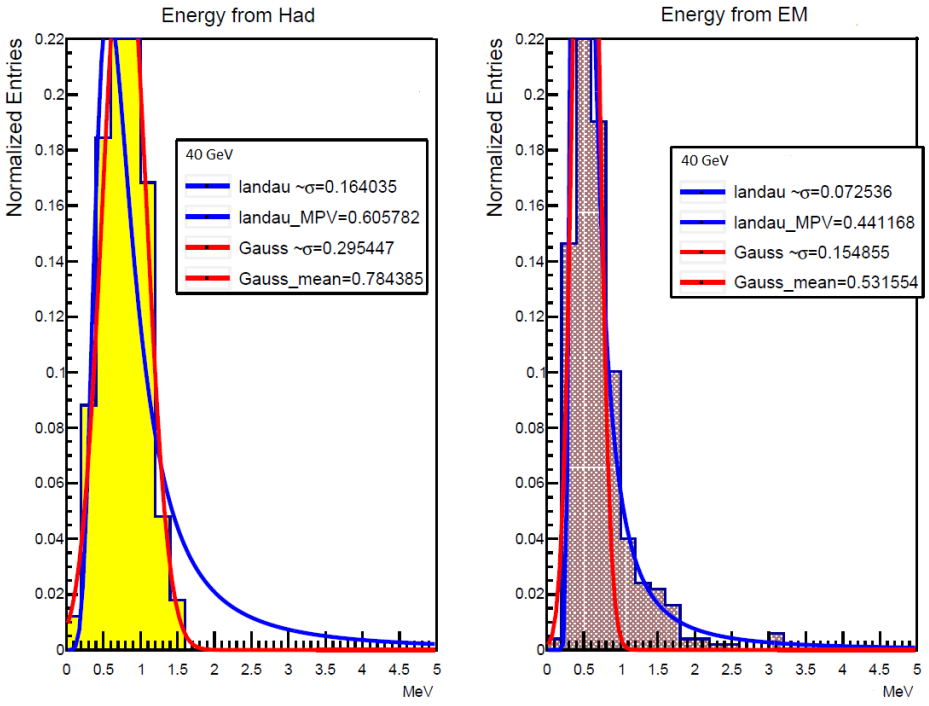


Figure 47: Energy distribution in detector with beam energy on 60, 80 and 120 GeV.

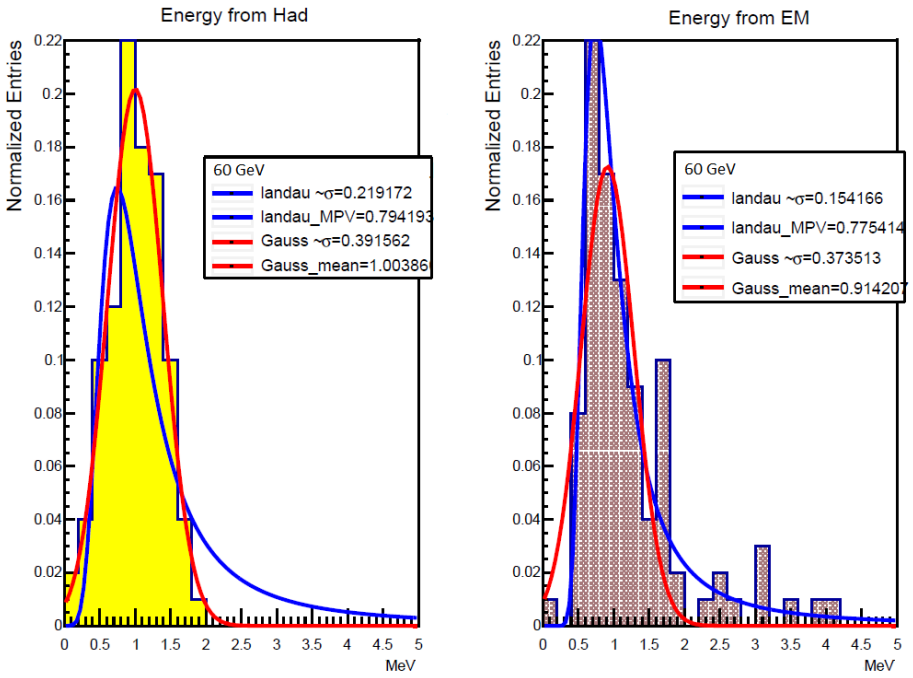


(a) Beam energy: 20 GeV.

