Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

MASTERTHESIS

Neural Network based Pulse Shape Analysis with the Belle II Electromagnetic Calorimeter

vorgelegt von

Stella Katharina Wermuth

Fakultät: Physik

Studiengang: Physik Master of Science

Matrikelnummer: 7236313

Vorgelegt am 13. Juni 2021

Erstgutachter: Dr. Torben Ferber

Zweitgutachter: Prof. Dr. Oliver Gerberding

Abstract

The Belle II experiment, located at the SuperKEKB e^+e^- collider in Japan, uses pulse shape analysis techniques to distinguish electromagnetically and hadronically interacting particles within the CsI(Tl) electromagnetic calorimeter. The pulse shapes from the particle-dependent scintillation response are nominally analysed with a multi-template offline fit to measure the fraction of scintillation emission produced by hadrons. This fitting method allows for the determination of the total deposited energy, the total scintillation emission by hadrons, and the time of energy deposit. This thesis reports on a new approach to extract the total deposited energy, and the hadronic component of the scintillation emission from the pulse shapes using machine learning techniques. For this, a neural network is trained on pulse shapes produced in crystals from calorimeter clusters from simulated photons and pions, and is employed as a multivariate regression tool. Inferred on photons, the neural network outperforms the current fitting method in terms of crystal energy resolution and hadron intensity resolution. For pions the neural network shows a similar resolution compared with the current fitting method. Furthermore the neural network approach improves the discrimination of electromagnetic and hadronic interactions and is robust towards fluctuations in photon pile-up from beam backgrounds. Overall the neural network approach is promising, however additional fine tuning of the composition of the training sample could further improve its performance and robustness.

Zusammenfassung

Das Belle II-Experiment, am SuperKEKB e^+e^--Beschleuniger in Japan, verwendet Methoden der Pulsform Analyse, um elektromagnetisch und hadronisch wechselwirkende Teilchen innerhalb des CsI(Tl) elektromagnetischen Kalorimeters zu unterscheiden. Die Pulsformen der teilchenabhängigen Szintillationsemission werden nominell mit einem Multi-Template-Fit offline analysiert, um den Anteil der von Hadronen erzeugten Szintillationsemission zu messen. Diese Fit-Methode erlaubt die Bestimmung der gesamten deponierten Energie, der gesamten Szintillationsemission durch Hadronen und des Zeitpunkts der Energiedeposition. Diese Arbeit befasst sich mit einen neuen Ansatz, um die gesamte deponierte Energie und die hadronische Komponente der Szintillationsemission aus den Pulsformen mit Hilfe von maschinellem Lernen zu rekonstruieren. Dazu wird ein neuronales Netz mit Pulsformen trainiert, die in Kristallen von Kalorimeterclustern von simulierten Photonen und Pionen generiert wurden, und als multivariates Regressionsnetz eingesetzt. Für Photonen übertrifft das neuronale Netz die konventionelle Fit-Methode in Bezug auf die Auflösung der deponierten Energie und der Hadronenintensität. Für Pionen zeigt das neuronale Netz eine ähnliche Auflösung verglichen mit der konventionellen Fit-Methode. Darüber hinaus verbessert das neuronale Netz die Klassifizierung von elektromagnetischen und hadronischen Wechselwirkungen und ist robust gegenüber Fluktuationen von Photonen-Pile-up, welche aus Strahlhintergründen resultieren. Insgesamt ist der Ansatz des neuronalen Netzes vielversprechend, jedoch könnte eine zusätzliche Verfeinerung der Zusammensetzung des Trainingsets die Auflösung und Robustheit weiter verbessern.

Contents

1 Introduction **6**

2 Physics Motivation **7**

2.1 Standard Model of Particle Physics 7

2.2 Anomalous Magnetic Moment of the Muon 9

2.3 $e^+e^- \to \pi\pi$ Cross Section Measurements 12

3 Belle II Experiment **15**

3.1 Detector 15

3.2 Electromagnetic Calorimeter 18

3.2.1 Particle Interactions in the Calorimeter 18

3.2.2 Scintillation Emission in CsI(Tl) 22

3.2.3 Reconstruction 22

4 Simulated and Experimental Data Sets **27**

4.1 Event Simulation 27

4.1.1 Single-Particle Simulation 28

4.1.2 Full-Event Simulation 29

4.2 Data 29

4.3 Selection 29

4.3.1 Single-Photon and Single-Pion Selection 29

4.3.2 $e^+e^- \to \mu^+\mu^-\gamma$ Selection 31

4.3.3 $D^{*+} \to \pi^+ D^0[\to \pi^- K^+]$ Selection 32

5 Artificial Neural Networks **33**

5.1 Machine Learning and Neural Networks 33

5.2 Neural Network Architecture of this Work 35

6 Neural Network Performance **39**

6.1 Performance with Single-Particle Simulation 39

6.1.1 Crystal Energy Prediction 39

6.1.2 Hadron Intensity Prediction 43

6.1.3 Energy and Hadron Intensity Resolution 45

6.1.4 Separation of Photons and Pions 49
6.2 Robustness of the Neural Network Approach 51
6.2.1 Time Shift 51
6.2.2 Beam Induced Background 54
6.3 From Crystal Energy Resolution to Cluster Energy Resolution 56
6.4 Band Structure in Single Proton Simulation 58
6.5 Performance with Full-Event Simulation and Data 60
6.5.1 Photons from $e^+e^- \to \mu^+\mu^-\gamma$ 60
6.5.2 Charged Pions from $D^{*+} \to \pi^+ D^0[\to \pi^- K^+]$ 63
6.5.3 Separation of Photons and Pions 66

7 Outlook **69**

8 Summary and Conclusion **71**

A Appendix **73**
A.1 Resolution for each Fittype 73
A.2 Simulation Settings and Selection to study Cluster Energy . . 76
A.3 Simulation Settings and Selection for Single Protons 77

1 Introduction

The Standard Model of particle physics describes the known elementary particles and their interactions. Its predictions have been verified by a large number of experiments but it is known to be incomplete. One of the few Standard Model predictions in tension with experimental data is that for the anomalous magnetic moment of the muon. Its measurement is one of the most precise in physics, yet it deviates by 4.2σ from the Standard Model prediction, providing a possible hint to physics beyond the Standard Model [1]. To further investigate this discrepancy the precision of the theory prediction should be improved. For this it is necessary to precisely measure the $e^+e^- \to \pi^+\pi^-$ cross section, which limits the precision of the Standard Model prediction. One of the experiments with the goal to precisely measure the $e^+e^- \to \pi^+\pi^-$ cross section is the Belle II experiment, located at the SuperKEKB e^+e^- collider in Japan. The measurement requires good particle identification techniques in order to discriminate pions and muons for background-suppression. One method for particle identification at Belle II is pulse shape analysis using the electromagnetic calorimeter. Thus far the pulse shape is analysed via a multi-template fit which extracts information about the underlying interaction and allows one to discriminate electromagnetic and hadronic interactions. To potentially further improve particle identification with the Belle II electromagnetic calorimeter a neural network approach for pulse shape analysis is studied in this work. This thesis describes a study of how this new approach performs compared with the current multi-template fit and is structured in the following sections: The physics motivation is discussed in chapter 2 with focus on the anomalous magnetic moment of the muon. This is followed by chapter 3, which describes the Belle II experiment emphasising the electromagnetic calorimeter and the multi-template fit used for pulse shape analysis. The simulation parameters and data selection criteria used to generate pulse shapes for this study are presented in chapter 4. Chapter 5 explains the architecture of the neural network used. The results are presented in chapter 6, and an outlook and summary of this work are given in chapter 7 and chapter 8 respectively.

2 Physics Motivation

This chapter gives a short introduction to the Standard Model of particle physics, based on [2]. An important area of research for physics beyond the Standard Model is the anomalous magnetic moment of the muon, which is presented in the second part of the chapter. The focus lies on the theory of the anomaly and what limitation the theorists face in order to improve the precision of the calculations, as described in [3–6]. This is followed by a detailed look at one of the ways to improve the precision, which is the $e^+e^- \to \pi\pi$ cross section measurement [3, 7].

2.1 Standard Model of Particle Physics

The Standard Model of particle physics describes the known fundamental particles of the universe and their interactions. All particles, which are summarised in Figure 1, can be classed into fermions, which have a half-integer spin, and bosons with integer spin.

Fermions are the fundamental particles which make up matter. Each fermion has an anti-particle, which is identical in mass but has opposite quantum numbers, such as electric charge. Fermions are further grouped into quarks and leptons. Each of them has three generations with two particles per generation. The quarks come in six flavour states: the up, down, charm, strange, top and bottom quark. Experimentally only bound states of quarks, called hadrons, are observed. Free quarks have never been observed due to their colour confinement by the strong force. The most common hadrons at collider experiments are mesons, made of two quarks, and baryons, made of three quarks. Pions are the lightest mesons, which consist of one up and one down quark. The charged leptons are electron, muon and tau, which all carry an electric charge of -1, as well as neutrinos corresponding to each charged lepton with an electric charge of 0.

Within the Standard Model vector bosons are the force carriers and responsible for the interactions. The forces described by the Standard Model are the electromagnetic force, mediated by photons, the weak force, medi-

ated by $W^{\pm}$ and Z^0 bosons, and the strong force, mediated by eight different gluons. All of these bosons have spin 1. The photon is massless and couples to electric charge. All fermions except neutrinos interact electromagnetically. The electromagnetic interaction is described by quantum electrodynamics (QED). $W^{\pm}$ and Z^0 bosons couple to hypercharge, which all fermions carry. The electromagnetic and weak interaction can be combined to the electroweak interaction. Gluons couple to colour-charge, carried by quarks, not by leptons. The strong interaction is described by quantum chromodynamics (QCD). Additionally there is the Higgs particle with spin 0, which gives mass to the particles, and was discovered in 2012 at the Large Hadron Collider [8, 9].

In general the Standard Model is able to describe a vast amount of experimental observations. However, it is known that the Standard Model is incomplete, since it provides no explanation for phenomena like gravitation or the matter-antimatter asymmetry, and is not able to describe Dark Matter or Dark Energy. Furthermore one observes a tension between the Standard Model prediction and experimental measurement for the quantity $g-2$, which is described in detail in chapter 2.2.

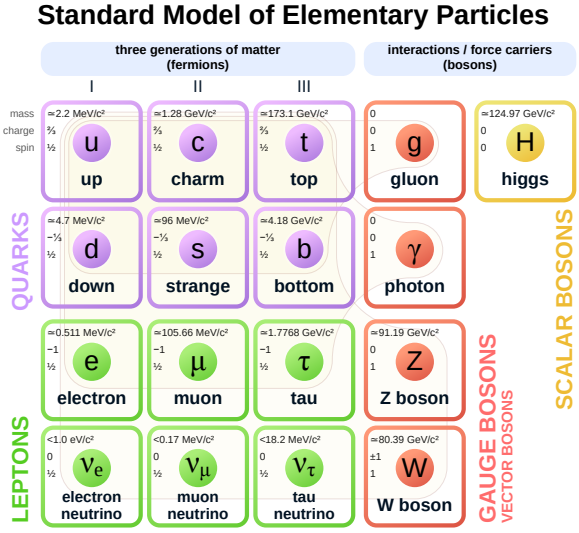


Figure 1: A Summary of the Standard Model of particle physics showing all particles and their mass, charge and spin, from [10].

2.2 Anomalous Magnetic Moment of the Muon

Since muons carry a spin $\vec{s}$ and electromagnetic charge e, they have a magnetic moment $\vec{M}_\mu$,

$$\vec{M}_\mu = g_\mu \frac{e}{2m_\mu} \vec{s} \tag{1}$$

where m_μ is the muon mass and g is the so-called *g-factor*. This magnetic moment makes muons precess like dipoles in an external magnetic field. In classical theory with point-like vertices, one would observe $g = 1$ however for relativistic quantum mechanics $g = 2$ is expected. Furthermore the Standard Model predicts additional higher order radiative corrections, which lead to a deviation from $g = 2$. This *anomalous magnetic moment* is expressed as

$$a_\mu = \frac{g-2}{2}. \tag{2}$$

Figure 2 shows different corrections to the interaction of a muon with a virtual photon of an external magnetic field. These corrections lead to the anomaly a_μ. Radiative QED corrections have the dominant contributions to a_μ. However, they are also easiest to calculate with high precision and have the smallest contribution to the uncertainty of the Standard Model prediction. Weak interactions have the smallest contribution to a_μ and are also well known. Contributions due to the strong interaction can be separated into hadronic vacuum polarisation (HVP) and hadronic light-by-light scattering (HLbL). HLbL is suppressed compared to HVP. The muon itself does not feel the strong force but it can emit a photon. In case of hadronic vacuum polarisation this photon creates a quark-antiquark pair, which then forms a hadron. The uncertainty on the HVP contribution dominates the uncertainty of the Standard Model calculation of a_μ.

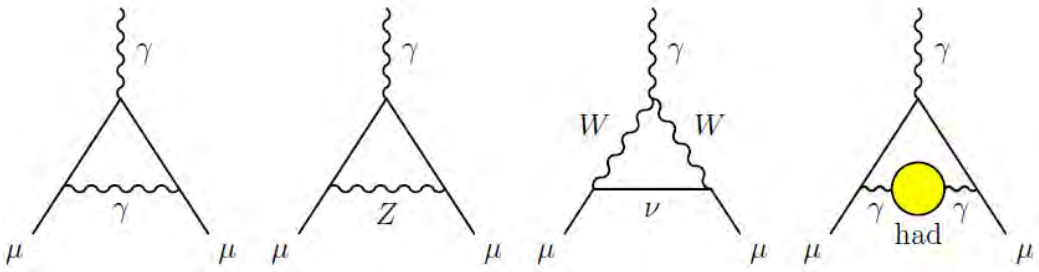


Figure 2: Processes which lead to the anomaly a_μ from left to right: Schwinger term predicted by QED, weak interaction with the Z and $W^\pm$ bosons and contribution from QCD hadron vacuum polarisation, Figure from [6].

Comparing the Standard Model prediction for a_μ with an experimental measurement provides a way to test the Standard Model theory very precisely. If a significant difference above 5σ between theory and experiment is found, it is evidence for new physics. The magnetic moment of the muon is measured by the high-precision g-2 experiment located at Fermilab, while the Muon g-2 theory initiative works on reducing the uncertainty of the theory prediction. At the time of writing a discrepancy of 4.2σ is observed between theory and experiment, as shown in Figure 3. If this discrepancy persists and the uncertainty of theory and experiment can be reduced, one might find a 5σ evidence for new physics. This makes this research very interesting and one of the most promising areas to reveal physics beyond the Standard Model.

The remainder of this chapter will focus on the limitations for the Standard Model theory calculations, which are described in [3]. As mentioned before, the HVP contribution a_μ^{HVP} limits the precision of the Standard Model calculation. One can see how different energy regions contribute to the anomaly in the left diagram in Figure 4. The largest contribution with more than 70% is the region from 0 GeV to 1 GeV with the ρ-meson decaying to two pions. This two-pion channel also contributes in large parts to the uncertainty of a_μ^{HVP}, as seen in the right diagram.

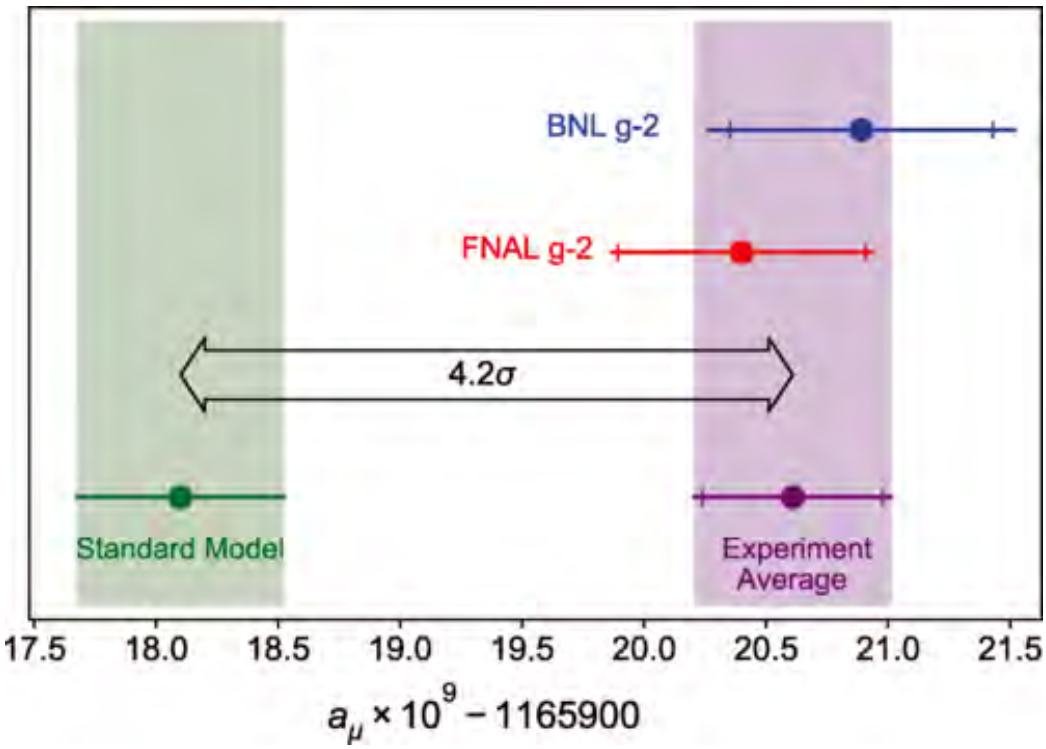


Figure 3: Comparison of the experimental average of the g-2 experiment, which was first located at the Brookhaven National Laboratory (BNL) and is now located at Fermi National Accelerator Laboratory (FNAL), and the latest theory calculation, published by the Muon g-2 theory initiative [3]. One finds a 4.2σ discrepancy. Figure was published in [1].

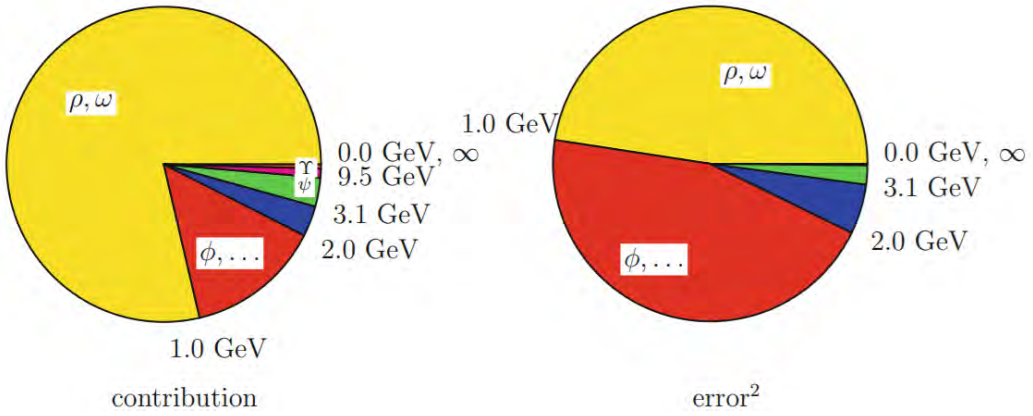


Figure 4: Shares in % of contribution (a) and the square of the error (b) for a_μ^{HVP} from different energy regions as well as important meson resonances within these energy regions, from [4].

In order to calculate a_μ^{HVP} non-perturbatively, one can use Eucledian lattice QCD, which needs large-scale computational resources [11]. Alternatively one can use a data-driven approach, which utilises measurements of cross sections. Since the two-pion channel is the channel with the largest contribution, the $\mu^+\mu^- \to \pi\pi$ cross section is especially important. Due to lepton universality one can extract this at e^+e^- colliders by measuring the $e^+e^- \to \pi\pi$ cross section. By improving the precision of the cross section measurement, one can in turn improve the precision of the a_μ^{HVP} calculation. Since this is the limiting factor it will lead to an overall improved precision of the a_μ prediction and will make it possible to see if the discrepancy between theory and experiment holds up with improved precision, which might open a window for new physics.

2.3 $e^+e^- \to \pi\pi$ Cross Section Measurements

To measure the $e^+e^- \to \pi\pi$ cross section one needs to scan the $\pi\pi$-invariant-mass spectrum. For this one can use the scan method, where the centre-of-mass energy of the collision is directly adjusted to different energies, as done at CMD-2 [12] and SND [13]. However, this approach requires a good beam energy resolution and it only allows for discrete data taking with gaps within the spectrum. Alternatively one can use initial state radiation (ISR) in order to indirectly adjust the centre-of-mass energy, known as radiative return. In this case the electron or positron emits a photon before the annihilation. The ISR leads to a decreased collision energy, making it possible to scan a whole range of energies with a fixed beam energy. For this method, one selects events with a $\pi\pi\gamma$ final state. This was done at KLOE [14, 15], at BABAR [16, 17], and more recently at BESIII [18] and CLEO-c [19].

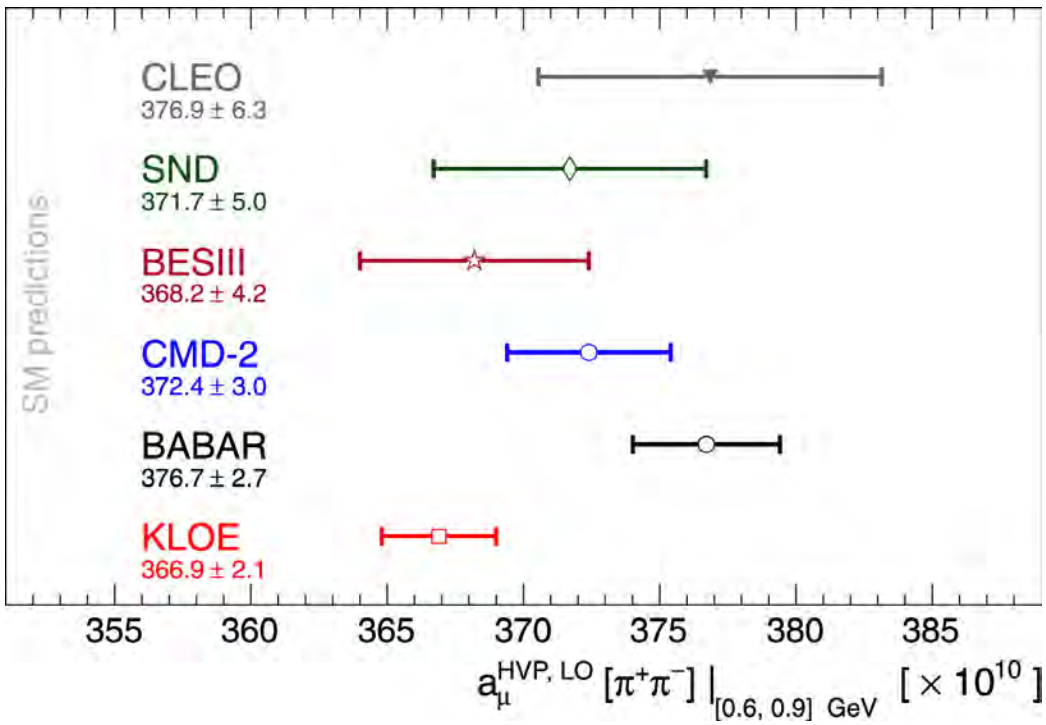


Figure 5: Comparison of the a_μ^{HVP} calculation for the two-pion channel for the range of 0.6 GeV to 0.9 GeV with the input of different experiments [3].