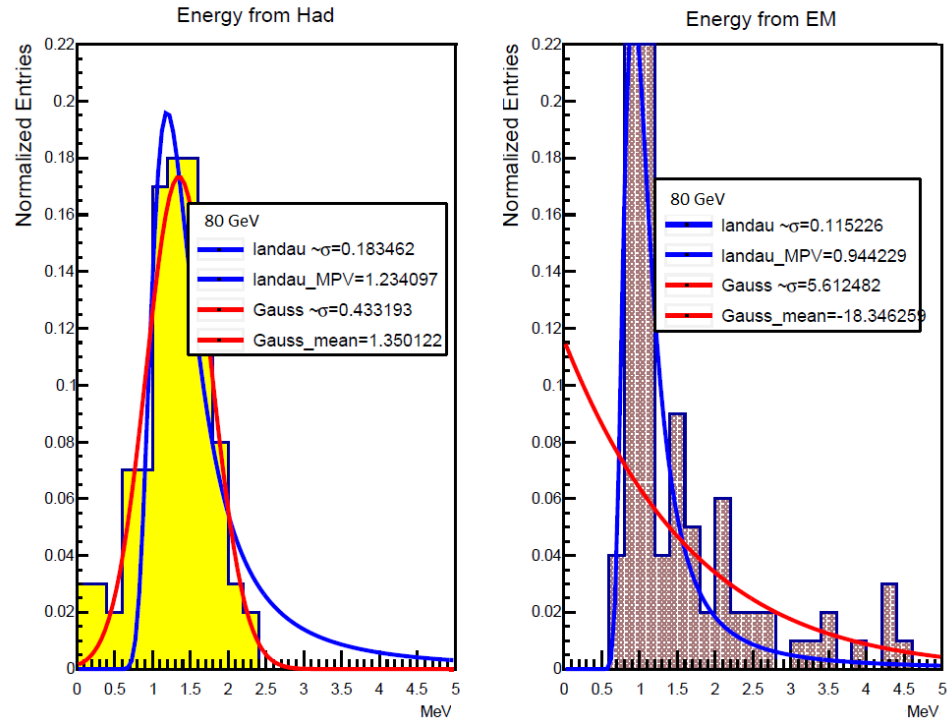


(b) Beam energy: 40 GeV.

Figure 48: The energy distribution of the hadronic part (left) and the electromagnetic part (right) of the hadron shower with beam energy on 20 and 40 GeV.



(a) Beam energy: 20 GeV.



(b) Beam energy: 40 GeV.

Figure 49: The energy distribution of the hadronic part (left) and the electromagnetic part (right) of the hadron shower with beam energy on 60 and 80 GeV.