The calculated a_μ^{HVP} at leading order with input from these different experiments is shown in Figure 5. The most precise measurements where done by KLEO and BABAR, however, these measurements do not agree. This leads to an unresolved tension and the need for a new and precise measurement of the $e^+e^- \to \pi\pi$ cross section.

One of the most promising experiments to resolve the discrepancy between the experiments and provide improved precision for the a_μ^{HVP} calculation is the Belle II experiment, which will be introduced in detail in chapter 3. Belle II also uses radiative return. In Figure 6 data from the first physics run from early 2018 is shown. One can see the comparison of data and simulation for the invariant mass spectrum of two charged particles, where each was assigned the pion mass. In this plot, one challenge of the $e^+e^- \to \pi\pi\gamma$ cross section measurement is visible, which is the background from $e^+e^- \to \mu\mu\gamma$. In order to suppress this background and reduce the

systematic uncertainty for this measurement good particle identification is needed to separate muons and pions. One of the detectors which contributes to the identification of these particles is the electromagnetic calorimeter, which is presented in chapters 3.2.2 and 3.2.3.

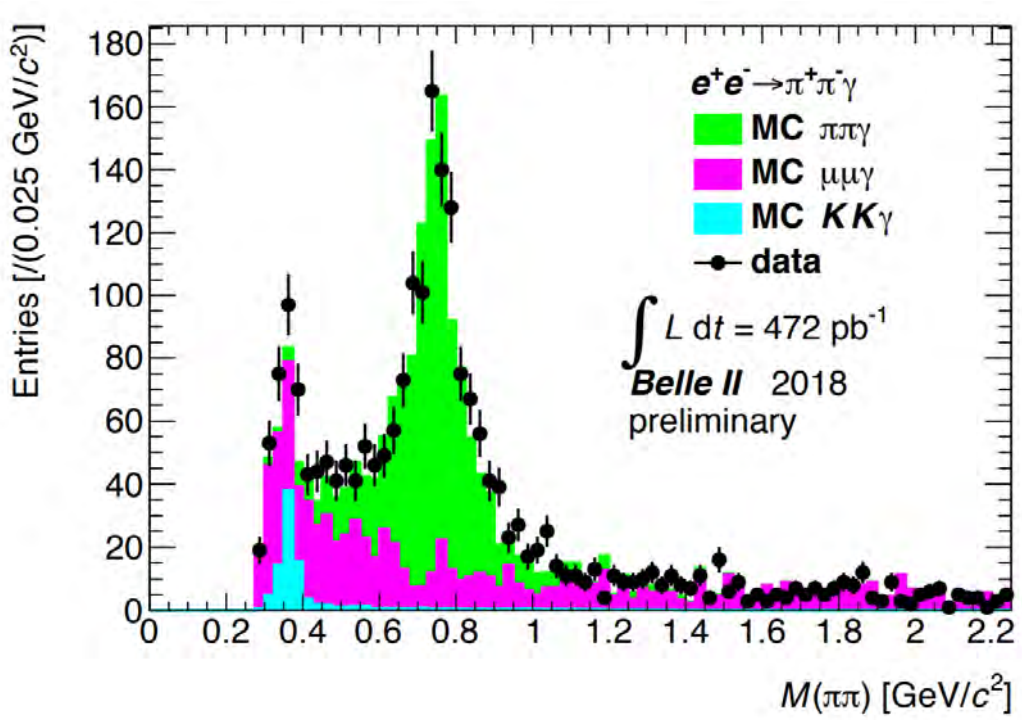


Figure 6: Mass spectrum of two pions observed at Belle II in 2018 compared with the Monte Carlo simulation of the signal and the main backgrounds from the $\mu\mu\gamma$ and $KK\gamma$ processes, [7].

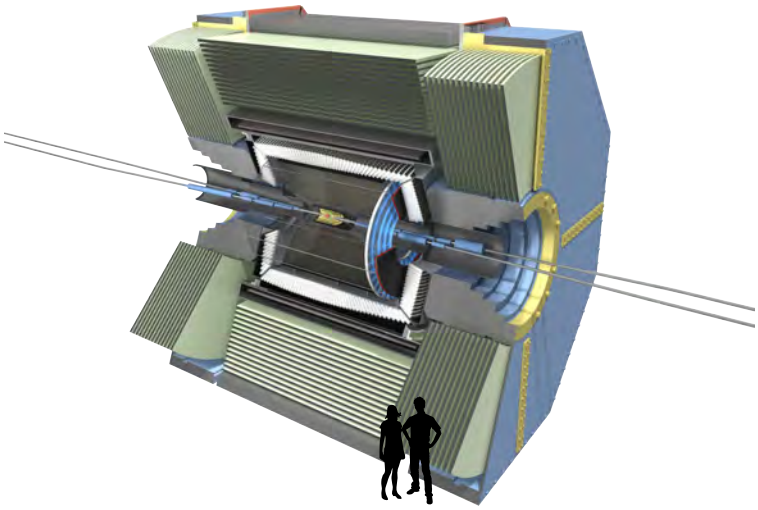


Figure 7: The Belle II detector [20].

3 Belle II Experiment

The Belle II experiment is located at SuperKEKB e^+e^- collider in Tsukuba, Japan. With a collision energy of 10.58 GeV at the $\Upsilon(4S)$ resonance Belle II is designed as a B-factory, since the $\Upsilon(4S)$ mainly decays into B-meson pairs. The main goals of Belle II are precision measurements, especially within the flavour sector, the search for rare processes and searches for physics beyond the Standard Model, like Dark Matter. This chapter gives a short overview of the Belle II detector in the first part and the second part focuses on the interactions, processes and signal reconstruction within the electromagnetic calorimeter.

3.1 Detector

The Belle II detector is described in detail in [21–23], which this chapter is based on. Its geometry is cylindrical around the interaction point, as seen in Figure 7. From the innermost system outward the sub-detectors are the tracking system, the electromagnetic calorimeter surrounded by a 1.5 T solenoid and the $K_L - \mu$ Detector. They are detailed in the following sub sections.

Vertex Detector

The vertex detector is located at a radius of 14 mm. It consists of two layers of pixel detectors followed by four layers of double-sided silicon-strip detectors, known as the silicon vertex detector. This detector system enables the reconstruction of the vertex by measuring the transverse coordinates of hits from the tracks of charged particles.

Central Drift Chamber

Spanning radii of 160 mm to 1100 mm the central drift chamber is the main tracking device. The large volume drift chamber is filled with a gas mixture of 50% He and 50% C_2H_6 and small drift cells. Charged particles will ionise the gas resulting in free electrons. These electrons are accelerated towards wires at high voltage and create an electron avalanche by further ionising the gas. From combining these wire signals, the trajectories of charged particles can be reconstructed. By measuring the curvature of the trajectory within the magnetic field of the surrounding solenoid the momentum of the particles can be calculated. Furthermore, the energy loss over distance travelled for particles with low momentum is used for particle identification.

Particle Identification System

In order to distinguish charged particles, especially kaons and pions, Belle II uses a particle identification system of two different Cherenkov detectors. Charged particles radiate Cherenkov photons within detectors and from the angle of these photons and the refractive index of the material one can reconstruct the velocity of the initial particle. From the velocity and the momentum of the particle the mass can be calculated and the particle can be identified. In the barrel a Cherenkov-based time-of-propagation detector with a fast time resolution provides two-dimensional information about the Cherenkov ring image. Photo-detectors reconstruct time and position of incoming particles. In the forward endcap particle identification is provided by an additional aerogel ring-imaging Cherenkov detector.

Electromagnetic Calorimeter

The electromagnetic calorimeter (ECL) with an inner radius of 1250 mm in the barrel part is made out of 8736 CsI(Tl) scintillator crystals. Photons and charged particles interact with the detector material and lead to the emission of scintillation light. This makes it possible to detect photons, measure the energy of particles and separate electromagnetic and hadronic interactions due to a different scintillation light response. Each crystal in the barrel has a trapezoidal geometry with a front face area of $4.5 \times 4.5\,\mathrm{cm}^2$, a rear face area of $5 \times 5\,\mathrm{cm}^2$ and a nominal length of 30 cm. Two photodiodes are glued to the rear face of each crystal. Crystals are also installed in the forward and backward endcaps and the ECL provides an overall coverage of 90% of the solid angle.

$K_L - \mu$ Detector

The $K_L - \mu$ detector is the outermost detector placed around the superconducting solenoid. Its sandwich structure is made out of alternating 4.7 cm thick iron plates and active detector elements. This subdetector detects and helps to distinguish muons and K_L particles. Muons will create a track within the active elements, while K_L particles shower hadronically. Together with the central drift chamber information both particles can distinguished. The iron plates additionally function as a magnetic flux return for the solenoid. In order to cope with high background hit rates the innermost layer and both endcaps use scintillator strips for the active elements. The outer layers use glass-electrode resistive plate chambers as active material.

3.2 Electromagnetic Calorimeter

This chapter explains the processes within the ECL in more detail. The first part focuses on the mechanisms by which particles interact with the detector material and deposit energy within. This energy deposition is converted into an electrical signal from the scintillation light, as described in the second part. The last part of this chapter explains the signal chain and reconstruction of the recorded signals.

3.2.1 Particle Interactions in the Calorimeter

At Belle II, the most important ways particles deposit energy within the ECL are ionisation, electromagnetic showers, and hadronic showers. All of these processes are described in great detail in the literature. This chapter is based on [24].

Ionisation

When charged particles enter a material they lose energy through ionisation. The particles transfer energy to the atoms of the detector material, which then eject electrons. The mean energy loss of the initial particle per unit length is described by the Bethe-Bloch equation, which depends on characteristics of the particle and the material. The energy loss versus particle momentum, according to the Bethe-Bloch equation for CsI(Tl), is shown in Figure 8. Heavier particles, like protons, lose more energy on a given path length due to ionisation than lighter particles, like muons. Particles, which are minimal ionising, will deposit roughly 200 MeV in the 30 cm long crystal and will travel further into the $K_L - \mu$ detector. Additionally lower momentum and higher charge leads to higher dE/dx. This makes α particles highly ionising.

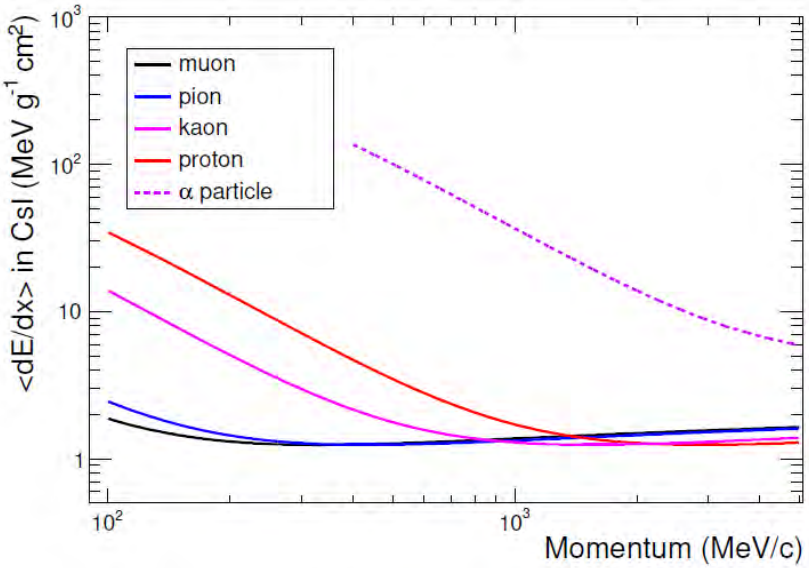


Figure 8: Mean energy loss per unit length, calculated with the Bethe-Bloch equation with the material properties of CsI(Tl), from [25].

Electromagnetic Shower

Electrons lose energy mainly by bremsstrahlung, due to their small mass. In this case, an electron emits a photon in the Coulomb field of a nucleus. Photons deposit energy by pair production, Compton scattering or the photoelectric effect, depending on the energy of the photon. For the energies at Belle II pair production is the most relevant. In this case the photon interacts with a nucleus and creates an electron-positron pair. The electron and positron can then again undergo bremsstrahlung. This repetition of photon emission and pair production leads to a cascade of electrons, positrons and photons, which is known as an electromagnetic shower. Such a shower is visualised in Figure 9. An important quantity of the material is the radiation length X_0, which describes the mean distance in the material over which an electron loses $1/e$ of its energy as well as the mean free path of pair production. X_0 depends on characteristics of the material and is 1.65 cm for CsI(Tl). The shower ends, if the energy of the photons is below 1.02 MeV and they are not able to undergo further pair production, and if the energy of the electrons is below a material dependent critical energy threshold E_c.

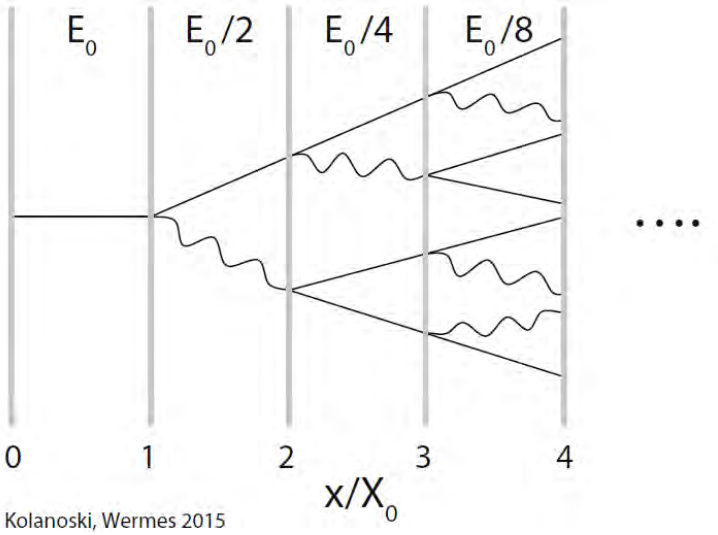


Figure 9: Sketch of an electromagnetic shower with an initial electron with an energy E_0, from [24].

Hadronic Shower

Hadronic showers are more complex than electromagnetic showers, since a large number of different processes happen simultaneously, as visualised in Figure 10. An incoming hadron interacts strongly with the protons and neutrons of the detector material. In this internuclear interaction secondary particles are produced and the nucleus can be excited. The most common secondary particles are pions, kaons, neutrons and protons. Charged secondary particles, like $\pi^{\pm}$, will ionise the detector material and can themselves interact with other nuclei, leading to a hadronic cascade. Due to elastic collisions with nuclei, those particles can also change their direction. Neutral secondary particles, like neutrons, propagate without ionising the material and can interact far from the original cluster, leading to so-called *hadronic split-offs*. Short-lived neutral hadrons, like K_S^0, will only travel for a short time until they decay into charged particles, which further interact with the detector. Moreover, a π^0 will decay immediately into two photons, which start an electromagnetic shower. In addition, when the nuclei de-excite, they can either undergo evaporation or fission. In case of nuclear evaporation low energy hadrons are emitted, which are highly ionising due to their low momentum.

For fission, the unstable nucleus will emit multiple neutrons and will break apart. Furthermore, there is also an invisible component within the shower, from energy which is absorbed to break up nuclei and from particles which escape without interaction.

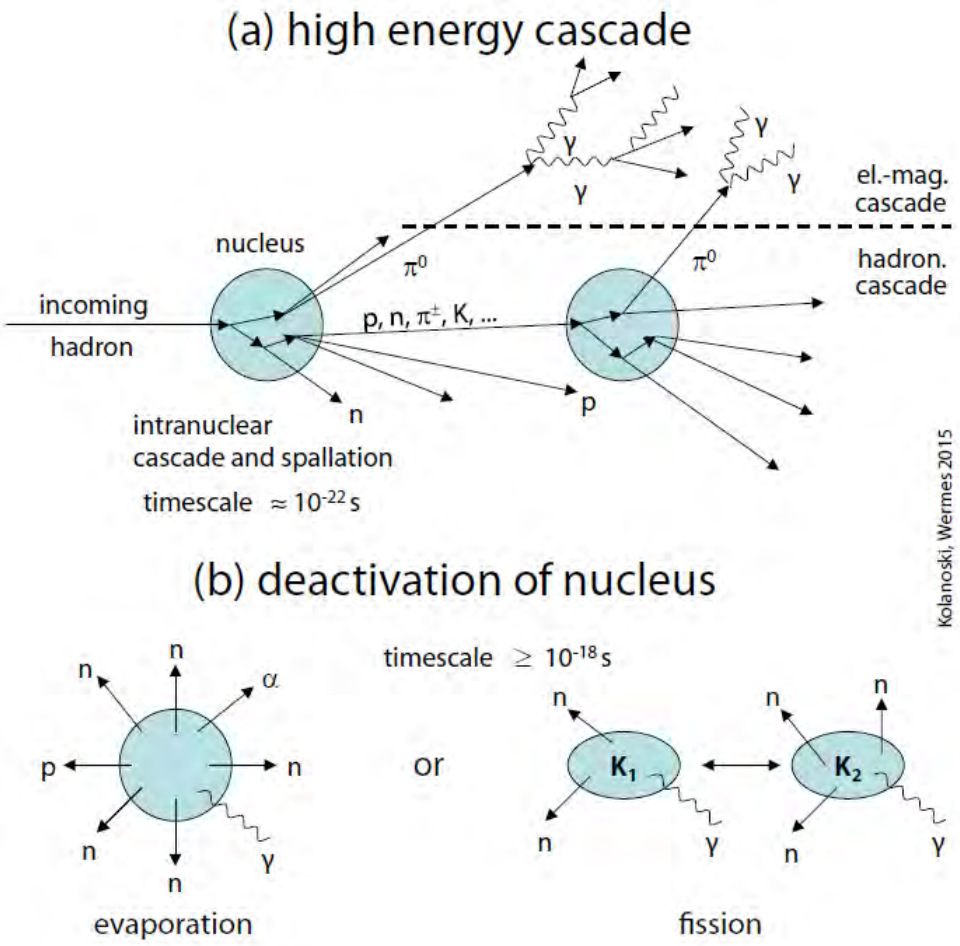


Figure 10: Sketch of a hadronic shower, from [24]

3.2.2 Scintillation Emission in CsI(Tl)

The deposited energy stimulates the CsI(Tl) crystals of the ECL detector to emit scintillation light, as described in [24]. If energy is deposited and absorbed within the crystal, an electron is excited from the valence band to the conduction band, leaving a positively charged hole in the valence band. The electron de-excites after a characteristic time scale by emitting a photon. In order to prevent reabsorption of this photon one uses doped materials, in case of Belle II the crystals are doped with thallium.

The light is converted into an electrical signal by photomultiplier tubes at the end of the crystal. Due to the photoelectric effect, electrons are ejected if the photo-cathode absorbs light. These electrons are accelerated by an electric field and focused onto an electron multiplier. This way a small number of electrons is amplified into a current, which can be measured. This current is proportional to the light intensity of the scintillation light and therefore to the energy deposition inside the crystal.

The scintillation response of CsI(Tl) is empirically known to depend on the interaction type underlying the energy deposition [26, 27]. The time structure of the scintillation light output can be modelled by a sum of exponential functions with different decay times due to different scintillation components. Highly-ionising particles have an additional fast scintillation component compared to particles with low dE/dx. Due to this, the shape of the scintillation signal depends on the nature of the energy deposition. This makes it possible to discriminate between electromagnetic and hadronic interactions based on the corresponding pulse shapes [26]. Furthermore, it makes it possible to distinguish between hadrons with different dE/dx, like protons and α-particles [27].

3.2.3 Reconstruction

In the next step, the signal of the scintillation light is processed in order to reconstruct the deposited energy within each crystal. This is described in detail in [25, 28] and summarised in the following.

Signal Chain

The signal chain is described in [29]. Two photodiodes per crystal convert the scintillation light into electrical signals which are amplified by a pre-amplifier circuit. The sum of the signals from both diodes are further shaped by a shaping amplifier, called a "ShaperDSP". This signal is digitised by Analog-to-Digital converters (ADC) at a sampling rate of 1.76 MHz with 31 consecutive samples being read out per event. Two examples for simulated pulse shapes, which are already shaped by the signal chain, are shown in Figure 11. It can be seen that the pulse shapes differ due to the different interactions of the corresponding simulated particles. The pion pulse has a faster rise and fall time and has an undershoot, while the photon pulse shape falls more slowly. Furthermore one can observe that the amplitude of the pion pulse shape is larger, even though both particles deposit the same energy in the crystal.

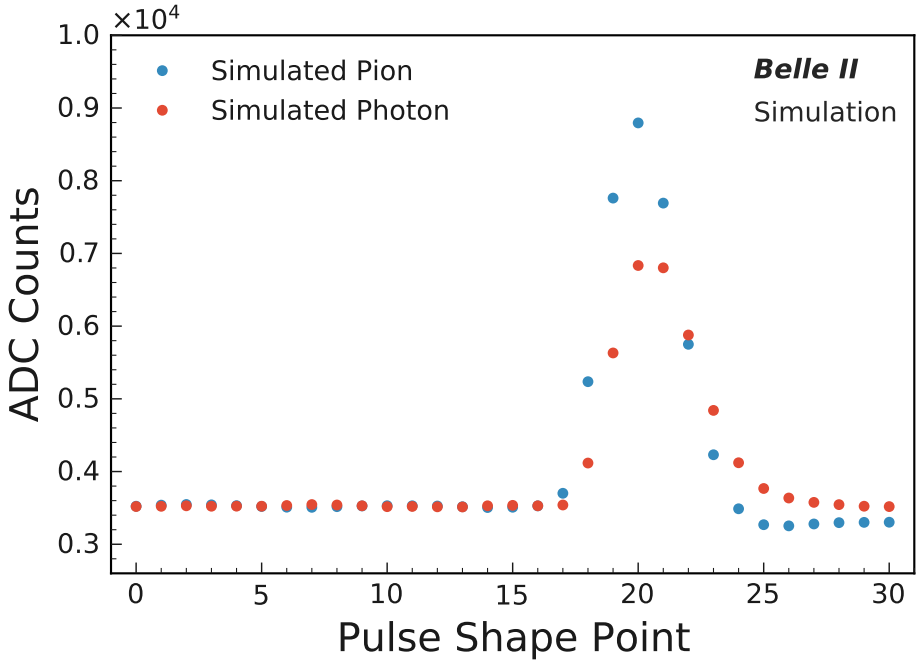


Figure 11: Pulse shapes from a simulated photon and a simulated pion. Both particles deposit an energy of 0.153 GeV in the crystal.

FPGA Fit

The digitised pulse shapes are fitted at the time of data acquisition with a single template in a Field Programmable Gate Array (FPGA). The template is calibrated to model the full signal chain of an energy deposit from a photon, called the *photon template*. The fit reconstructs the energy deposited within the crystal $E_{\mathrm{FPGA}}^{\mathrm{crystal}}$ and the time of the deposition. Since the signal chain was designed before pulse shape discrimination was studied in detail at Belle II, the firmware is not optimised for pulse shape discrimination. However, all digitised pulse shapes with $E_{\mathrm{FPGA}}^{\mathrm{crystal}} > 50\,\mathrm{MeV}$ are saved offline. The energy cut is necessary due to bandwidth limitations.

Offline multi-template fit

Offline, a multi-template fit is performed with two templates for crystals for which pulse shapes were recorded. One template is the photon template already used in the FPGA, the other template, called a *hadron template*, models the response of the signal chain for purely hadronic interactions. These templates need to be calibrated for each crystal individually, which requires the optimisation of 11 parameters, due to variations from crystal to crystal. With the fitted total deposited energy within the crystal $E_{\mathrm{Multi-template}}^{\mathrm{crystal}}$, the hadron component of the deposited energy $E_{\mathrm{Multi-template}}^{\mathrm{hadron}}$ and timing information are reconstructed. Finally, the hadron intensity is calculated, which is the fraction of the scintillation emission emitted by hadronic interactions relative to the total scintillation emission

$$\text{Hadron Intensity} = \frac{E_{\mathrm{Multi-template}}^{\mathrm{hadron}}}{E_{\mathrm{Multi-template}}^{\mathrm{crystal}}}. \tag{3}$$

In an ideal case of a photon-like pulse shape, the hadron intensity is zero and in case of highly ionising particles the hadron intensity is greater than zero. In Figure 12, two examples of pulse shapes and their corresponding multi-template fits are shown.

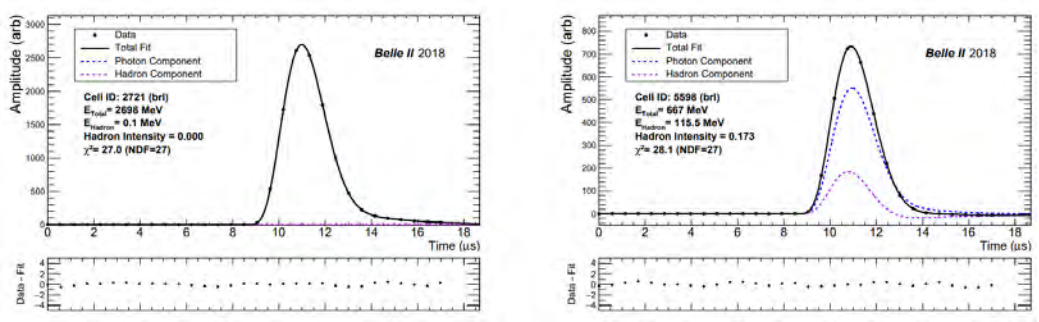


Figure 12: Typical pulse shapes from Belle II fitted with a photon and hadron template. The left side shows a pulse shape, typical for electromagnetic showers, with a hadron intensity close to zero. The right side shows an example for a pulse shape from a hadronic shower, with a higher hadron intensity, from [28]

Fittype

Additionally the multi-template fit assigns a fittype to each pulse shape. The fittype can be 0, 1, 2 or -1 and describes the quality of the pulse shape. Examples for each fittype, are visualised in Figure 13. For a fit with $\chi^2 < 60$, the pulse shape is assigned a fittype of 0. In the case $\chi^2 \geq 60$ the fit is performed again with an additional photon template with a time-offset to account for a pile-up photon, which causes a second peak in the baseline of the pulse. If the second fit fulfils $\chi^2 < 60$, the pulse shape is assigned a fittype of 1. However, if the χ^2 still exceeds the limit, the pulse shape is fit using a photon template and a diode-crossing template, which models when a photon deposits energy directly in the diodes at the end of the crystals. If this fit passes $\chi^2 < 60$, the pulse shape is assigned a fittype of 2 and the hadron intensity of the pulse shapes is set to 0. When the χ^2 requirement fails again, fittype -1 is assigned and for this pulse shape no reliable prediction of $E^{\text{crystal}}_{\text{Multi-template}}$ and $E^{\text{hadron}}_{\text{Multi-template}}$ is made. Currently, around 30% of all pulse shapes have a fittype of -1 and can not be used for pulse shape discrimination. However, it is expected that, with more luminosity this fraction will increase due to higher background levels.

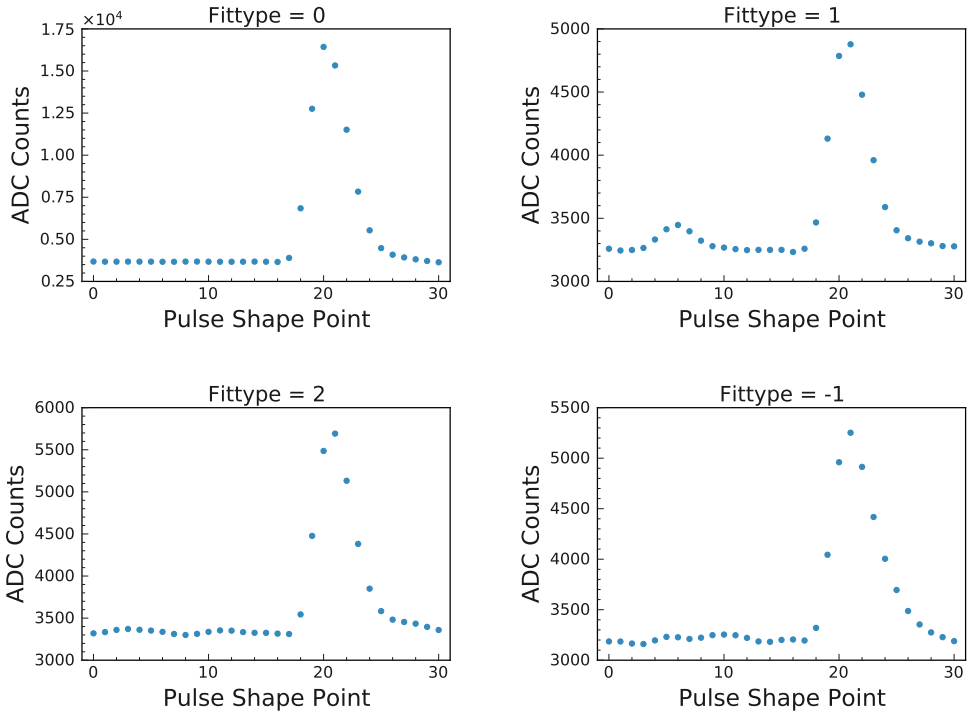


Figure 13: Simulated photon pulse shapes as examples for the four different fittypes.

Even though pulse shape discrimination works well at Belle II, it is worth investigating how to improve the pulse shape analysis. Current downsides of the multi-template fit are: the calibration, which is computationally expensive and needs to be done for every crystal separately, as well as pulse shapes with high noise where the fit fails, which is becoming more and more relevant. This thesis studies an alternative method using a neural network approach, since neural networks have the additional potential to be implemented on an FPGA for real time pulse shape discrimination. For this, various pulse shape samples are needed for training and testing procedures. How these pulse shape samples are created is the topic of the following chapter.

4 Simulated and Experimental Data Sets

Various types of pulse shapes are used to investigate improvements to the Belle II pulse shape discrimination technique. In particular pulse shapes associated with electromagnetic and hadronic interactions are needed. This study focuses on pulse shapes from photons as examples of electromagnetic interactions and pulse shapes from charged pions as examples of hadronic interactions. The neural network based pulse shape analysis technique employed in this work is based on supervised learning and requires labelled data with associated reference quantities. In order to generate labelled training sets, simulated events are used. This enables comparisons to the true energy $E_{\text{Truth}}^{\text{crystal}}$ and the true hadron intensity. The performance of the pulse shape analysis method can also be studied using simulation and to some degree also on data. This chapter gives an overview of the event simulation and the experimental data sets used. The first part of the chapter focuses on the simulation methods and settings as well as the data description. The second part presents the selection criteria, which are applied to select samples of photons and charged pions.

4.1 Event Simulation

The interaction of particles with the Belle II detector is simulated with GEANT4 [30]. GEANT4 simulates the event topology and the interaction of particles in the detector as well as the resulting detector response. Since GEANT4 does not simulate the dE/dx dependent scintillation light response of CsI(Tl) by default, it is added to the simulation as described in [28, 31]. In order to add detector noise to the pulse shapes the simulation includes run dependent beam induced background overlays. These overlays are taken from randomly triggered data events. Sources for beam induced background are described in detail in [32]. The Monte Carlo *truth matching* algorithm is run on all simulation samples. This associates the reconstruction level objects, for example tracks and clusters, with the underlying generated particles.

4.1.1 Single-Particle Simulation

The Belle II Particle Gun event generator simulates events with only a single propagating particle [33]. This way one can simulate its interactions with the detector as well as the detector response. This study used the prerelease-05-00-00b of the Belle II software [34]. A set of 1000000 photon events and a set of 1000000 charged pion events were generated. For both sets the following simulation settings were applied:

- Momentum generated uniformly in the range of $0.05\,\mathrm{GeV/c} - 6\,\mathrm{GeV/c}$
- Polar angle θ generated uniformly in the range of $30° - 125°$
- Azimuthal angle ϕ generated uniformly in the range of $0° - 360°$
- Beam induced background overlay from experiment 12 run 3363, 3402, 4074 or 5649
- Option to add random time shift, which is uniformly distributed in the range of $\pm 2000\,\mathrm{ns}$

At the time the pulse shape is recorded by the FPGA, the full reconstruction of the interaction time t_0 is not completed, which can lead to a shift of the pulse shape, as shown in Figure 14. The random time shift accounts for this pulse shape shifting.

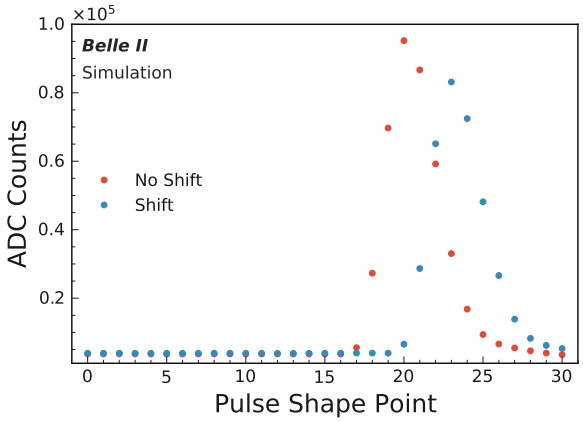


Figure 14: Example of shifted pulse shapes. Both pulse shapes are from different simulated photon events each with fittype 0 and $E_{\mathrm{Truth}}^{\mathrm{crystal}} = 3.23\,\mathrm{GeV}$.

4.1.2 Full-Event Simulation

In addition to single particle simulations, realistic physics events were simulated with the Belle II software release 04-02-09. Samples of generic modes ($B^0\overline{B^0}$, B^+B^-, $u\overline{u}$, $d\overline{d}$, $s\overline{s}$, $c\overline{c}$ and $\tau^+\tau^-$) and $e^+e^- \to \mu^+\mu^-\gamma$ were generated using background overlays from experiment 10 data. The simulation is equivalent to an integrated luminosity of $10\,\mathrm{fb}^{-1}$.

4.2 Data

The data used was taken in 2019 in experiment 10. All good runs were used covering run number 3129 to 5902, which corresponds to an integrated luminosity of $3.741\,\mathrm{fb}^{-1}$. These runs were reprocessed to reconstruct the pulse shapes using the same release as for the simulations (release 04-02-09 of the Belle II software).

4.3 Selection

Two kinds of processes are studied in data and full-event simulation:

$$e^+e^- \to \mu^+\mu^-\gamma. \tag{4}$$

as a sample of photons and

$$D^{*+} \to \pi^+ D^0\,[\to \pi^- K^+] \tag{5}$$

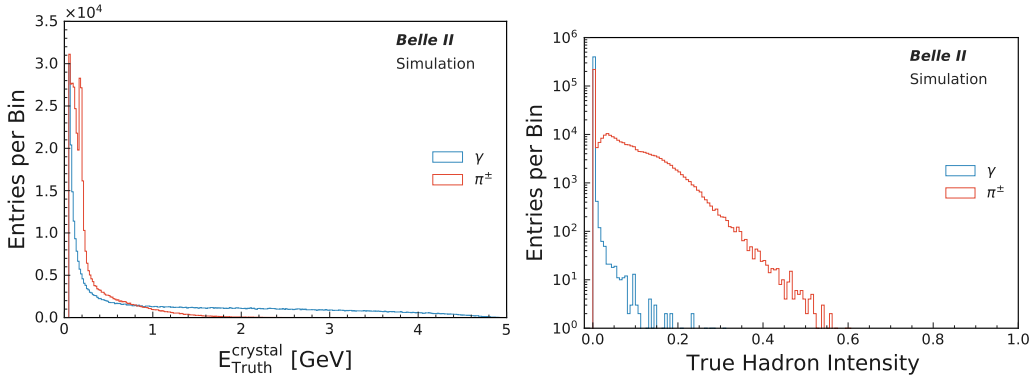
as a sample of charged pions. These two processes are selected by applying the following criteria.

4.3.1 Single-Photon and Single-Pion Selection

For the single-particle simulation of photons and pions the same selection criteria were applied:

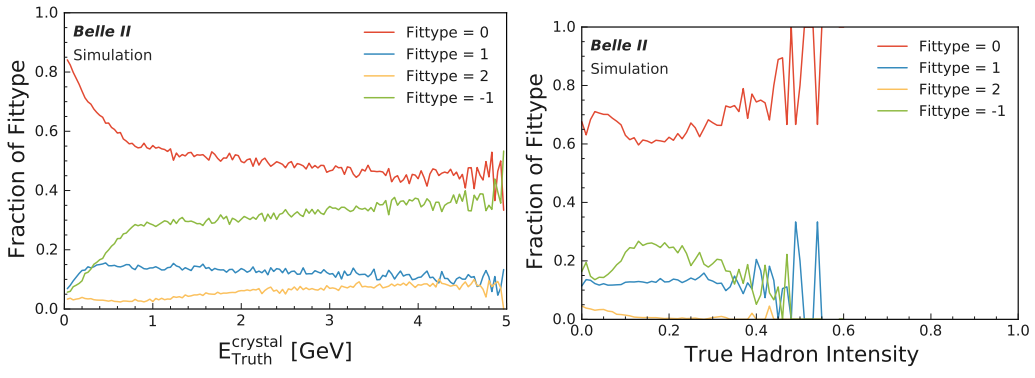
- $E_{\mathrm{Truth}}^{\mathrm{crystal}} > 50\,\mathrm{MeV}$
- Crystal within the barrel: ECL Cell ID: 860 to 8022

In total 800000 pulse shapes were selected to study the neural network approach, of which 400000 photon and 400000 pion pulse shapes. A time shift was applied to half of all pulse shapes, while the rest have no time shift. The distributions of crystal energy and hadron intensity of this pulse shape selection as well as the fraction of fittypes is shown in Figure 15 and Figure 16.



(a) (b)

Figure 15: Distribution of the $E^{\text{crystal}}_{\text{Truth}}$ (a) and the true hadron intensity (b) sampled by photon and pion pulse shapes.



(a) (b)

Figure 16: Fraction of fittypes versus the $E^{\text{crystal}}_{\text{Truth}}$ (a) and versus the true hadron intensity (b).

4.3.2 $e^+e^- \to \mu^+\mu^-\gamma$ Selection

As a sample for electromagnetic interactions, radiative muon events were selected from data and full-event simulation. Two tracks with muon mass hypothesis and a trackless energy deposition within the calorimeter were reconstructed. The following criteria were applied for the event selection:

- Minimal reconstructed cluster energy: $E_{\text{cluster}} > 1\,\text{GeV}$
- Cluster within the angular acceptance of the barrel: $32.2^\circ < \theta_{\text{cluster}} < 128.7^\circ$
- Total invariant mass of $\mu\mu\gamma$: $M_{\text{virtual photon}} > 9\ \text{GeV/c}^2$
- Both tracks have a muon identification probability: Muon ID > 0.1
- Tracks originate from interaction point: transverse distance of vertex to interaction point $|dr| < 2\,\text{cm}$ and longitudinal distance of vertex to interaction point $|dz| < 5\,\text{cm}$
- Both muon momenta: $p_\mu > 1\,\text{GeV/c}$

The pulse shapes from those remaining events need to further satisfy:

- $E^{\text{crystal}}_{\text{FPGA}} > 50\,\text{MeV}$
- Crystal within the barrel: ECL Cell ID: 860 to 8022
- Multi-template fit did not fail ($E^{\text{crystal}}_{\text{Multi-template}} \neq -1\,\text{GeV}$)

This selection results for simulation in about 600000 photon pulse shapes and for data in roughly 300000 photon pulse shapes.

Backgrounds of the selection were not studied. The largest contribution is expected to be from muons, which also interact electromagnetically. Thus the sample remains largely uncontaminated from hadronic contributions to the pulse shapes.

4.3.3 $D^{*+} \to \pi^+ D^0[\to \pi^- K^+]$ Selection

In order to have charged hadron pulse shapes a D^* selection was performed on full-event simulation and on data, inspired by [35]. A D^0 was reconstructed with two opposite charged particles with kaon and pion mass hypothesis. The D^0 was combined with a pion of the opposite charge to the other pion to reconstruct a D^* meson. The following criteria needed to be fulfilled by the event:

- Tracks originate from interaction point: transverse distance of vertex to interaction point $|dr| < 2\,\mathrm{cm}$ and longitudinal distance of vertex to interaction point $|dz| < 4\,\mathrm{cm}$
- $D^0 \to K^-\pi^+$ with a reconstructed D^0 mass: $1.8\,\mathrm{GeV/c^2} < M_{D^0} < 1.95\,\mathrm{GeV/c^2}$
- D^* momentum in the centre-of-mass frame: $p^*_{D^*} > 2.5\,\mathrm{GeV/c}$
- Difference of the reconstructed D^* mass and reconstructed D^0 mass: $0.1438\,\mathrm{GeV/c^2} < M_{D^*} - M_{D^0} < 0.1468\,\mathrm{GeV/c^2}$

The pulse shapes from those remaining events need to further satisfy:

- $E^{\mathrm{crystal}}_{\mathrm{FPGA}} > 50\,\mathrm{MeV}$
- Crystal within the barrel: ECL Cell ID: 860 to 8022
- Multi-template fit did not fail ($E^{\mathrm{crystal}}_{\mathrm{Multi-template}} \neq -1\,\mathrm{GeV}$)

This results for simulation in about 20000 charged pion pulse shapes and for data in about 10000 charged pion pulse shapes.

Backgrounds of the selection were not studied. The largest contribution is expected to be from kaons, which also have a hadronic component.

5 Artificial Neural Networks

Artificial neural networks are a multivariate method of information processing. They are found in increasingly many areas of data analysis and often show improved performance compared with conventional methods. In this work, neural networks are evaluated as an analysis tool for the ECL crystal pulse shapes. The first part of this chapter gives a short introduction to the relevant fundamental concepts of neural networks based on [36–40]. The second part focuses on the specific neural network architecture used for this study.

5.1 Machine Learning and Neural Networks

Neural networks are models for information processing, consisting of a set of neurons with weighted connections between them. These weights are adjusted to a specific task using a (machine) learning algorithm. The architecture of a network describes how the neurons are organised into layers and how they are connected. This thesis focuses on *fully-connected feed-forward* neural networks. These networks process information in only one direction, from the input layer to the output layer, and every single neuron is fully connected to the neurons from the adjacent layers.

A simple example of a fully connected network is sketched in Figure 17. The first layer of the network is the input layer, where data enters the network. This is followed by a series of hidden layers. These neurons receive a number of inputs coming from the previous layer. Figure 18 visualises, how the input is processed in a single neuron. Each input x_i is multiplied with a weight w_i, the weighted inputs are summed and a bias Θ is subtracted. The resulting value y is passed through an activation function f. This way the output $f(y)$ of one hidden layer becomes the input of the next hidden layer. The last layer is the output layer, which processes the outputs of the previous hidden layers and returns the networks response.

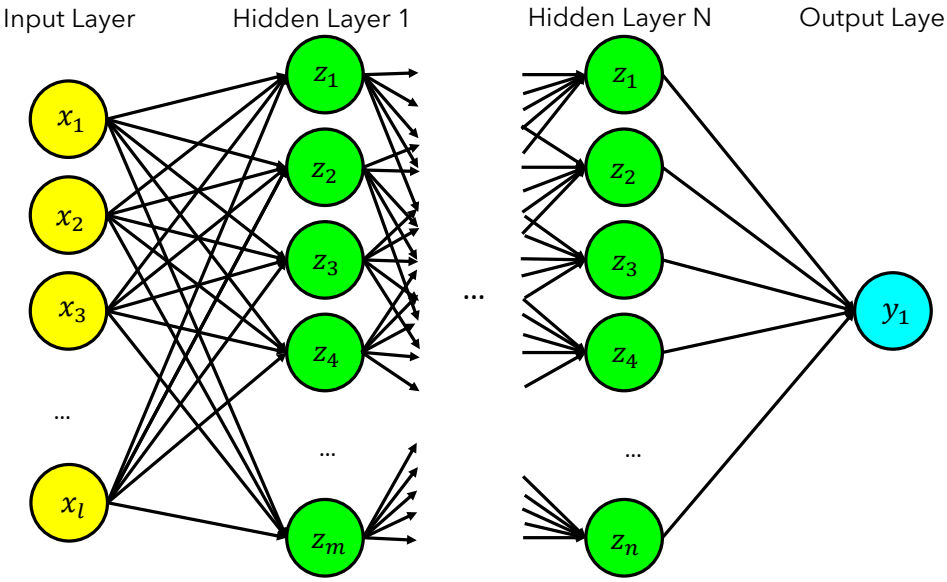


Figure 17: Schematic drawing of a fully-connected feed-forward network, modified from [41].

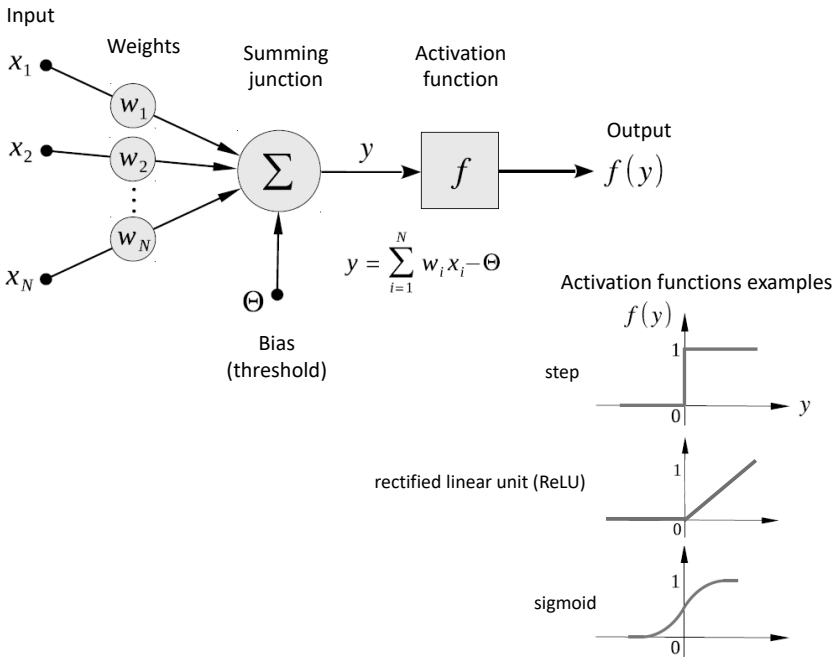


Figure 18: Sketch of how a single neuron processes N inputs to a single output $f(y)$, modified from [42].