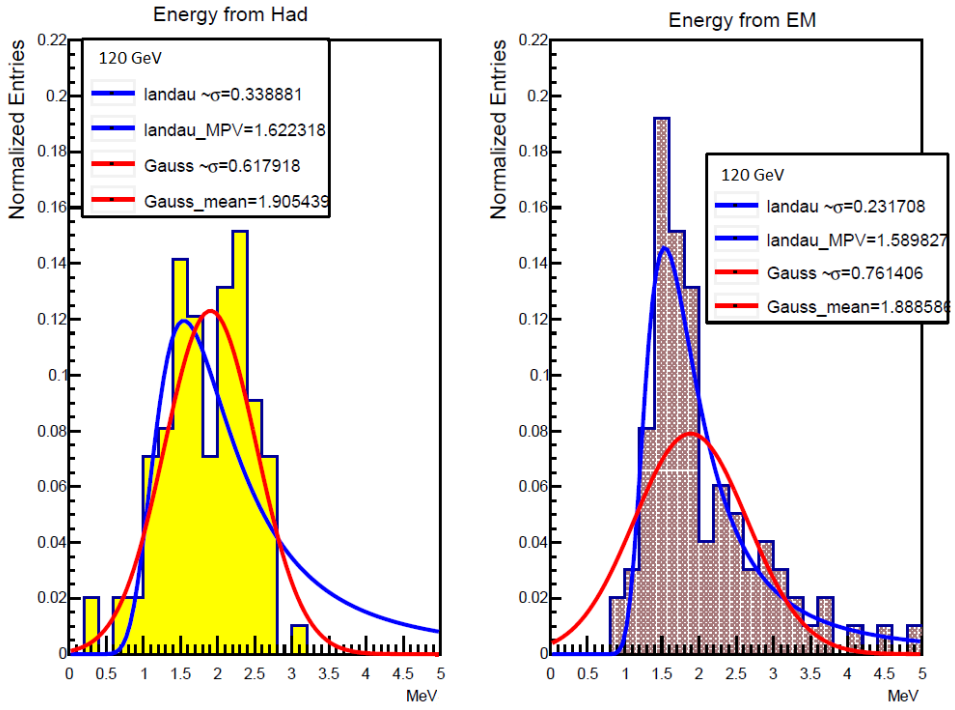


Figure 50: The energy distribution of the hadronic part (left) and the electromagnetic part (right) of the hadron shower with beam energy on 120 GeV.