During an iterative learning process, called training, the weights and biases are adjusted. There are different methods for learning and the following discussion focuses only on supervised learning. During supervised training the network is presented with a set of training inputs and desired outputs, called targets. A loss function evaluates the difference of targets and the corresponding responses of the network. The goal of the training is to minimise the loss by adjusting the weights and biases. This is done by a back-propagation algorithm, which calculates the gradient of the loss in respect to the weights and biases. Those parameters are optimised for the next training cycle in the direction of minimal loss. The step size of this is defined by the so-called *learning rate*. Networks tend to perform better when the learning rate is reduced over the training cycles, since the slope of the loss function usually decreases closer to the minimum of the loss. In order to reduce computational memory usage during this training process, the training set is split into batches. One full cycle, where the network saw all training inputs once, is called an *epoch*. The number of epochs is not directly limited, however, training can become computationally expensive and a network might run into overtraining, if the number of epochs is too large.

5.2 Neural Network Architecture of this Work

The goal for a neural network based pulse shape analysis is to feed a pulse shape to a regression network, which returns relevant quantities, or to a classifier network, which separates pulse shapes into different classes. Several research groups have demonstrated that neural networks can be applied to pulse shape analysis [41–45]. The applications and architectures differ, since some groups use this approach to reconstruct quantities like pulse height or rise time [43, 44], while others use it to classify pulse shapes for background rejection [41, 42, 45]. However, all groups found a similar or improved performance compared with analytical methods and faster computation.

For the application described in this thesis, the goal was to reconstruct the crystal energy and hadron intensity for each pulse shape. The task was

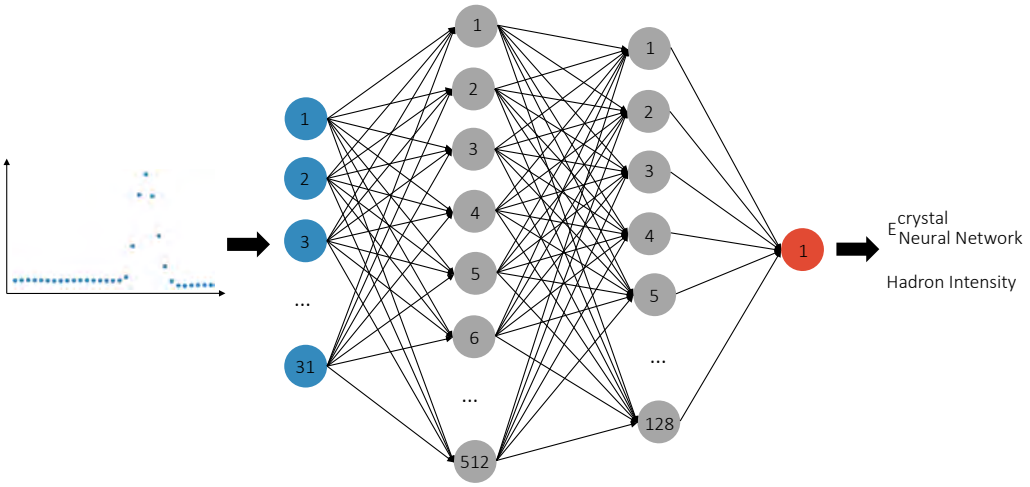
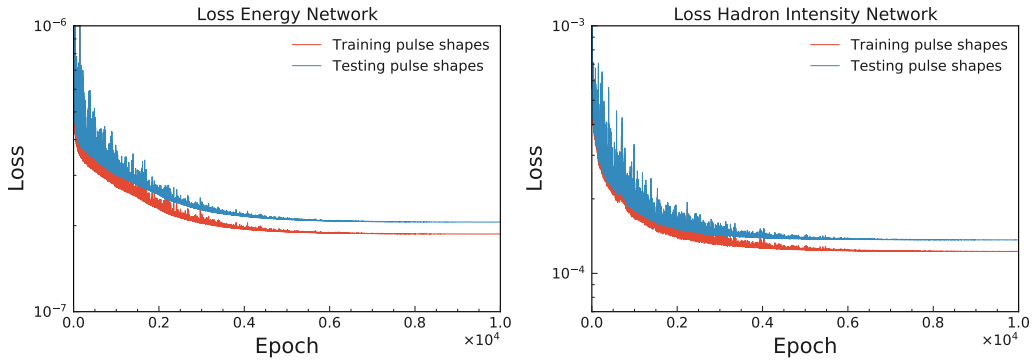


Figure 19: Sketch of the architecture for the energy network and hadron intensity network.

split between two regression networks, in order to adjust the weights for each quantity individually. One network returns the crystal energy, referred to as the *energy network*, and the other network returns the hadron intensity, referred to as the *hadron intensity network*. Instead of a regression network, one could set up the hadron intensity network as classifier, which is trained to separate pulse shapes from electromagnetic and hadronic interactions. However, a regression network gives the option to not only differentiate between electromagnetic and hadronic interactions but also between particles with different dE/dx, for example protons and α particles.

This study used the open-source machine learning framework PyTorch [46]. During the study a range of architectures was tested. This included variations in the number of layers and neurons, as well as different initial learning rates and activation functions. A good performance was observed, when both networks have the following architecture: the input layer has 31 neurons, two hidden layers, which have 512 and 128 neurons respectively, and the output layer has one neuron. This architecture is sketched in Figure 19. This leads to 82177 free parameters for each network. All hidden layers use a rectified



(a) (b)

Figure 20: Loss of the energy network (a) and the hadron intensity network (b). Training pulse shapes refer to pulse shapes used for training, while testing pulse shapes are unknown for the networks.

linear unit activation function [47]. As loss function both networks use the mean square error [48] and an adaptive moment estimation optimiser [49, 50]. The energy network has an initial learning rate of 0.0001 and the hadron network has an initial learning rate of 0.001. In both cases the learning rate is reduced by 4% in every 50^{th} epoch. Since neural networks tend to perform better if the input and target values are between 0 and 1, all pulse shapes were scaled by $\frac{1}{2^{18}}$, since 2^{18} corresponds to the maximal possible ADC count. The target for the energy network is converted from energy in GeV to ADC counts and scaled in the same way as the pulse shapes. The hadron intensity targets were not scaled since they are, by definition, between 0 and 1. The networks were trained with 720000 pulse shapes from simulated photons and pions in batches of 2000. The simulation and selection of those pulse shapes is described in detail in chapter 4. Both networks were trained for 10000 epochs on a CPU, which takes around 60 hours. The loss for both networks is shown in Figure 20, which stagnates after approximately 7000 epochs. A comprehensive optimisation of the hyper-parameters and the architecture needs to be done in a succeeding work.

6 Neural Network Performance

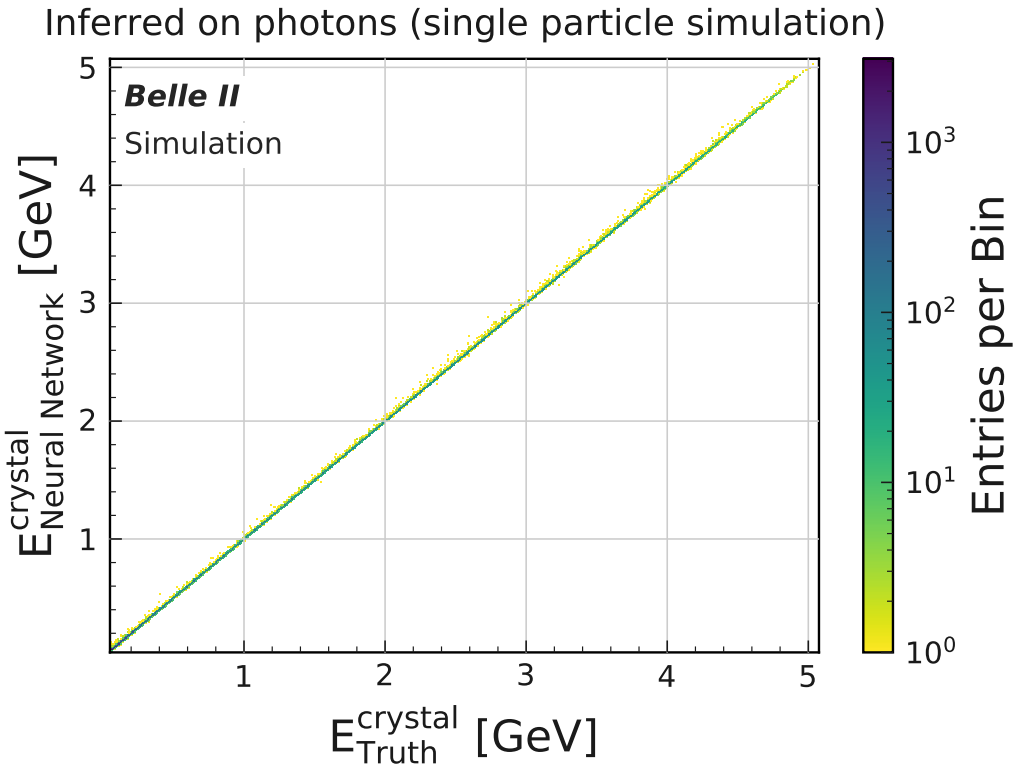
This chapter summarises the performance of the neural networks, introduced in chapter 5.2, inferred on the data sets, introduced in chapter 4. The first section describes the performance when inferred on single-particle simulation of photons and pions and compares it with the default methods of the FPGA and the multi-template fit. The performance is evaluated in terms of energy resolution, hadron intensity resolution, and separation power. The second section focuses on the robustness of the neural network when time shifts and unknown beam backgrounds are introduced. The third section describes how the neural network prediction for the crystal energy affects the cluster-energy resolution. The fourth section shows the neural network output when inferred on simulated single-proton events. The final chapter presents the performance of the neural networks inferred on full-event simulation and data.

6.1 Performance with Single-Particle Simulation

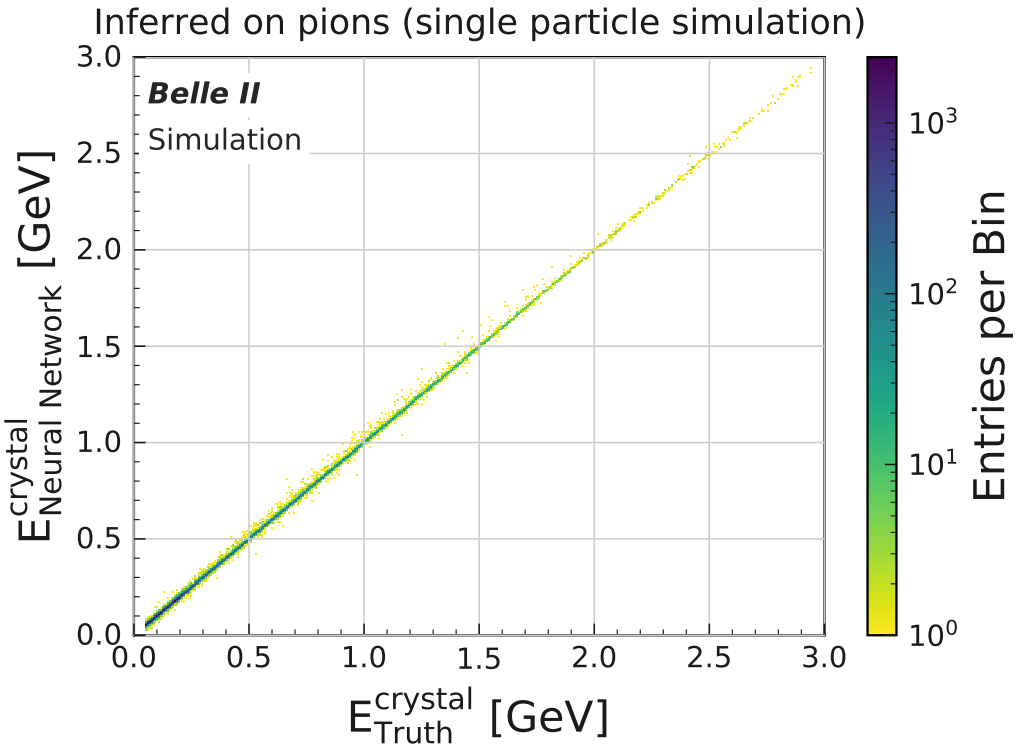
In this chapter the performance of the neural network is evaluated when inferred on single-particle simulation of photons and pions. The figures of merit are the crystal energy resolution, the hadron intensity resolution, and the separation power. The network outputs are compared with the FPGA and multi-template fit prediction. The comparison is performed separately for photons and pions and different fittype categories respectively.

6.1.1 Crystal Energy Prediction

A first step to study the performance of the neural networks is to compare the network response with the Monte Carlo truth value, as visualised in Figure 21 for photons and pions. For both particle types the neural network response is overall linear with respect to $E_{\text{Truth}}^{\text{crystal}}$, whereas pions appear occasionally as outliers. The input pulse shapes are already subjected to time shifts and represent all fittypes, this therefore shows a high degree of robustness in the response.



(a)



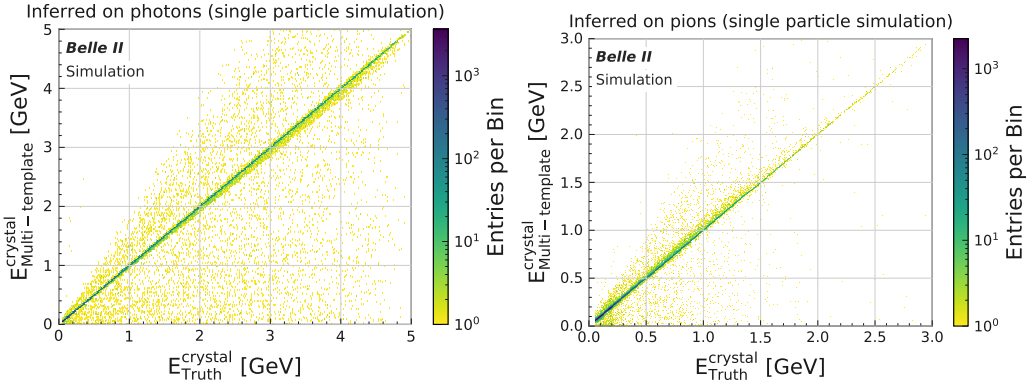
(b)

Figure 21: Linearity of the energy prediction from the neural network compared with the true crystal energy for photons (a) and charged pions (b).

The same comparison is done for the response of the multi-template fit in Figure 22 and in Figure 23, where for Figure 23 pulse shapes with fittype -1 are excluded. These plots show that the multi-template fit does not perform well for pulse shapes with fittype -1. For this reason the multi-template fit is only applied to pulse shapes with fittype $\neq$ -1 in the following.

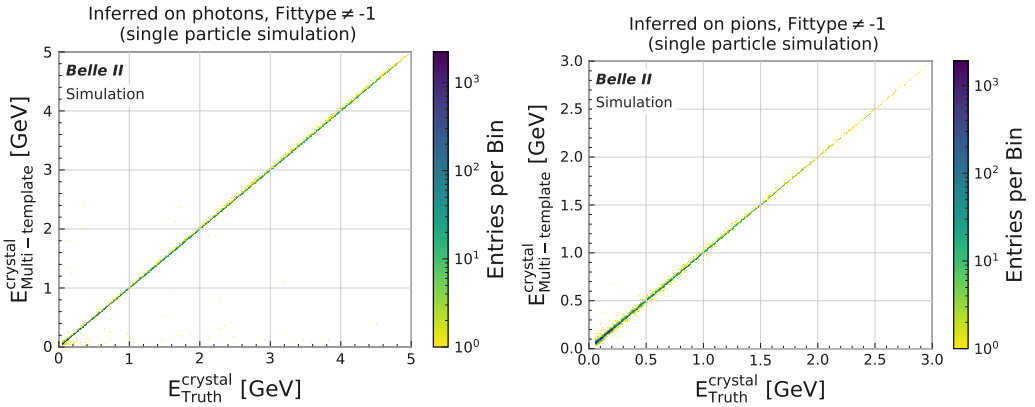
Figure 24 shows the comparison of the FPGA response to the Monte Carlo truth values for the crystal energy. The outliers of the FPGA fit for photons result from shifted pulse shapes, which were described in chapter 4.1.1. Since the FPGA is designed to use only photon templates for the fit, the energy reconstruction for pions is less accurate as seen in the broadening of the line. This is the reason why in the following the FPGA is only considered for photons.

Overall one finds a linear relation between the crystal energy reconstruction and the Monte Carlo truth values for all three methods. However the FPGA fit can only be applied to photons and shows shortcomings if the pulse shapes are shifted while the multi-template fit is restricted to fittypes 0, 1, and 2 in terms of the crystal energy reconstruction. The neural network has a good response without restrictions on the pulse shapes.



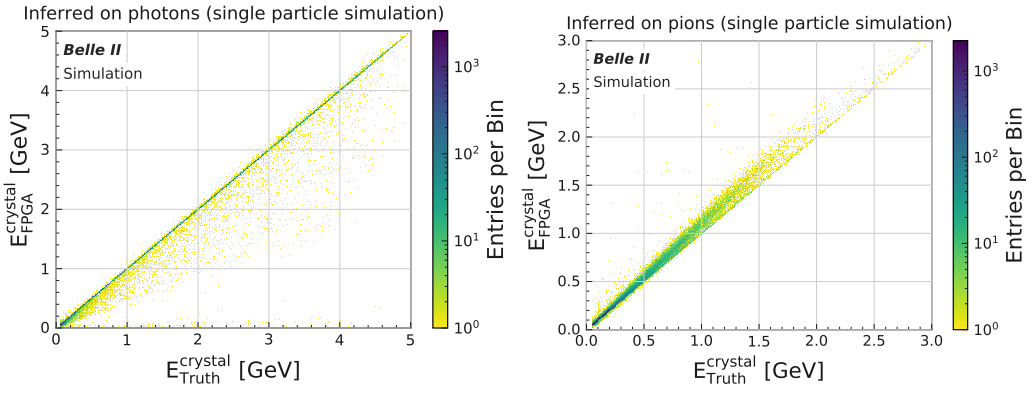
(a) (b)

Figure 22: Linearity of the energy prediction from the multi-template fit compared with the true crystal energy for photons (a) and charged pions (b) including all fittypes.



(a) (b)

Figure 23: Linearity of the energy prediction from the multi-template fit compared with the true crystal energy for photons (a) and charged pions (b) excluding fittype -1.



(a) (b)

Figure 24: Linearity of the energy prediction from the FPGA compared with the true crystal energy for photons (a) and charged pions (b).

6.1.2 Hadron Intensity Prediction

Since the multi-template fit makes no reliable hadron intensity prediction for pulse shapes with fittypes -1 and 2, these fittypes are excluded in all following plots.

For photons, the hadron intensity is expected to be zero without significant tails, as shown by the true hadron intensity in Figure 25. The multi-template fit and the neural network peak at zero, but the multi-template fit predicts negative hadron intensities more frequently than the neural network. One finds that the neural network has smaller tails than the multi-template fit. Overall the neural network shows a more precise hadron intensity prediction for photons than the multi-template fit.

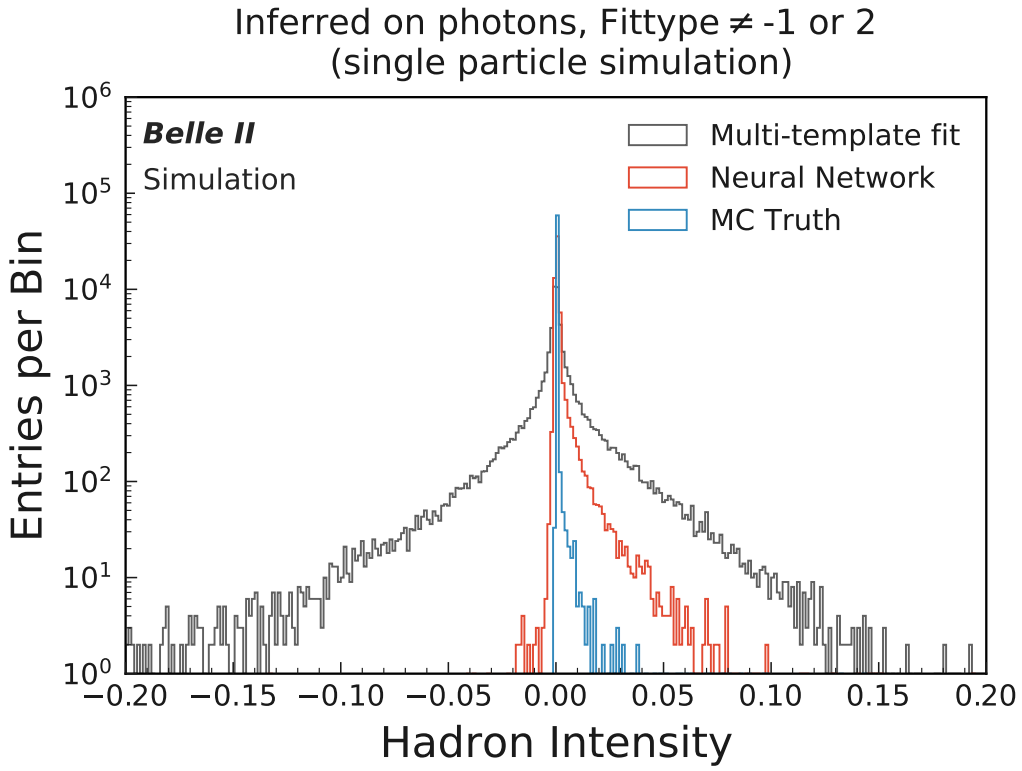


Figure 25: Hadron intensity for photons, excluding pulse shapes with fittype -1 and 2.

For pions one expects a predicted hadron intensity in the range of 0 to 0.5 with a peak at zero resulting from pions which only ionise and do not undergo hadronic interactions. This is shown in Figure 26. The multi-template fit predicts negative hadron intensities more often than the neural network and also underestimates the number of pulse shapes with a hadron intensity of zero. In Figure 27, the hadron intensity prediction of the neural network (a) and the multi-template fit (b) are shown as a function of the true hadron intensity. Both methods show a linear relation and again, one finds that the neural network predicts negative hadron intensities less often than the multi-template fit and results in fewer outliers overall.