F Python Functions

```
  # -*- coding: utf-8 -*-
"""
Created on Tue Dec  7 15:43:17 2021

@author: Sebastian
"""

import numpy as np
import matplotlib.pyplot as plt

#%%

def getcoords(filename):
    alle = np.load(filename) # load data

    Coords = np.zeros([48,2]) # create and define coords arrays for the sipm
    Coords[0:16][:,0],Coords[16:48][:,0] = alle[0:16][:,1],alle[32:64][:,1]
    Coords[0:16][:,1],Coords[16:48][:,1] = alle[0:16][:,2],alle[32:64][:,2]
    Coords[:,0] = (Coords[:,0] - 7)*-1
    return Coords

def FullLoad(filename,Timefile,bool=False):
    print("new")
    ped0 = np.loadtxt("myped.txt")
    ped = np.delete(ped0,np.s_[16:32],axis=0)

    alle = np.load(filename) # load data
    NE = len(alle)//64 # number of events
    print("Number of events no reduction: " + str(NE))
    AllChannelArray = np.zeros([64,NE]) # make an array for the output of each channel
    for i in range(NE):
        AllChannelArray[:,i]=alle[64*i:64*(i+1),3]
    DumpChannels = AllChannelArray[16:32] # dump the unused channels
    ChannelArray = np.delete(AllChannelArray,np.s_[16:32],axis=0)
#sync
    Time = np.load(Timefile)
    TB0Search = np.where(Time[:,0]==0)
    TB1Search = np.where(Time[:,1]==0)
    Syncindex = np.concatenate((TB0Search[0], TB1Search[0]))
    Syncindex = np.sort(Syncindex)
    SyncNum = len(Syncindex)
    print("Number of Sync events: " + str(SyncNum) + " in %: " + str((SyncNum/NE)*100)
```

```
    SyncEvents = np.zeros([48,SyncNum])
    for index in range(SyncNum):
        SyncEvents[:,index] = ChannelArray[:,Syncindex[index]]
    NoSync = np.delete(ChannelArray,Syncindex,axis=1)

    NoSyncNE = len(NoSync[0])
    Empty = np.zeros(NoSyncNE)
    for i in range(NoSyncNE):
        Empty[i]= np.sum(NoSync[:,i])
    emptyindex = np.where(Empty==0.0)
    emptynum = len(emptyindex[0])
    NoSync = np.delete(NoSync,emptyindex,axis=1)
    NoSyncNE = len(NoSync[0])
    print("Number of empty events: " + str(emptynum))
    print("Number of events (No empty): " + str(NoSyncNE))

#Limit
    LimitFull = np.empty(0,int)
    for i in range(48):
        Limitsearch = np.where(np.round(NoSync[i],3)==4089-ped[i])
        Limitsearch = np.array(Limitsearch[0])
        LimitFull = np.append(LimitFull,Limitsearch)
    Limitindex = np.unique(LimitFull)
    NoLimit = np.delete(NoSync,Limitindex,axis=1)
    LimitNum = len(Limitindex)
    print("Number of limit events: "+str(LimitNum) + " (" +
                                str((LimitNum/NE)*100) + ("%)"))
    LimitEvents = np.zeros([48,LimitNum])
    for index in range(LimitNum):
        LimitEvents[:,index] = NoSync[:,Limitindex[index]]
    Board0Array = NoLimit[0:16]
    Board1Array = NoLimit[16::]
    Sum0,Sum1 = np.zeros(NE-SyncNum-LimitNum-emptynum),np.zeros(NE-SyncNum-
                            LimitNum-emptynum)
    NonzeroB0,NonzeroB1 = np.zeros(NE-SyncNum-LimitNum-emptynum),np.zeros(NE-SyncNum
                            -LimitNum-emptynum)

    for i in range(NE-SyncNum-LimitNum-emptynum):
        Sum0[i]= np.sum(Board0Array[:,i])
        Sum1[i]= np.sum(Board1Array[:,i])
        NonzeroB0[i] = np.count_nonzero(Board0Array[:,i])
        NonzeroB1[i] = np.count_nonzero(Board1Array[:,i])

    Board0 = np.delete(Sum0,np.where(Sum0[:]==0),axis=0)
```

```
    Board1 = np.delete(Sum1,np.where(Sum1[:]==0),axis=0)
    NonzeroB0 = np.delete(NonzeroB0,np.where(NonzeroB0[:]==0),axis=0)
    NonzeroB1 = np.delete(NonzeroB1,np.where(NonzeroB1[:]==0),axis=0)
    print("Final number of events: " + str(len(NoLimit[0])))
    if bool:
        return LimitEvents #, DumpChannels
    else:
        return NoLimit, Board0, Board1,NonzeroB0,NonzeroB1



def hist(ChannelArray,NE) :
    EnergyHist = np.zeros(NE)
    Nonzero = np.zeros(NE)
    for i in range(NE):
        EnergyHist[i]= np.sum(ChannelArray[:,i])
        Nonzero[i] = np.count_nonzero(ChannelArray[:,i])
    return EnergyHist,Nonzero

def my2dhist(x0,y0,x1,y1,bins=10,title="title",title1="title",title2="title",
                label0="label",label1="Label", label2="LAB",bool=False):
    #fig values
    fig, (ax0,ax1) = plt.subplots(1,2,figsize=(20,8))
    fig.suptitle(title)

    #plot 1 values
    H0 = ax0.hist2d(x0,y0, bins=bins, label=label0,cmap="plasma")
    ax0.set_title(title1 + ", bins=" + str(bins))
    ax0.set(xlabel='ADC')
    ax0.set(ylabel='SIPM Fired')
    cbar = fig.colorbar(H0[3], ax=ax0)
    cbar.set_label(label2)
    if bool:
        ax0.legend()

    #plot 2 values
    H1 = ax1.hist2d(x1,y1, bins=bins, label=label1,cmap="plasma" )
    ax1.set_title(title2 + ", bins=" + str(bins))
    ax1.set(xlabel='ADC')
    ax1.set(ylabel='SIPM Fired')
    cbar1 =fig.colorbar(H1[3], ax=ax1)