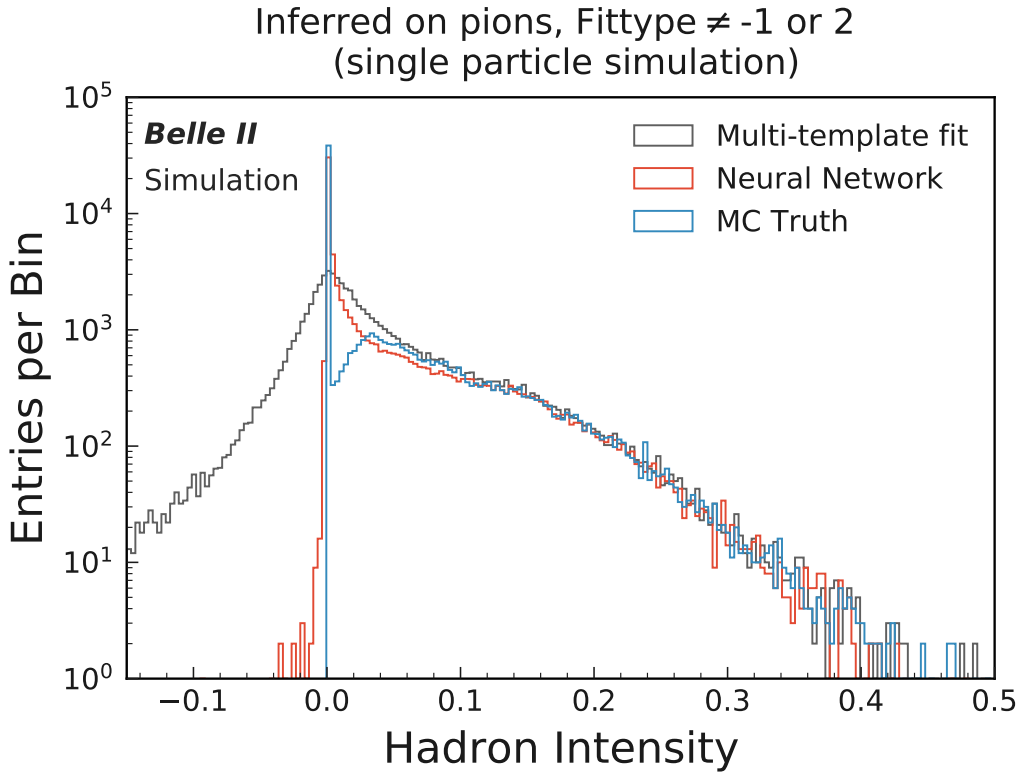
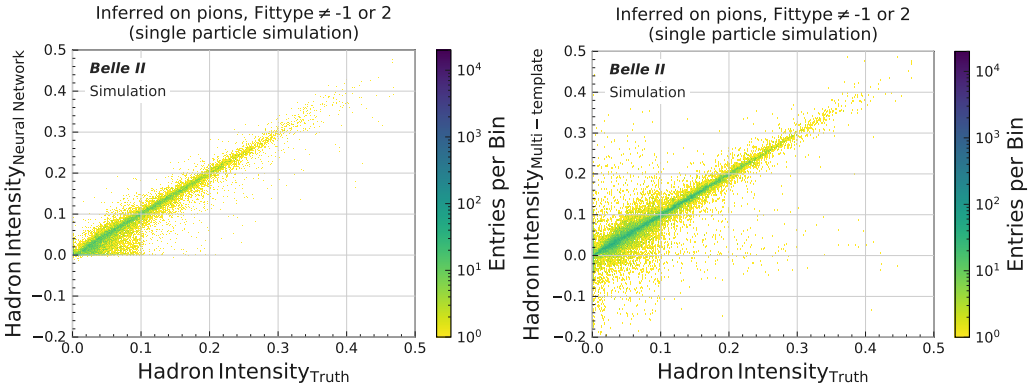


Figure 26: Hadron intensity of pions, excluding pulse shapes with fittype -1 and 2.

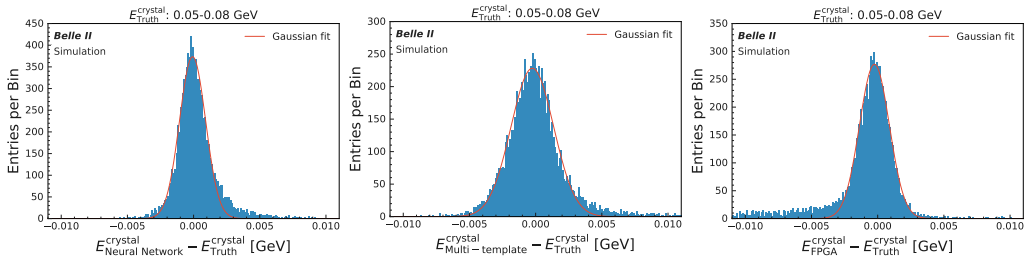


(a) (b)

Figure 27: Hadron intensity prediction for pions using the hadron intensity network (a) and the multi-template fit (b) compared with the true hadron intensity, excluding pulse shapes with fittype -1 and 2.

6.1.3 Energy and Hadron Intensity Resolution

To calculate the energy resolution, the difference of the predicted and the true value is evaluated in bins of true energy for all three methods. As an example, the resulting distributions for the range $\mathrm{E}_{\mathrm{Truth}}^{\mathrm{crystal}} = 50$ to 80 MeV are shown in Figure 28. The distributions are fitted with a Gaussian.



(a) (b) (c)

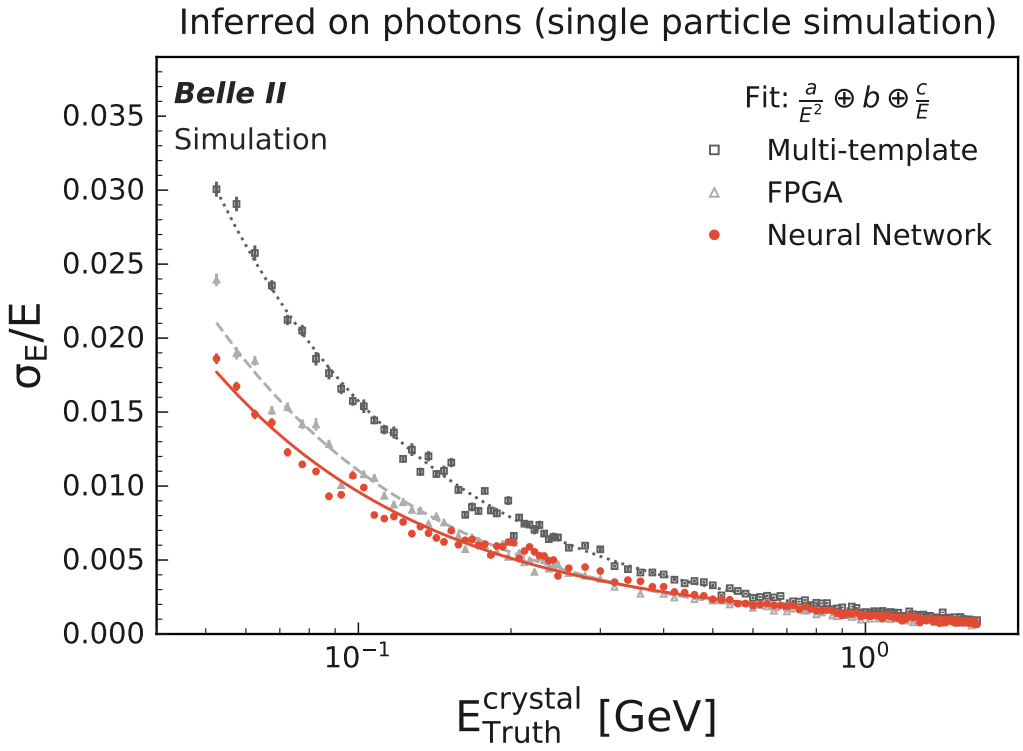
Figure 28: Difference between the predicted crystal energy and true crystal energy for photon crystals with a true energy of 50 MeV to 80 MeV, as predicted by the energy network (a), the multi-template fit (b) and the FPGA (c). Fitted with Gaussians.

The relative crystal energy resolution, defined as σ_E/E, is shown in Figure 29. For photons with $\mathrm{E}_{\mathrm{Truth}}^{\mathrm{crystal}} < 250\,\mathrm{MeV}$, the neural network improves the crystal energy resolution compared with the multi-template (FPGA) fit by up to 40% (14%). Above this energy, the neural network resolution is equal to that of the FPGA fit, which is slightly superior to the multi-template fit. For pions with $\mathrm{E}_{\mathrm{Truth}}^{\mathrm{crystal}} < 300\,\mathrm{MeV}$, the neural network improves the crystal energy resolution by up to 36%. Above this energy the resolution degrades slightly compared with the multi-template fit. Here one should note that the better performance in the low energy range could result from the composition of the training data, shown in Figure 15. The training data contained more low-energy events to account for the expected energy distribution during physics data taking.

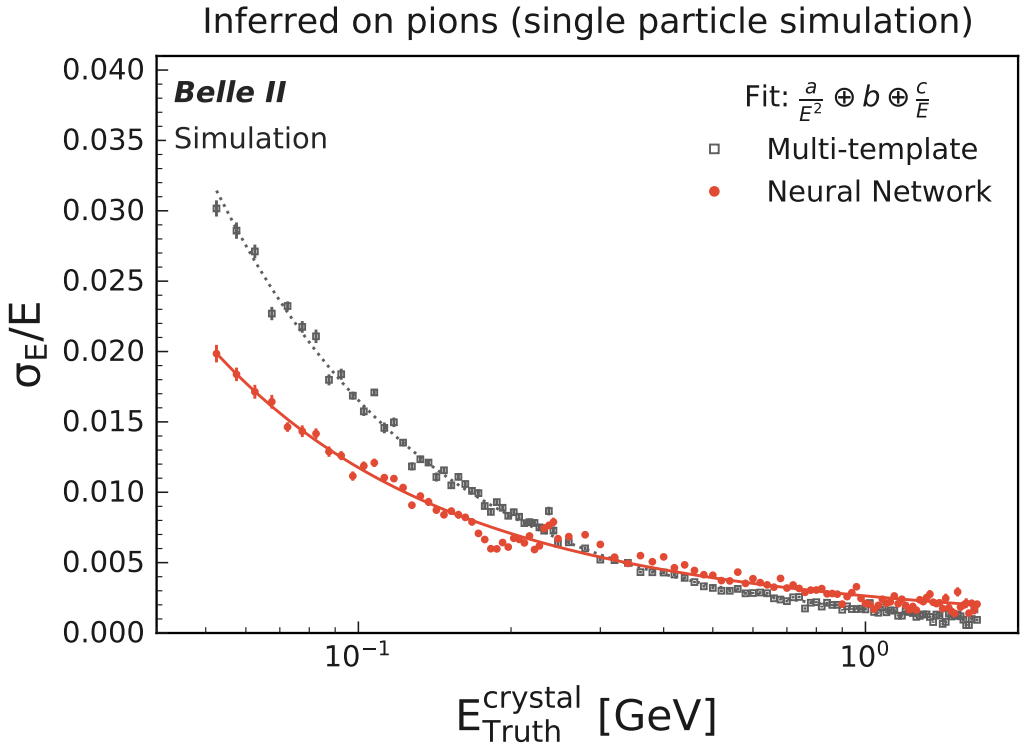
Analogously, the hadron intensity resolution for photons is calculated as a function of the crystal energy, shown in Figure 30a. The hadron intensity resolution is drastically improved by the neural network across all crystal energies. The reader is reminded that, in contrast to the multi-template fit, the neural network contains pulse shapes of all fittypes, which also shows the high degree of robustness of this result.

Furthermore the relative hadron-energy resolution is calculated for pions in dependence of the true hadron energy, shown in Figure 30b. One finds an improved resolution for pulse shapes with a small $\mathrm{E}_{\mathrm{Truth}}^{\mathrm{hadron}} < 5\,\mathrm{MeV}$. For larger hadron energies, the resolution of the multi-template fit and the neural network approach converge.

For completeness, all resolution plots are shown for each fittype individually in appendix A.1.

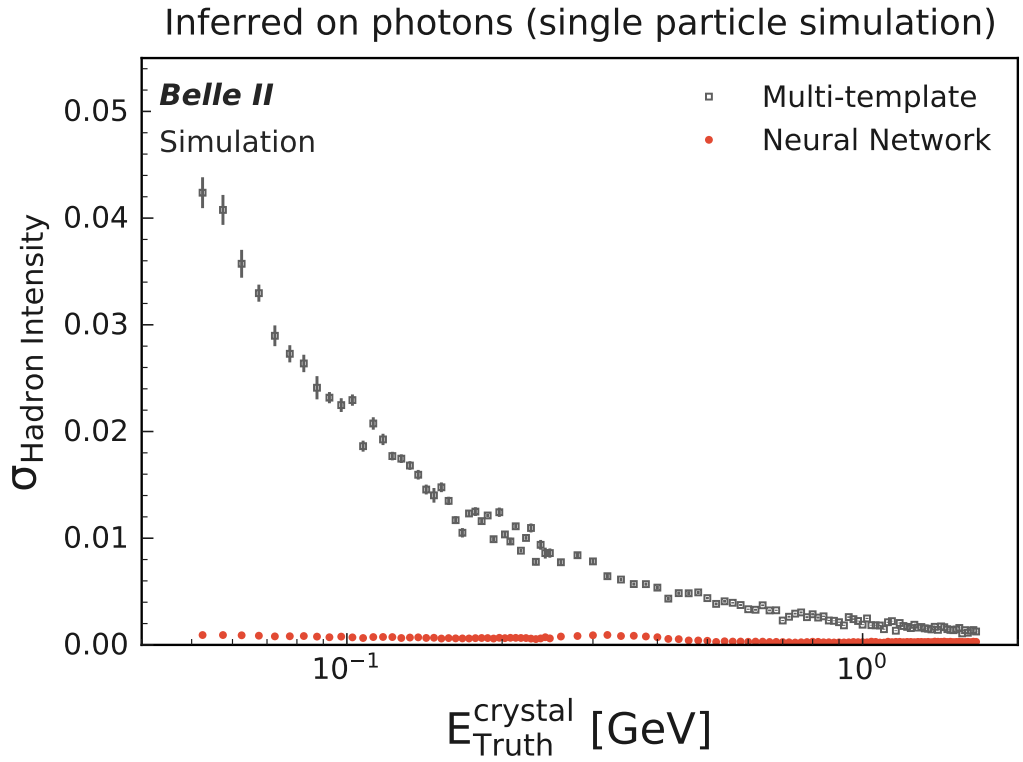


(a)

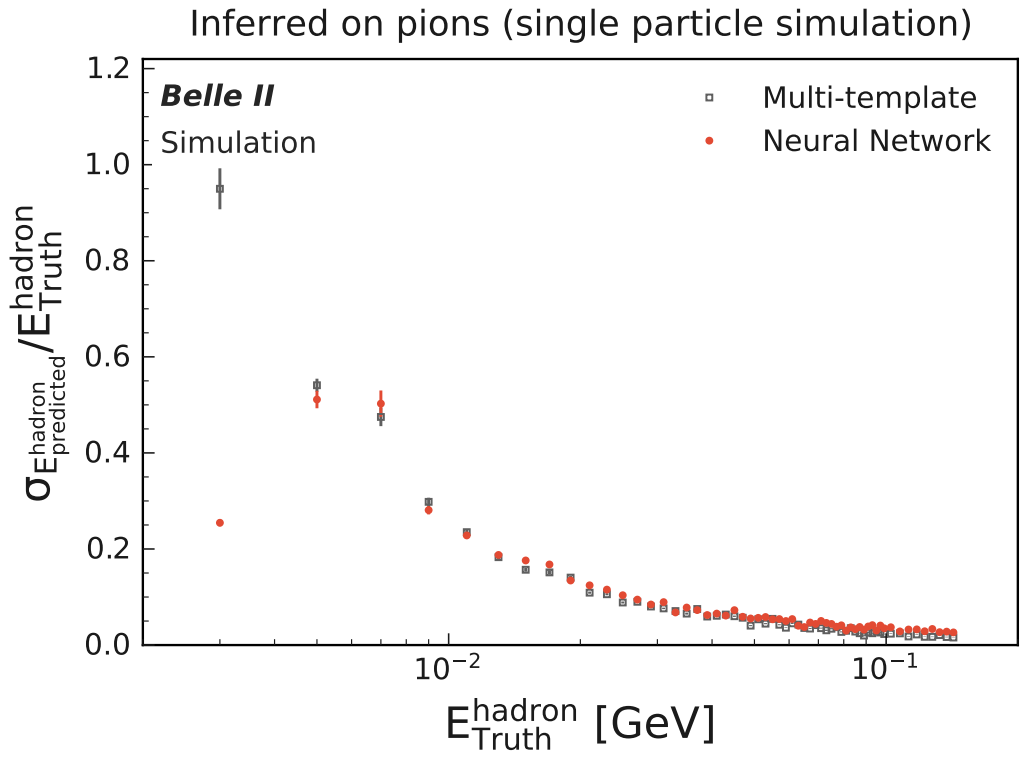


(b)

Figure 29: Relative crystal energy resolution for photons (a) and charged pions (b) fitted with the empirical energy-resolution function.



(a)



(b)

Figure 30: Hadron intensity resolution for photons (a) and and relative hadron energy resolution for charged pions (b).

6.1.4 Separation of Photons and Pions

In order to use the neural networks to discriminate between hadronic and electromagnetic interactions, a classification variable C, with

$$C = \frac{\text{Hadron Intensity}_{\text{predicted}} - \text{Hadron Intensity Offset}(E^{\text{crystal}}_{\text{predicted}})}{\sigma_{\text{Hadron Intensity}}(E^{\text{crystal}}_{\text{predicted}})}, \quad (6)$$

is defined as the hadron intensity, corrected for an energy-dependent offset, divided by the energy-dependent hadron intensity resolution from Figure 30a. The energy-dependent offset of the hadron intensity prediction is taken as the mean of the hadron intensity prediction for photons. The variable C is calculated for each pulse shape and the resulting distribution is plotted in Figure 31. Figure 32 shows the pion rejection rate versus the photon efficiency that would result from different cut values on C. This is shown for the different fittypes. The neural network outperforms the multi-template fit for fittypes 0 and 1. For fittype 2 and -1, the multi-template fit cannot be used but the neural network provides good discrimination.

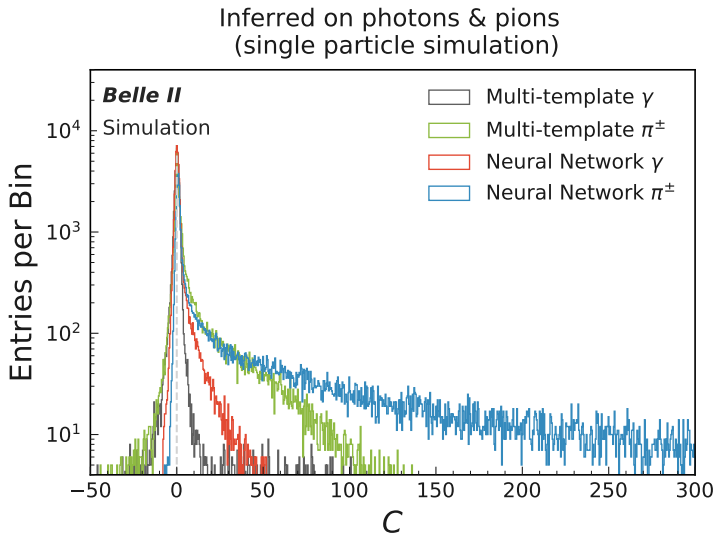


Figure 31: Normalised distribution of the classification variable C for the neural network and the multi-template-fit.

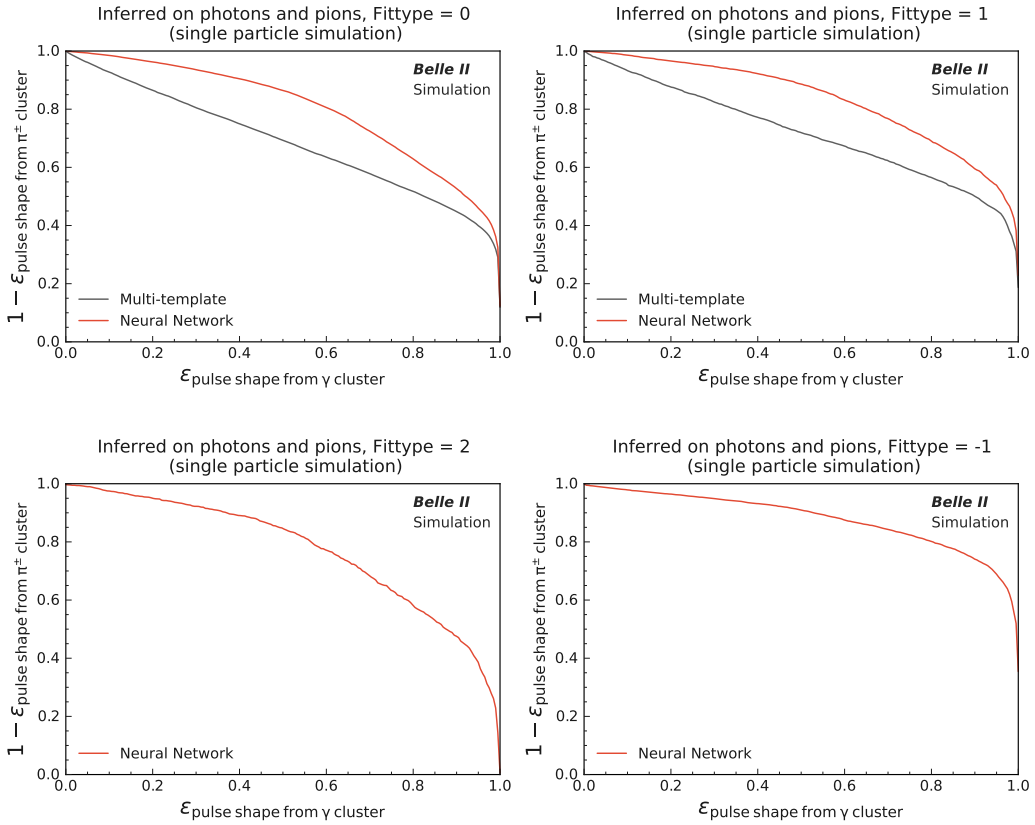


Figure 32: Pion rejection rate versus the photon efficiency for different fittypes.

6.2 Robustness of the Neural Network Approach

This chapter investigates how well the neural network performs when time shifts and unknown beam backgrounds are introduced. The figures of merit are crystal energy and hadron intensity resolution which are compared with the performance presented in 6.1.3.

6.2.1 Time Shift

A time shift, as described in 4.1.1, results in a shift of the pulse peak. If no time shift is applied during the generation of the pulse shape, the position of the maximum is in 85% of the cases at the 20^{th} position and in 14% of the cases at the 21^{st} position. For the training of the neural networks a time shift was applied to half of the pulse shapes. The position of the pulse maxima within this training set was 72% at the 20^{th} position, 17% at the 21^{st} position and 8% at the 19^{th} position. A separate set of pulse shapes was selected, excluding pulse shapes with a maximum at position 20 or 21. This set is used for testing. The peak position distribution of these two sets are shown in Figure 33.

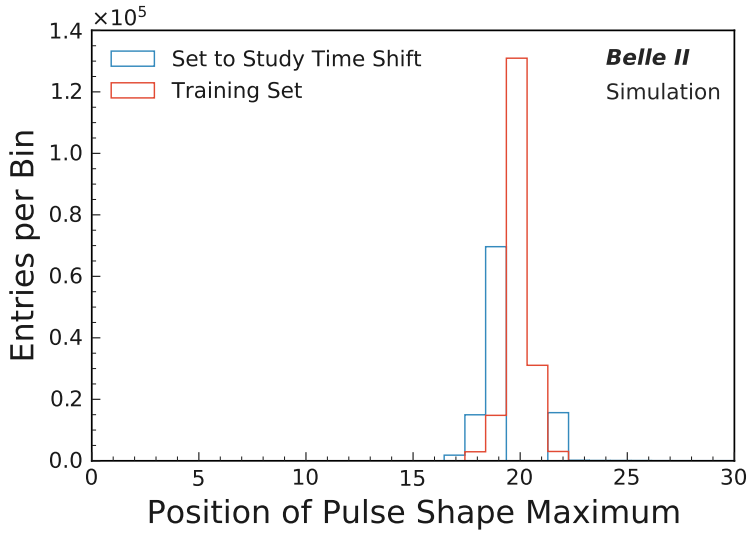


Figure 33: Position of pulse shape maximum of the training set and of the set of pulse shapes used to study the effect of shifted pulse shapes.

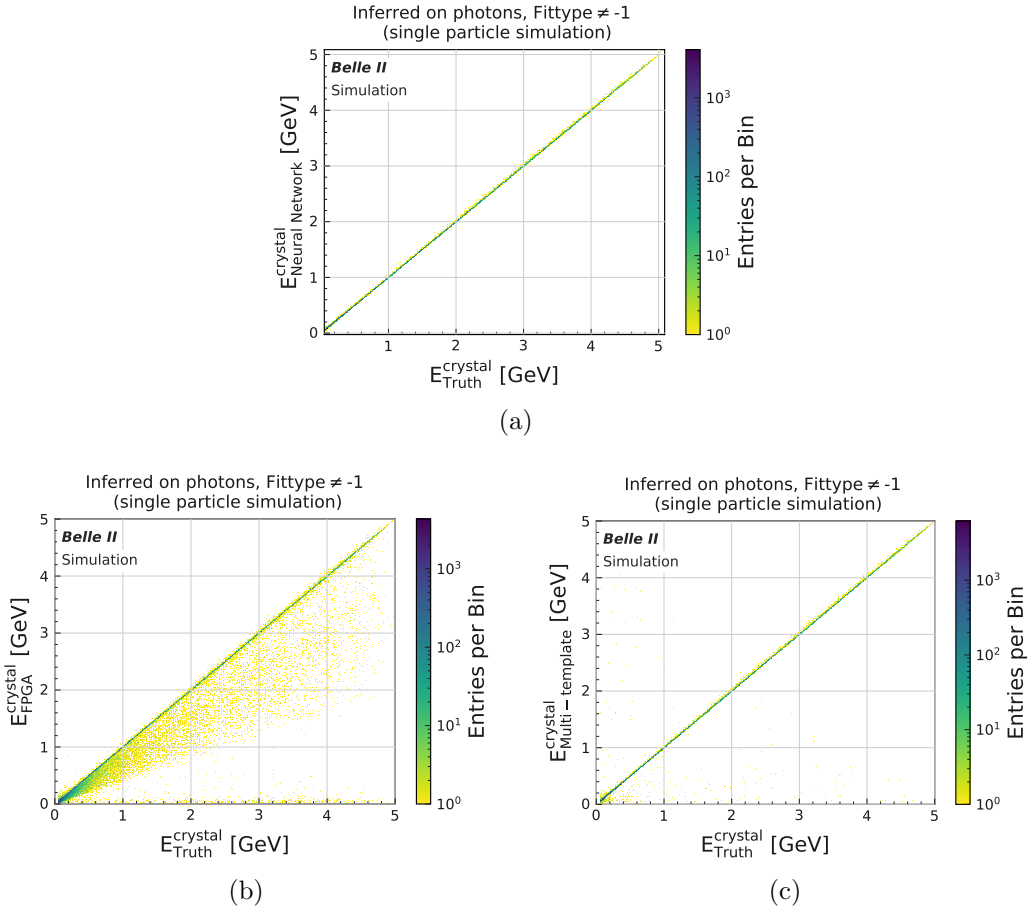
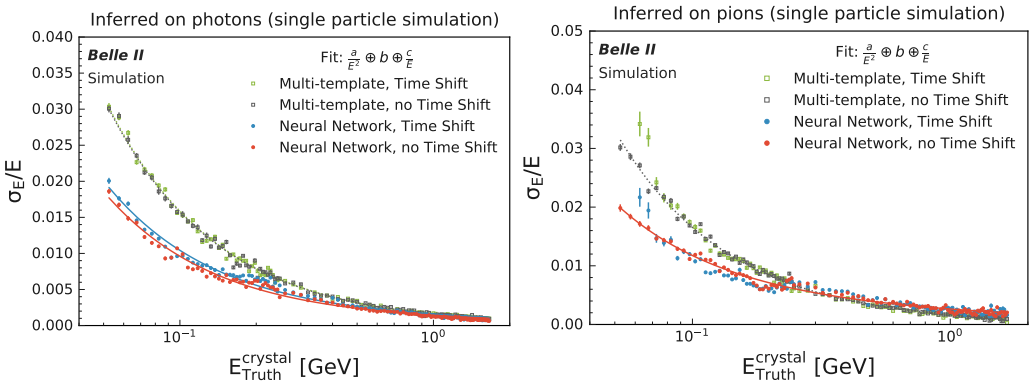


Figure 34: Predicted crystal energy by the neural network (a), the FPGA (b) and the multi-template fit (c) compared with the true crystal energy for photons, excluding fittype -1.

As already seen in chapter 6.1.1, the FPGA fit is less robust when exposed to time shifted pulse shapes. With the intentionally introduced time shifts in the training and testing sets, this is again clearly visible, as illustrated in Figure 34.

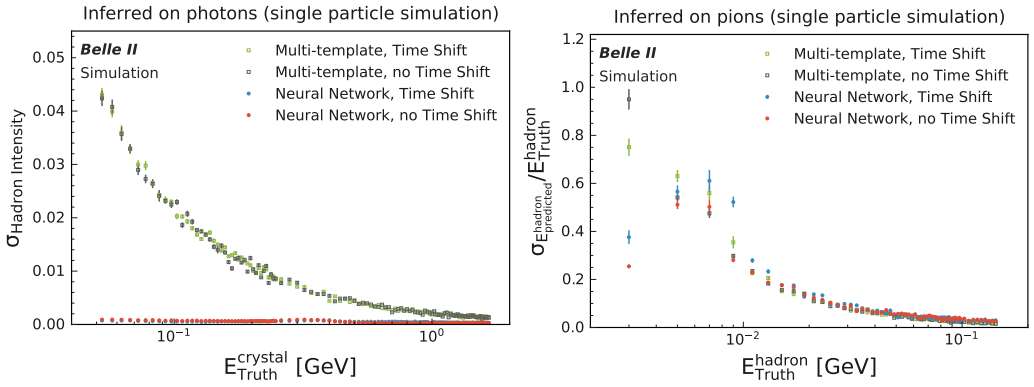
The crystal energy resolution for the multi-template fit and the neural network approach are plotted in Figure 35. For photons the resolution of the neural network is slightly degrading for shifted pulse shapes compared with no shifted pulse shapes, while the multi-template fit is not affected. For pions one observes for small true crystal energies that the multi-template fit and the neural network resolution are worse when time shifts are introduced. For

higher crystal energies both methods are barely affected by the time shift. For the hadron intensity resolution, shown in Figure 36, almost no effect for photons and a slightly worse resolution for pions is visible for the neural network. The multi-template fit is not affected in terms of photon resolution, and for pions the result in inconclusive.



(a) (b)

Figure 35: Crystal energy resolution for photons (a) and pions (b) comparing the effect of time shifted pulse shapes.

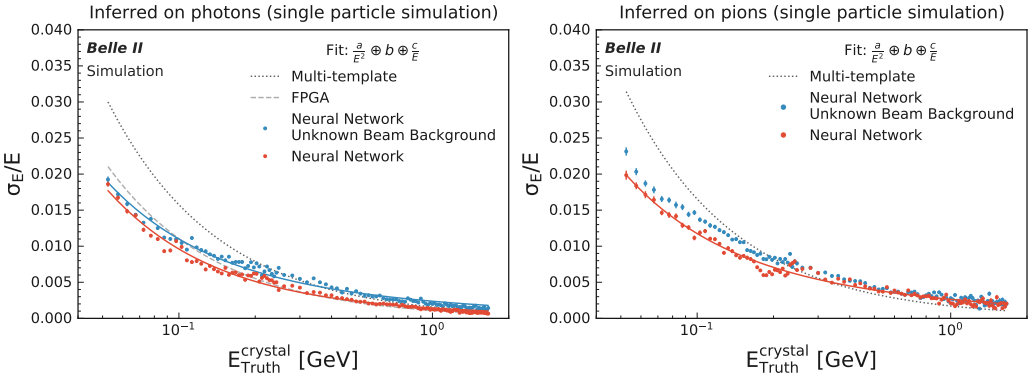


(a) (b)

Figure 36: Hadron resolution for photons (a) and pions (b) comparing the effect of time shifted pulse shapes.

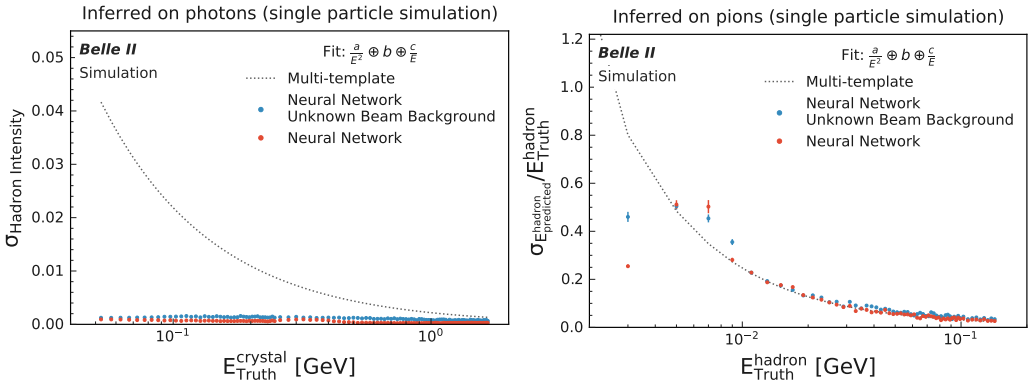
6.2.2 Beam Induced Background

In order to investigate how the neural networks perform, if faced with beam induced background overlays, which are not included in the training, a set of pulse shapes was selected which only used the background overlays from run 3363, 3402 and 4074 for the training. This way the network has not seen any pulse shapes with run-dependent beam induced background from run 5649 during training. When inferred on pulse shapes generated with this unknown background file, the energy resolution is overall worse compared with the resolution if inferred on pulse shapes with the background overlay that the neural networks were exposed to during training, as shown in Figure 37. The hadron intensity is shown in Figure 38 and also degrades for both photons and pions. This observation suggests that the neural network does not easily generalise to new beam backgrounds without retraining. It shows that a variety of pulse shapes and background overlays is important during training for an improved robustness.



(a) (b)

Figure 37: Crystal energy resolution for photons (a) and pions (b) inferred on pulse shape sets of known and unknown beam backgrounds.



(a) (b)

Figure 38: Hadron resolution for photons (a) and pions (b) inferred on pulse shape sets of known and unknown beam backgrounds

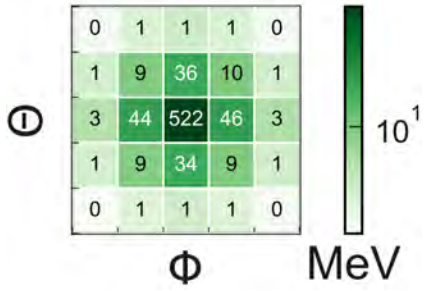


Figure 39: Example of a simulated photon cluster with a Monte Carlo true photon energy below 1 GeV, from [51].

6.3 From Crystal Energy Resolution to Cluster Energy Resolution

In chapter 6.1.3 it was shown that the neural network can improve the energy resolution on crystal level for photons with small $E_{\text{Truth}}^{\text{crystal}}$. This chapter discusses the effect of the neural network on the *cluster* energy resolution for photons. A photon cluster, as visualised in Figure 39, consists of up to 5x5 crystals. The majority of energy is deposited in the central crystal and up to a few 10 MeV in the surrounding crystals.

To study the effect of the neural network on the cluster energy resolution the Particle Gun Simulation was modified to either use the neural network prediction, the FPGA prediction or the Monte Carlo truth information for the crystal energy in order to reconstruct the cluster energy. With this modified simulation, photon events were generated with four different fixed photon momenta. The complete simulation settings are reported in appendix A.2. The reconstructed cluster energy is shown in Figure 40. The reconstruction using the FPGA and the Monte Carlo truth information agree, suggesting that the cluster resolution is not limited by the precision by which the crystal energy is known. The reconstruction using the neural network is shifted and shows a worse resolution compared with the FPGA.

The reason for this shift is that the neural network is only trained with pulse shapes from crystals above 50 MeV. When reconstructing the cluster energy the neural network is in this case inferred on crystals $E_{\text{Truth}}^{\text{crystal}} < 50\,\text{MeV}$. In Figure 41 one can see that for crystals below 50 MeV the network has

a worse performance than the FPGA. The neural network tends to predict crystal energies above the true crystal energy, which translates into the shift of the cluster energy. This is expected since the neural network was not trained with pulse shapes in this energy range. However as in the example cluster shown in Figure 39 most crystals within a cluster are below the 50 MeV threshold, suggesting why the effect of the training constraint is rather severe. Therefore it is important to include crystals below 50 MeV within the training. This should be investigated in a subsequent work.

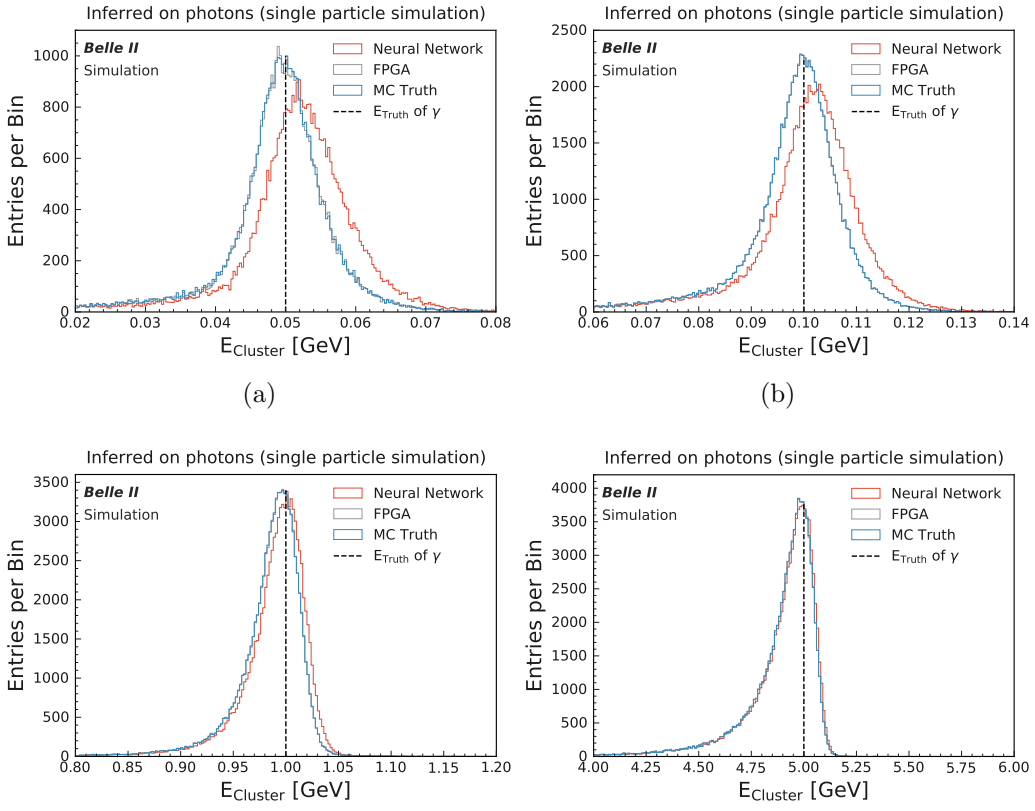
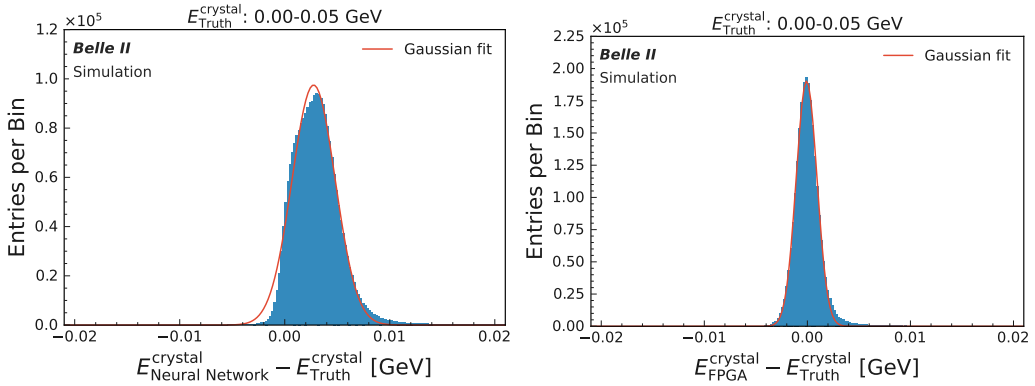


Figure 40: Reconstructed cluster energy of simulated photon events with a true photon energy of 50 MeV (a), 100 MeV (b), 1 GeV (c) and 5 GeV (d).



(a) (b)

Figure 41: Difference of crystal energy prediction and true crystal energy for photon crystals with a true crystal energy of 0 MeV to 50 MeV, as predicted by the energy network (a) and the FPGA (b).

6.4 Band Structure in Single Proton Simulation

To discriminate between different hadrons, like protons and alpha particles, one can use the hadron intensity distribution as a function of the crystal energy. In this distribution different hadron types lead to different bands [27]. As an example, single protons were simulated with Particle Gun. The full simulation settings are presented in appendix A.3. The band structure of these single-proton events is plotted in Figure 42. The true hadron intensity as a function of the true crystal energy (a) clearly shows the single proton band in the lower left. The other bands result from different secondary particles such as deuterons. For the neural network (b) and multi-template fit (c) the band structures are blurred out significantly. This is due to the limited hadron intensity resolution for hadrons which is similar for both methods, as presented in Figure 30b. Whether this can be improved by the addition of protons and other hadrons to the training sample is another possible topic for a subsequent work.

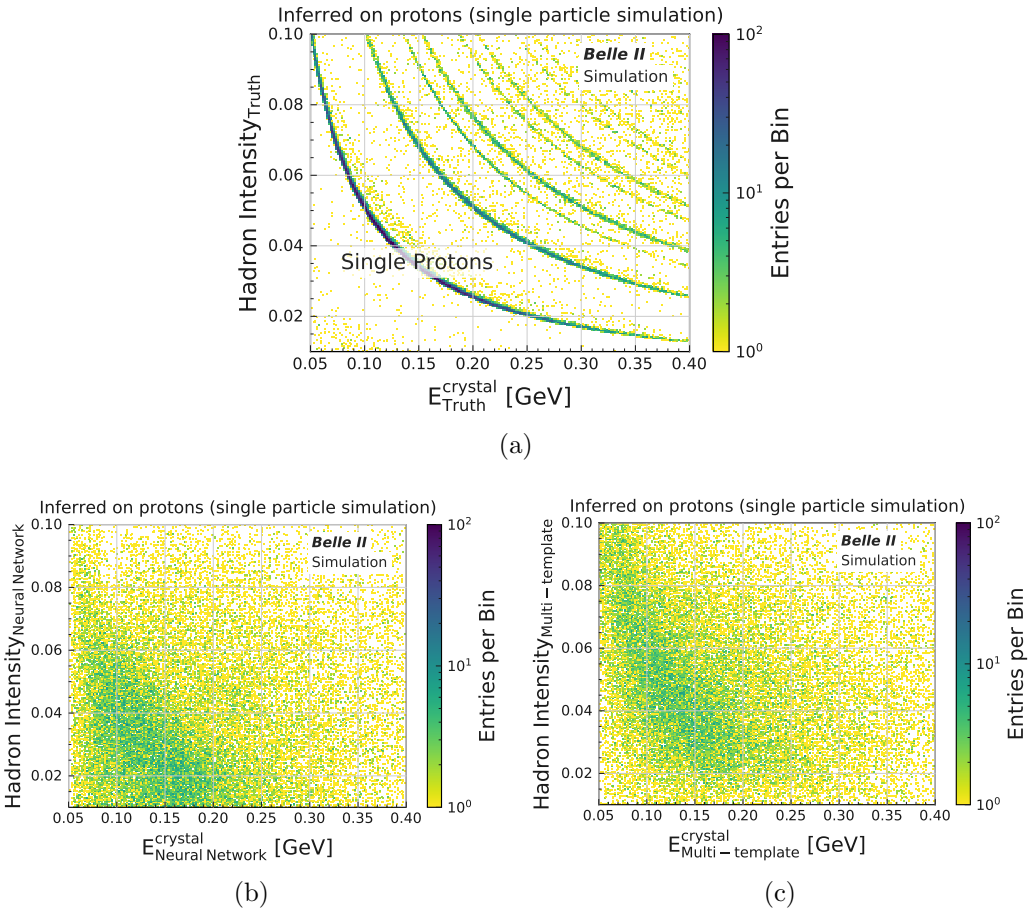


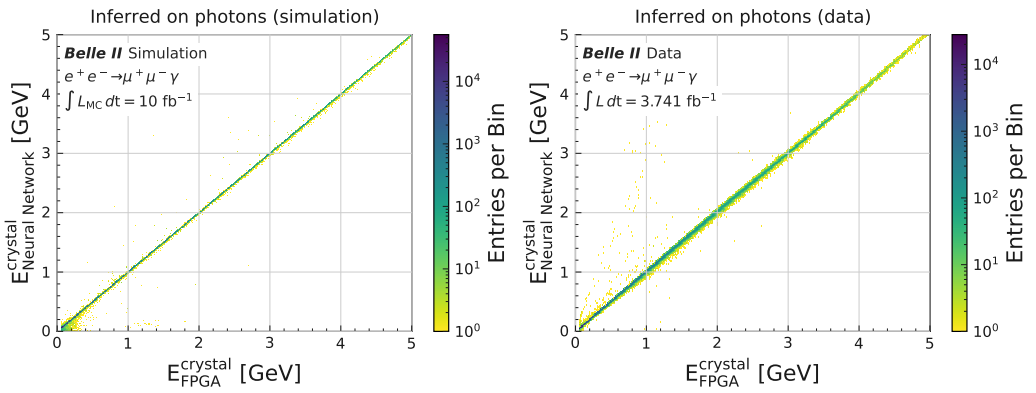
Figure 42: Band structure in the correlation of hadron intensity and crystal energy for simulated single protons using the truth values (a), the neural network response (b) and the multi-template fit (c).