    cbar1.set_label(label2)
    if bool:
        ax1.legend()
```

```
    plt.show()
    return fig

def MaxChX(data,Numberofchannels):
    maxindex = []
    maxcount= np.zeros(Numberofchannels)
    for i in range(len(data[0])):
        maxindex.append(np.where( data[:,i]== np.amax(data[:,i]))[0][0])
    maxindex = np.array(maxindex)
    for i in range(Numberofchannels):
        maxcount[i]=np.count_nonzero(maxindex == i)
    return maxcount

def Chmax(data,ch):
    chmax=[]
    for i in range(len(data[0])):
        if ch == np.where( data[:,i]== np.amax(data[:,i]))[0][0]:
            chmax.append(data[:,i])
    chmax = np.array(chmax)
    newchmax = np.zeros([48,len(chmax)])
    for i in range(len(chmax[0])):
        newchmax[i,:] = chmax[:,i]
    return newchmax


def mapmatrix(data,coords):
    matrix = np.zeros([6,8])
    for i in range(len(coords)):
        matrix[int(coords[i,1]),int(coords[i,0])] = data[i]
    return matrix

def totchsum(data,single):
    totsum= np.zeros(48)
    tsingle= np.zeros(48)
    for i in range(48):
        totsum[i] = np.sum(data[i,:])
    tsingle = data[:,single]
    return totsum, tsingle

def colorplot(data,title="title",cbartitle="Counts",cmap="plasma"):
    charray = np.array([["Ch10","Ch8","Ch38","Ch32","Ch26","Ch20","Ch6","Ch0"],
                        ["Ch11","Ch44","Ch39","Ch33","Ch27","Ch21","Ch16","Ch1"],
                        ["Ch12","Ch45","Ch40","Ch34","Ch28","Ch22","Ch17","Ch2"],
                        ["Ch13","Ch46","Ch41","Ch35","Ch29","Ch23","Ch18","Ch3"],
```

```
                    ["Ch14","Ch47","Ch42","Ch36","Ch30","Ch24","Ch19","Ch4"],
                    ["Ch15","Ch9","Ch43","Ch37","Ch31","Ch25","Ch7","Ch5"]
                    ])
    fig, ax = plt.subplots(figsize=(8,8))
    ax.set_title(title,fontsize=20)
    cax = ax.matshow(data, cmap=cmap,interpolation='nearest')
    for (i, j), z in np.ndenumerate(charray):
        ax.text(j, i, z, ha='center', va='center')
    plt.axis('off')
    cbar = fig.colorbar(cax)
    cbar.set_label(cbartitle)
    return fig


def datafit(Histdata,cutpoint1=0,cutpoint2=0):
    newdataarray = []
    for i in range(len(Histdata)):
        if cutpoint1>Histdata[i]>cutpoint2:
            newdataarray.append(Histdata[i])
    newdataarray = np.array(newdataarray)
    return newdataarray

def Scale_factor(x,Nbins,xmin,xmax):
    fig, ax = plt.subplots(figsize=(9,7))
    hist, ignore, ignore2 = ax.hist(x, bins=Nbins, range=(xmin,xmax))
    histn, ignoren, ignore2n = ax.hist(x, bins=Nbins, range=(xmin,xmax), density=True)
    N = max(hist)/max(histn)
    plt.close(fig)
    return N
def gauss_extended(x, N, mu, sigma):
    """Non-normalized Gaussian"""
    return N * gauss_pdf(x, mu, sigma)


def gauss_pdf(x, mu, sigma):
    """Normalized Gaussian"""
    return 1. / np.sqrt(2. * np.pi) / sigma * np.exp(-(x - mu) ** 2. / 2. / sigma ** 2.
def GaussianFit(Nbins,xmin,xmax,data,SebBins=0):
    """Given inputs of data and corresponding histogram
    returns Chi2 regression Gaussian fit from Minuit"""
    fig, ax = plt.subplots()
    hist, bins, _ = ax.hist(data,bins=Nbins,range=(xmin,xmax))
    plt.close(fig)
```

```
# Defining variables
dis_bins = bins[1] - bins[0]
x_hist = np.arange(xmin+dis_bins/2,(xmax),dis_bins)
x_var =  np.linspace(xmin-dis_bins/2,(xmax),1000)

# Only inputs from bins >0
histcor = hist[hist>SebBins]
x_hist = x_hist[hist>SebBins]
error = np.array([1/i for i in histcor])
N_scale =  max(hist)/max(gauss_pdf(x_var, np.mean(data), np.std(data)))

# Fit
chi2_object = Chi2Regression(gauss_extended, x_hist, histcor, weights=error)
minuit = Minuit(chi2_object, pedantic=False, N=N_scale, mu=np.mean(data),
            sigma=np.std(data), print_level=0)
minuit.migrad();
#minuit.hesse();
NDOF = len(histcor) - 3
Chi2_fit = minuit.fval
Prob_fit = stats.chi2.sf(Chi2_fit, NDOF)

#Parameters of fit
N_fit, mean_fit, sigma_fit = minuit.args
N_fite, mean_fite, sigma_fite = minuit.errors.values()


# Significant decimals
decmu, decstd = decimals(mean_fite), decimals(sigma_fite)

# Text string to use as a label or info box later
text_string = (r"Gaussian $\chi^2$ in numbers:" + "\n"
                + r"$\mu$ = " + str(round(mean_fit,decmu)) +
                r"$\pm$" + str(round(mean_fite,decmu)) + "\n"
                + r"$\sigma$ = " + str(round(sigma_fit,decstd)) +
                r"$\pm$" + str(round(sigma_fite,decstd)) + "\n"
                + r"$\chi^2$ = " + str(round(Chi2_fit,2)) + "\n"
                + "NDOF = " + str(NDOF) + "\n"
                + "Prob = " + str(round(Prob_fit*100,3)) + "%")

return minuit, x_var, gauss_extended(x_var, *minuit.args), text_string
```