6.5 Performance with Full-Event Simulation and Data

The last part of this thesis focuses on the performance of the neural networks if inferred on pulse shapes from full-event simulation and experimental data. The first section describes the performance if inferred on photons, of which the selection criteria were presented in chapter 4.3.2. The subsequent section summarises the performance regarding the charged pions, selected as described in chapter 4.3.3. In the final section the classification variable, which was introduced in section 6.1.4, is used to study the discrimination of photons and pions in full-event simulation and in experimental data.

6.5.1 Photons from $e^+e^- \to \mu^+\mu^-\gamma$

As a first step, the crystal energy prediction of the neural network and the FPGA are compared as shown in Figure 43. An overall linear relation is observed for simulation and for data. For data the line is broader and shows more outliers.

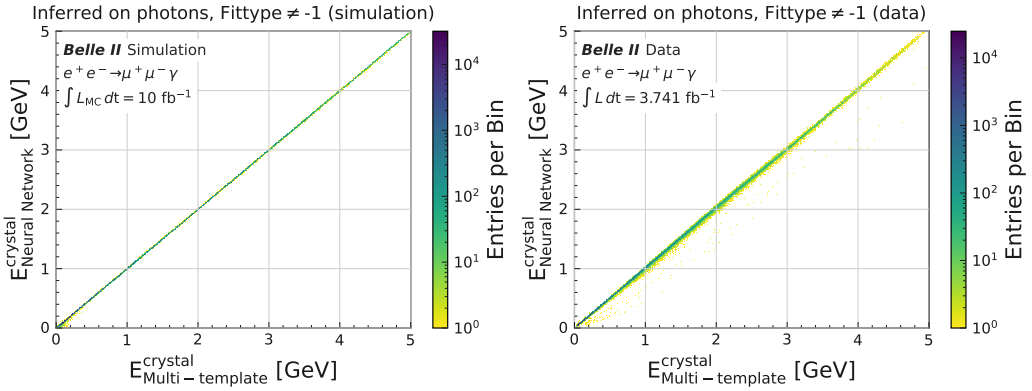


(a) (b)

Figure 43: Crystal energy prediction of the neural network versus the FPGA fit for simulation (a) and data (b).

The neural network response is further compared with the multi-template fit in Figure 44. For simulation the relation is linear and shows fewer outliers than in the comparison of the neural network with FPGA. In the case of

data, the relation is linear with a broader line. For the outliers the neural network tends to predict smaller crystal energies than the multi-template fit.



(a) (b)

Figure 44: Crystal energy prediction of the neural network versus the multi-template fit for simulation (a) and data (b).

The hadron intensity prediction of the neural network and the multi-template fit are shown in Figure 45. For both sets, the neural network predicts zero for more pulse shapes compared with the multi-template fit. In simulation, both methods peak at zero. However the multi-template fit has larger tails and reconstructs a negative hadron intensity for a larger fraction of pulse shapes than the neural network. The neural network shows a second smaller peak around the hadron intensity of 0.00126. In case of data one observes that the multi-template fit peaks slightly above zero, while the neural network peaks at zero. The multi-template fit reconstructs a negative hadron intensity for more pulse shapes than the neural network. The tails for a hadron intensity prediction of both methods above approximately 0.004 align in the logarithmic scale. The neural network often predicts a hadron intensity of 0.00433 for simulation and even more frequently for data. The reason for this is not understood, and would be a good topic of further investigation.

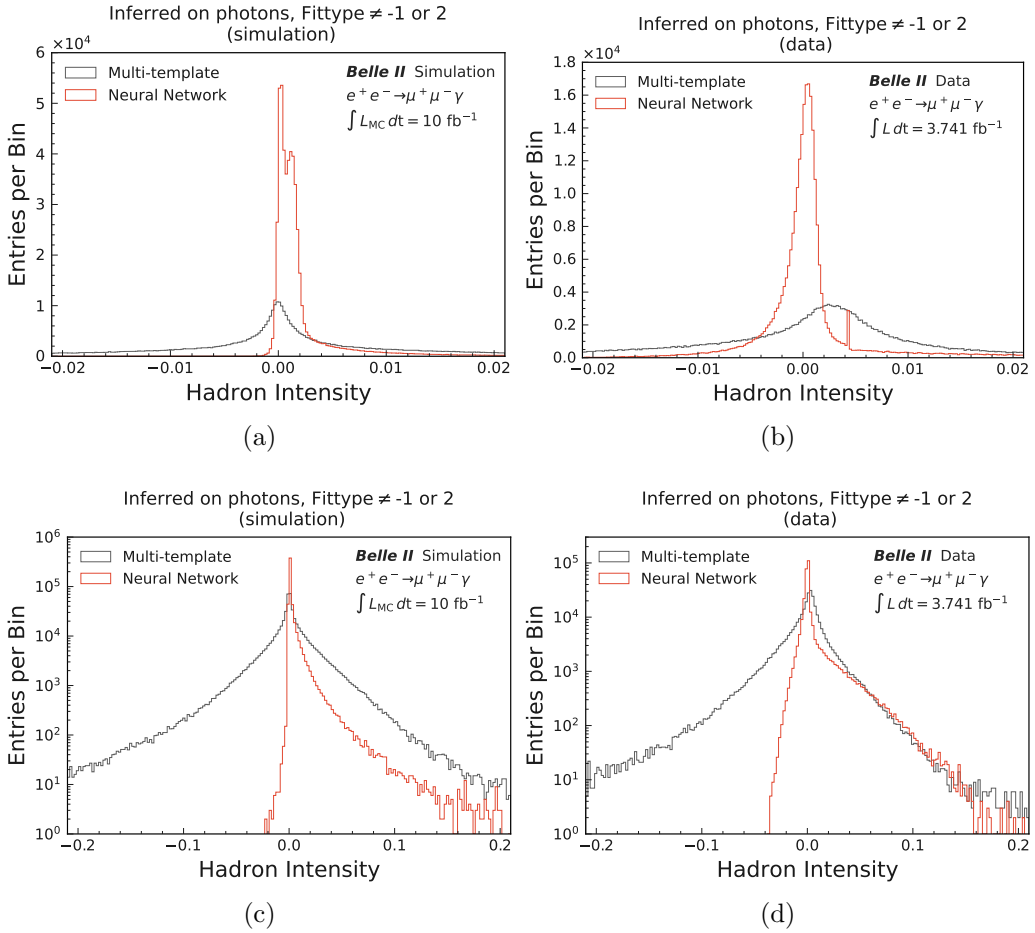
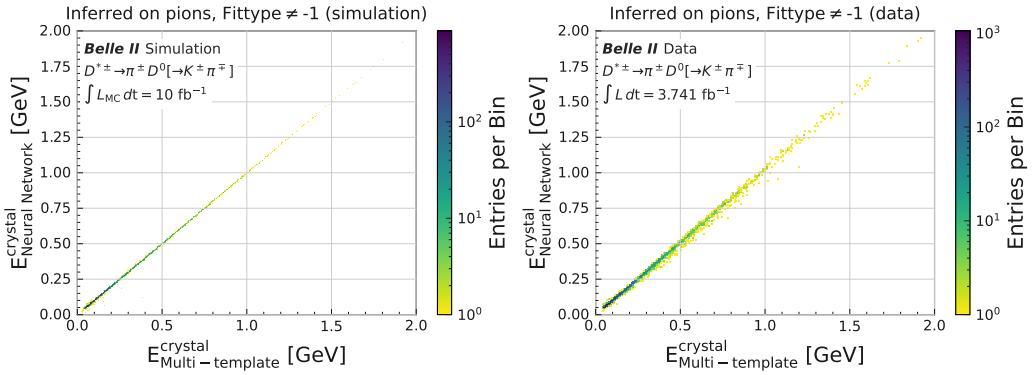


Figure 45: Hadron intensity distribution of the multi-template fit and neural network inferred on simulation (a, c) and data (b, d). Plotted on a linear (a, b) and logarithmic scale (c, d). Excluding pulse shapes of fittype -1 and 2 in all plots.

6.5.2 Charged Pions from $D^{*+} \to \pi^+ D^0[\to \pi^- K^+]$

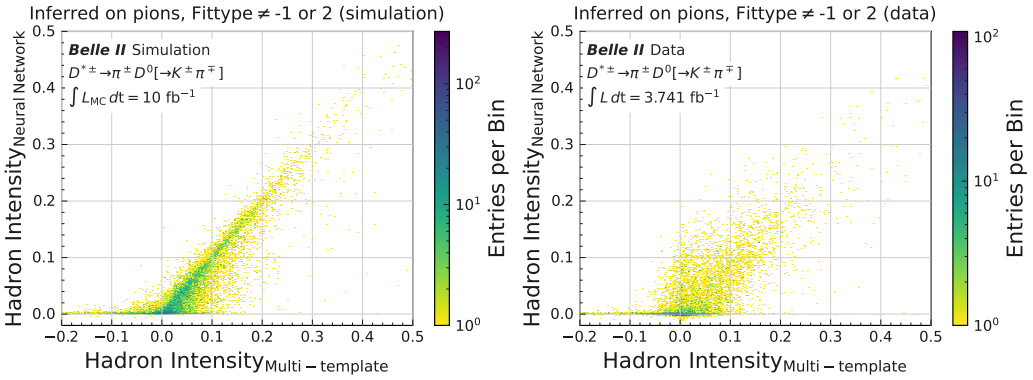
In Figure 46, the neural network crystal energy response is compared with the response of the multi-template fit for charged pions from the D^{*+} selection, which was presented in chapter 4.3.3. The relation is linear for simulation and for data, with a broader spread for data.



(a) (b)

Figure 46: Crystal energy prediction of the neural network versus the multi-template fit for simulation (a) and data (b).

Figure 47 compares the hadron intensity prediction of the neural network and the multi-template fit. For simulated pulse shapes which have a positive hadron intensity prediction from the multi-template fit, the correlation between neural network and multi-template fit is linear. Pulse shapes with negative hadron intensity reconstructed by the multi-template fit have a corresponding neural network prediction close to zero. For data a linear relation is not clearly visible. The predicted hadron intensities are further compared in Figure 48. The neural network predicts negative hadron intensities less often than the multi-template fit. Instead, the neural network predicts zero more frequently.



(a) (b)

Figure 47: Hadron intensity prediction by the neural network compared with the prediction of the multi-template fit for simulation (a) and data (b).

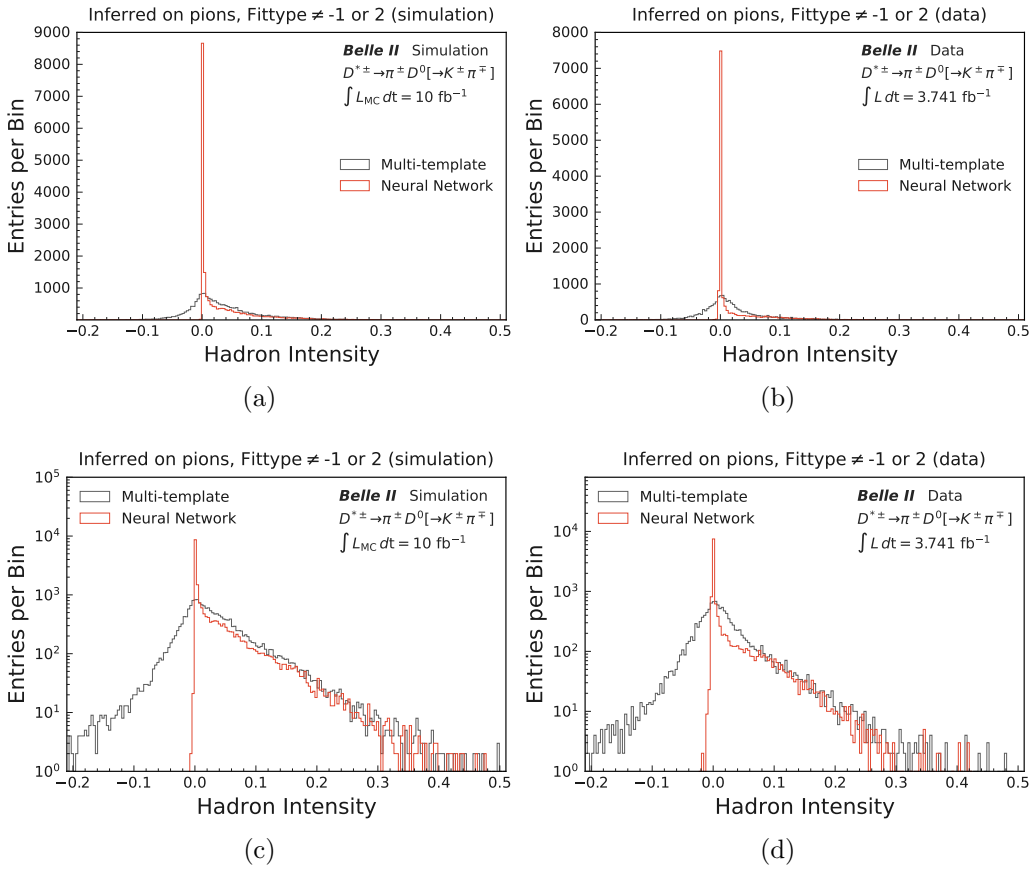
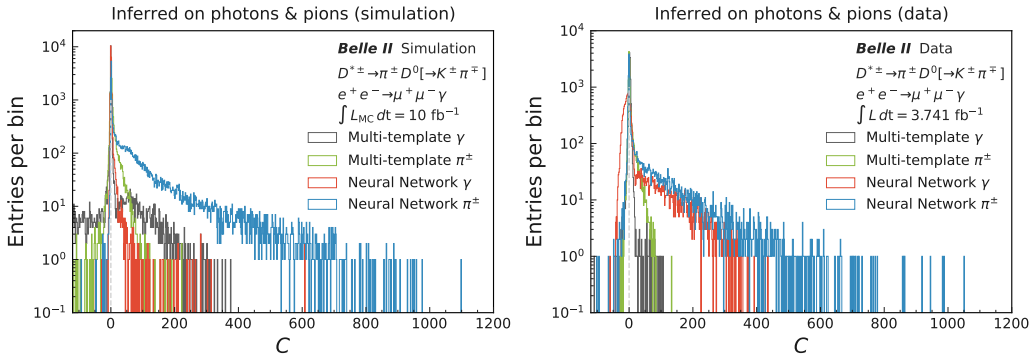


Figure 48: Hadron intensity distribution of the multi-template fit and neural network inferred on simulation (a, c) and data (b, d). Plotted on a linear (a, b) and logarithmic scale (c, d). Excluding pulse shapes of fittype -1 and 2 in all plots.

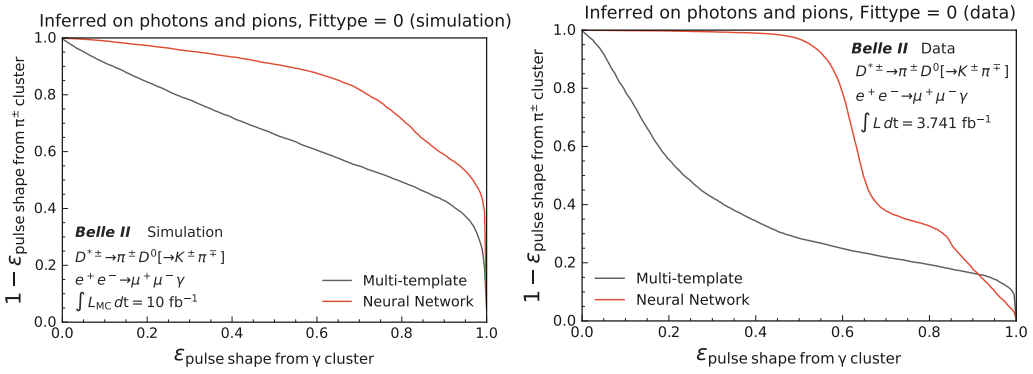
6.5.3 Separation of Photons and Pions

Using equation 6 one can calculate the classification variable C for each pulse shape. The distribution of C is shown in Figure 49. Under the assumption that the selections described in chapter 4.3 provide sets of pulse shapes from only photons or hadrons, one can further calculate the pion rejection rate versus the photon efficiency, as shown in Figure 50 - 52. Fittype 2 is not presented, since not enough pulse shapes with this fittype were available. In case of simulation the neural network has a stronger separation power than the multi-template fit for all fittypes. For data unexpected turning points are observed, especially for fittype 0. These might be a result from the difference of simulation and data, for example regarding the background. This can lead to a less precise prediction by the neural network if inferred on data, since it is trained on simulation. This could be further investigated in a subsequent work.



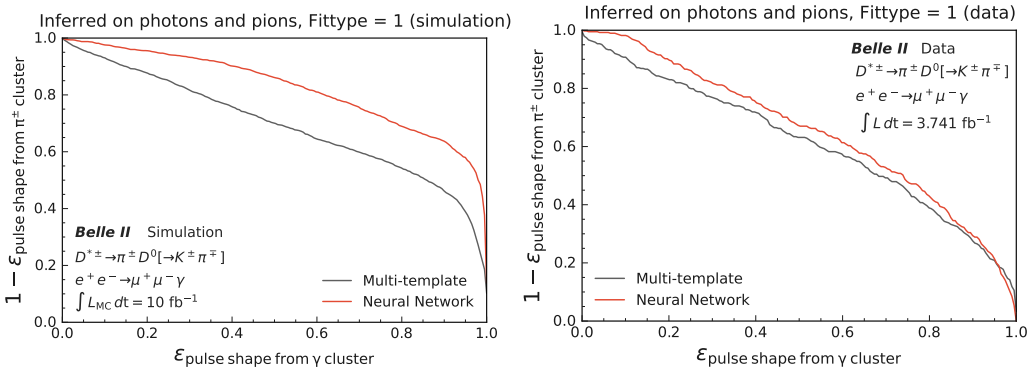
(a) (b)

Figure 49: Normalised distribution of the classification variable C for simulation (a) and data (b) for different pulse shape analysis techniques.



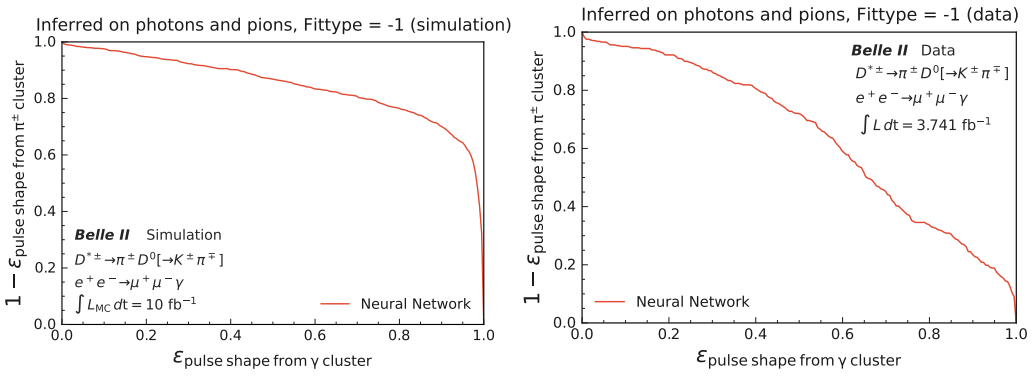
(a) (b)

Figure 50: Pion rejection rate versus the photon efficiency for fittype 0, simulation (a) and data (b).



(a) (b)

Figure 51: Pion rejection rate versus the photon efficiency for fittype 1, simulation (a) and data (b).



(a) (b)

Figure 52: Pion rejection rate versus the photon efficiency for fittype -1, simulation (a) and data (b).

7 Outlook

The work described in this thesis could be extended and further optimisation could be investigated. One important area of investigation is the composition of the training sample and its impact on the neural network performance. The neural network presented in this thesis, which was trained on 50% photons and 50% pions, showed a better crystal energy resolution and hadron intensity resolution for photons than for pions. This suggests that the prediction for pions is more challenging. To account for this, pions might need to make up a larger fraction within the training sample. Furthermore, it might be worthwhile to train with a larger variety of beam induced background overlay files and time shifts, since this increases the robustness of the neural network, as described in chapter 6.2. The robustness of the neural network might also be improved by adjusting the training parameters or architecture to allow for more generalisation. For example, one can try to implement drop out layers during the training. As discussed in chapter 6.3, the training needs to include crystals below 50 MeV to avoid the shift seen in the reconstructed cluster energy as well as the degrading of the cluster energy resolution. Moreover, other hadrons, like protons and alpha particles, could be included to the training sample, which might improve the hadron intensity resolution and might positively affect the resolution of the band structure, which was presented in chapter 6.4. While changing the composition of the training, it is important to investigate if it influences the performance of the neural networks if inferred on data and full-event simulation, where some artefacts are not understood, as discussed in chapter 6.5. To further study the neural network inferred on data and full-event simulation, more pulse shapes are needed to reach a sufficient statistical precision, for example for pion pulse shapes and rare fittypes. A different approach worth pursuing is implementing the neural network as a classifier instead of as a regression tool in which case the neural network could be trained on data.

At the time of writing, the FPGA-implementation of the neural networks is already under study. An FPGA-implementation would make a real-time

application of the neural networks possible. Similar FPGA implementations of neural networks for pulse shape analysis were found to improve the particle identification power for other experiments [43]. The architecture of the neural network might need to be adjusted in order to fit the resources and memory constraints of an FPGA. For example a single neural network, which predicts the crystal energy and hadron intensity simultaneously, could be more efficient in terms of precision, network size, and calculation speed. The optimum of network size, calculation speed, and resolution needs to be determined.

Furthermore, the neural networks can be applied in the $e^+e^- \to \pi\pi\gamma$ analysis to study if the precision of the $e^+e^- \to \pi\pi\gamma$ cross section measurement is improved. Since the neural network inference showed an improved separation power for electromagnetic and hadronic interactions compared with the multi-template fit, the neural network approach might help to improve the particle identification in the $e^+e^- \to \pi\pi\gamma$ analysis. This can lead to a reduction of the $e^+e^- \to \mu\mu\gamma$ background and reduction of the systematic uncertainty of the $e^+e^- \to \pi\pi\gamma$ cross section measurement. As discussed in chapter 2.2 and 2.3, this measurement is important for a more precise calculation of a_μ^{HVP} and a more precise calculation of the Standard Model prediction of a_μ, which could eventually reveal physics beyond the Standard Model.

8 Summary and Conclusion

The goal of this work was to investigate if neural networks are a feasible approach for pulse shape analysis for the Belle II ECL. To answer this question pulse shapes from photon and pion events were simulated and selected as the training sample for a neural network. Additional simulation events were generated and experimental events were selected as testing samples to study several aspects of the performance of the neural network, like the robustness, the effect on cluster energy reconstruction, the performance with simulated single-proton events and the inference on full-event simulation and experimental data.

Two feed forward neural networks with a simple architecture of two fully connected hidden layers with a total of 672 neurons were trained as a prototype network and achieved a good initial performance. For photons the neural network approach improves the crystal energy resolution compared with the multi-template fit and FPGA fit. The hadron intensity resolution is drastically improved by the neural network. In the case of pions with small $E_{\text{Truth}}^{\text{crystal}}$, the neural network outperforms the multi-template fit in terms of crystal energy resolution. The hadron intensity resolution is similar for both methods. The separation power of electromagnetic and hadronic interactions of the neural network outperforms the multi-template fit. One additional advantage of the neural network is that it has no restriction on the pulse shapes in terms of particle type or noise level, while the FPGA fit is restricted to photons without time shift and the multi-template fit is restricted to pulse shapes with low levels of noise (fittype 0 and 1). This advantage will become even more important in the future, since it is expected that the fraction of pulse shapes with pile-up noise (fittype -1) will increase with higher luminosity.

In addition the robustness of the neural network was studied by presenting it with shifted pulse shapes and unknown beam induced background. Here the neural network showed good but potentially improvable performance, and methods to further improve the robustness were discussed. The effect of the neural network on the reconstructed cluster energy was studied. However due

to an energy threshold in the training sample a degrading in the cluster energy resolution was observed, which can be mitigated using a lower threshold data set. Moreover, this work demonstrated that the neural network can be used for inference on full-event simulation and data, for which further optimisations in the training procedure should be considered.

In conclusion, the neural network approach shows promising improvements in terms of resolution, separation power and robustness, especially towards pulse shapes with high noise levels and diode crossing. To achieve operation readiness, next steps should be taken for example by improving the training, implementing the neural networks on an FPGA and applying the neural networks as part of the particle identification in the $e^+e^- \to \pi\pi\gamma$ analysis.

A Appendix

A.1 Resolution for each Fittype

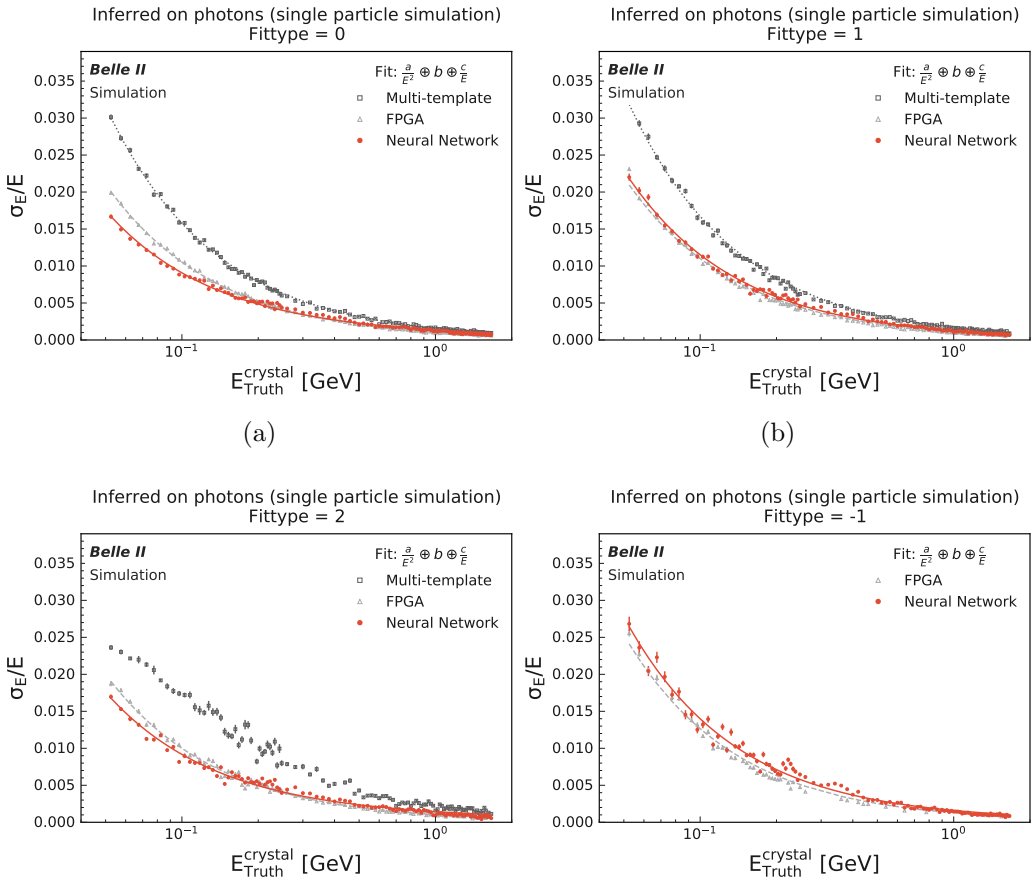


Figure 53: Crystal energy resolution for photons with fittype 0 (a), fittype 1 (b), fittype 2 (c) and fittype -1 (d).

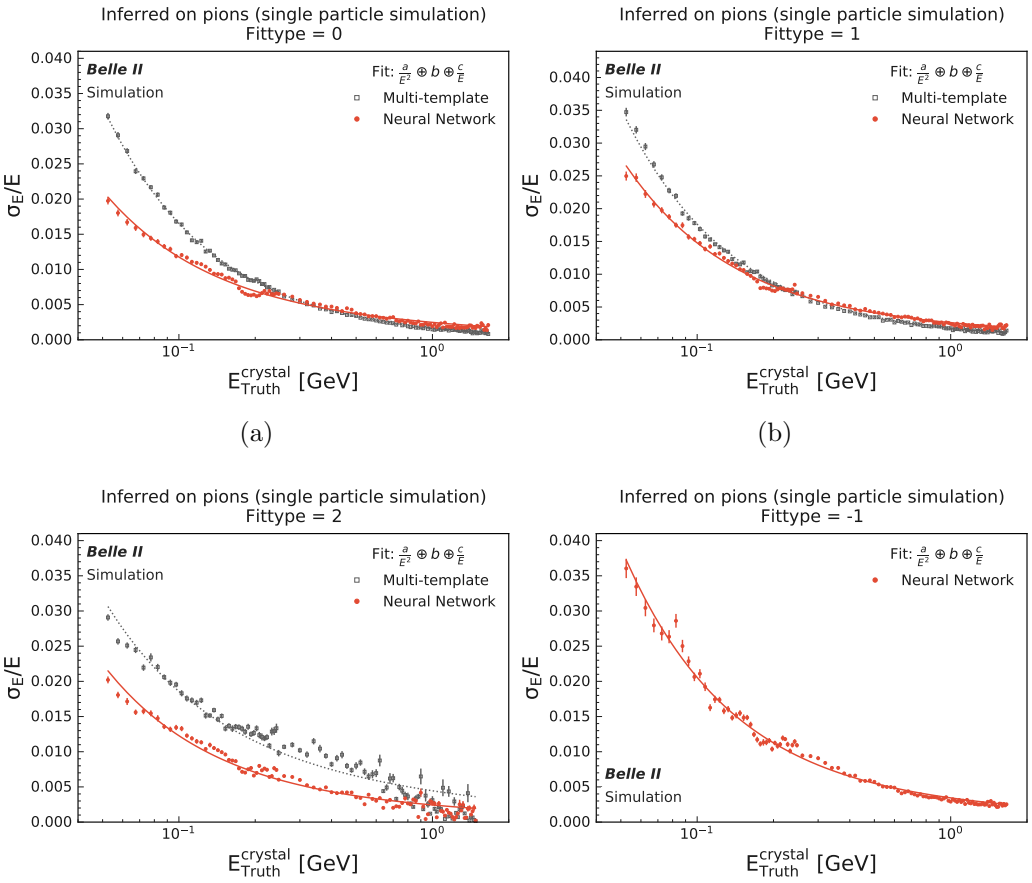


Figure 54: Crystal energy resolution for pions with fittype 0 (a), fittype 1 (b), fittype 2 (c) and fittype -1 (d).

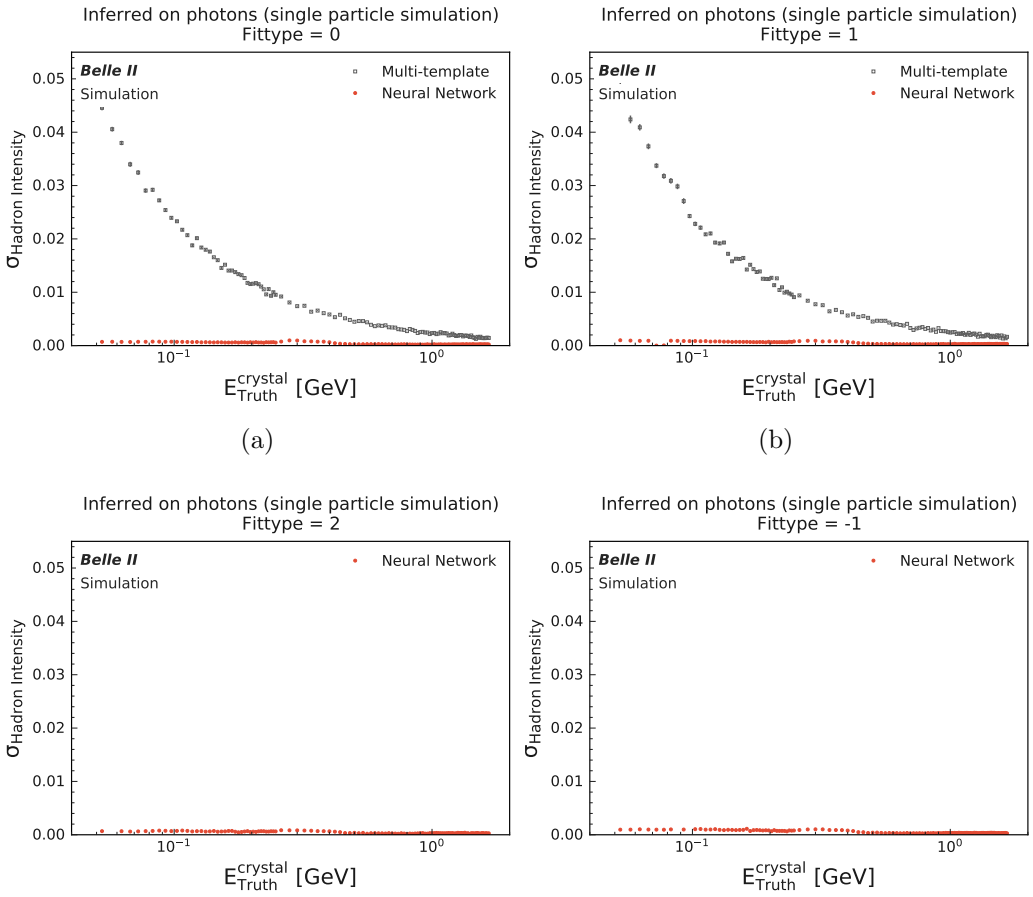


Figure 55: Hadron intensity resolution for photons with fittype 0 (a), fittype 1 (b), fittype 2 (c) and fittype -1 (d).

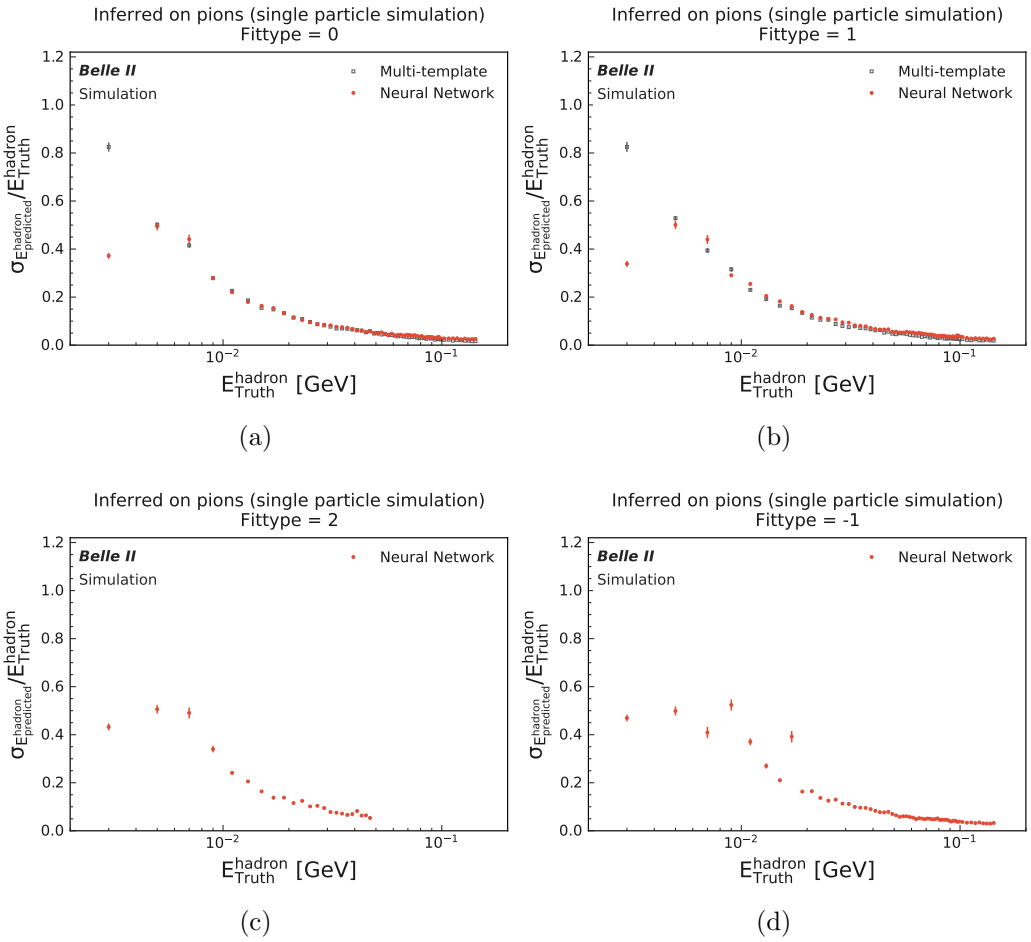


Figure 56: Hadron energy resolution for pions with fittype 0 (a), fittype 1 (b), fittype 2 (c) and fittype -1 (d).

A.2 Simulation Settings and Selection to study Cluster Energy

A set of 100000 photon events were generated with the following simulation settings:

- Fixed momentum of 0.05 GeV/c, 0.1 GeV/c, 1 GeV/c or 5 GeV/c
- Polar angle θ fixed to 110° (within the barrel)
- Azimuthal angle ϕ fixed to 200° (within the barrel)

- Beam induced background overlay from experiment 12 run 3363, 3402, 4074 or 5649
- No random time shift added

For the further selection Monte Carlo Truth Matching was applied, without any other selection criteria.

A.3 Simulation Settings and Selection for Single Protons

A set of 1000000 proton events were generated with the following simulation settings:

- Momentum generated uniformly in the range of $0.05\,\mathrm{GeV/c} - 6\,\mathrm{GeV/c}$
- Polar angle θ generated uniformly in the range of $30^\circ - 125^\circ$
- Azimuthal angle ϕ generated uniformly in the range of $0^\circ - 360^\circ$
- Add random time shift, which is uniformly distributed in the range of $\pm 2000\,\mathrm{ns}$
- Beam induced background overlay from experiment 12 run 3363, 3402, 4074 or 5649

The following selection criteria were applied:

- $E_{\mathrm{Truth}}^{\mathrm{crystal}} > 50\,\mathrm{MeV}$
- Crystal within the barrel (ECL Cell ID: 860 to 8022) resulting in 191946 pulse shapes from single proton events in total.

References

[1] B. Abi et al. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, Apr 2021. doi: 10.1103/PhysRevLett.126.141801. URL https://link.aps.org/doi/10.1103/PhysRevLett.126.141801.

[2] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[3] T. Aoyama, N. Asmussen, M. Benayoun, J. Bijnens, T. Blum, M. Bruno, I. Caprini, C.M. Carloni Calame, M. Cè, G. Colangelo, and et al. The anomalous magnetic moment of the muon in the Standard Model. *Physics Reports*, 887:1–166, Dec 2020. ISSN 0370-1573. doi: 10.1016/j.physrep.2020.07.006. URL http://dx.doi.org/10.1016/j.physrep.2020.07.006.

[4] F. Jegerlehner. *The Anomalous Magnetic Moment of the Muon (Second edition)*. Springer Spektrum, 2017. ISBN 978-3-319-63577-4. doi: 10.1007/978-3-319-63577-4.

[5] Alexander Keshavarzi, Daisuke Nomura, and Thomas Teubner. g-2 of charged leptons, $\alpha(m_Z^2)$, and the hyperfine splitting of muonium. *Physical Review D*, 101(1), Jan 2020. ISSN 2470-0029. doi: 10.1103/physrevd.101.014029. URL http://dx.doi.org/10.1103/PhysRevD.101.014029.

[6] A. Hoecker and W.J. Marciano. The muon anomalous magnetic moment. Aug 2017. URL https://pdg.lbl.gov/2017/reviews/rpp2017-rev-g-2-muon-anom-mag-moment.pdf.

[7] Y. Maeda. Plots for ee → pi+pi-gamma analysis with phase2 full data. Sep 2018. URL https://docs.belle2.org/record/1124?ln=en.

[8] S. Chatrchyan and et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):

30–61, Sep 2012. ISSN 0370-2693. doi: 10.1016/j.physletb.2012.08.021. URL http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[9] G. Aad and et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, Sep 2012. ISSN 0370-2693. doi: 10.1016/j.physletb.2012.08.020. URL http://dx.doi.org/10.1016/j.physletb.2012.08.020.

[10] Wikipedia: Standard Model. https://en.wikipedia.org/wiki/Standard_Model. Accessed: 2021-04-08.

[11] S. Borsanyi, Z. Fodor, J.N. Guenther, and et al. Leading hadronic contribution to the muon magnetic moment from lattice QCD. *Nature*, April 2021. doi: 10.1038/s41586-021-03418-1. URL https://doi.org/10.1038/s41586-021-03418-1.

[12] R.R Akhmetshin et al. Update: A reanalysis of hadronic cross section measurements at CMD-2. *Physics Letters B*, 578(3):285–289, 2004. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2003.10.108. URL https://www.sciencedirect.com/science/article/pii/S0370269303017179.

[13] M. N. Achasov et al. Update of the $e^+e^- \to \pi^+\pi^-$ cross section measured by the spherical neutral detector in the energy region $400 < \sqrt{s} < 1000$ mev. *Journal of Experimental and Theoretical Physics*, 103(3):380–384, Sep 2006. ISSN 1090-6509. doi: 10.1134/s106377610609007x. URL http://dx.doi.org/10.1134/S106377610609007X.

[14] F. Ambrosino et al. Measurement of $\sigma(e^+e^- \to \pi^+\pi^-\gamma(\gamma))$ and the dipion contribution to the muon anomaly with the KLOE detector. *Physics Letters B*, 670(4):285–291, 2009. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2008.10.060. URL https://www.sciencedirect.com/science/article/pii/S0370269308013269.

[15] A. Anastasi et al. Combination of KLOE $\sigma\left(e^+e^- \to \pi^+\pi^-\gamma(\gamma)\right)$ measurements and determination of $a_\mu^{\pi^+\pi^-}$ in the energy range $0.10 < s < 0.95$ GeV^2. 2018.

[16] B. Aubert et al. Precise Measurement of the $e^+e^- \to \pi^+\pi^-(\gamma)$ Cross Section with the Initial State Radiation Method at BABAR. *Phys. Rev. Lett.*, 103:231801, Dec 2009. doi: 10.1103/PhysRevLett.103.231801. URL https://link.aps.org/doi/10.1103/PhysRevLett.103.231801.

[17] J. P. Lees et al. Precise measurement of the $e^+e^- \to \pi^+\pi^-(\gamma)$ cross section with the initial-state radiation method at BABAR. *Phys. Rev. D*, 86:032013, Aug 2012. doi: 10.1103/PhysRevD.86.032013. URL https://link.aps.org/doi/10.1103/PhysRevD.86.032013.

[18] M. Ablikim et al. Measurement of the $e^+e^- \to \pi^+\pi^-$ cross section between 600 and 900 MeV using initial state radiation. *Physics Letters B*, 753:629–638, 2016. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2015.11.043. URL https://www.sciencedirect.com/science/article/pii/S0370269315008990.

[19] T. Xiao et al. Precision measurement of the hadronic contribution to the muon anomalous magnetic moment. *Phys. Rev. D*, 97:032012, Feb 2018. doi: 10.1103/PhysRevD.97.032012. URL https://link.aps.org/doi/10.1103/PhysRevD.97.032012.

[20] Belle II Collaboration. Belle II Archives. https://www.belle2.org/archives/. Accessed: 2021-03-09.

[21] I. Adachi, T.E. Browder, P. Križan, S. Tanaka, and Y. Ushiroda. Detectors for extreme luminosity: Belle II. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 907:46–59, 2018. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2018.03.068. URL https://www.

sciencedirect.com/science/article/pii/S0168900218304200. Advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn).

[22] E Kou, P Urquijo, W Altmannshofer, F Beaujean, G Bell, M Beneke, I I Bigi, F Bishara, M Blanke, C Bobeth, et al. The Belle II Physics Book. *Progress of Theoretical and Experimental Physics*, 2019(12), Dec 2019. ISSN 2050-3911. doi: 10.1093/ptep/ptz106. URL http://dx.doi.org/10.1093/ptep/ptz106.

[23] A. J. Bevan et al. The Physics of the B Factories. *The European Physical Journal C*, 74(11), Nov 2014. ISSN 1434-6052. doi: 10.1140/epjc/s10052-014-3026-9. URL http://dx.doi.org/10.1140/epjc/s10052-014-3026-9.

[24] H. Kolanoski and N. Wermes. *Particle Detectors: Fundamentals and Applications*. Oxford University Press, 2020. ISBN 978-0198858362. doi: 10.1093/oso/9780198858362.001.0001.

[25] S. Longo. *First application of CsI(Tl) pulse shape discrimination at an e^+e^- collider to improve particle identification at the Belle II experiment*. PhD thesis, University of Victoria, Oct 2019.

[26] R S Storey, W Jack, and A Ward. The Fluorescent Decay of CsI(Tl) for Particles of Different Ionization Density. *Proceedings of the Physical Society*, 72(1):1–8, jul 1958. doi: 10.1088/0370-1328/72/1/302. URL https://doi.org/10.1088/0370-1328/72/1/302.

[27] C.M. Bartle and R.C. Haight. Small inorganic scintillators as neutron detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 422(1):54–58, 1999. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(98)01062-6. URL https://www.sciencedirect.com/science/article/pii/S0168900298010626.

[28] S. Longo et al. CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron–positron colliders. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 982:164562, Dec 2020. ISSN 0168-9002. doi: 10.1016/j.nima.2020.164562. URL http://dx.doi.org/10.1016/j.nima.2020.164562.

[29] A. Sibidanov. Belle II readiness for Phase II collisions. Feb 2018. URL https://docs.belle2.org/record/800?ln=en.

[30] Geant4: a simulation toolkit. https://geant4.web.cern.ch/. Accessed: 2021-04-30.

[31] S. Longo and J. M. Roney. Hadronic vs. electromagnetic pulse shape discrimination in CsI(Tl) for high energy physics experiments. *Journal of Instrumentation*, 13, 2018.

[32] P.M. Lewis et al. First measurements of beam backgrounds at SuperKEKB. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 914:69–144, Jan 2019. ISSN 0168-9002. doi: 10.1016/j.nima.2018.05.071. URL http://dx.doi.org/10.1016/j.nima.2018.05.071.

[33] Doris Yangsoo Kim. Software and Physics Simulation at Belle II. Oct 2015. URL https://docs.belle2.org/record/533/export/hx?ln=en.

[34] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun. The Belle II Core Software. *Computing and Software for Big Science*, 3 (1), Nov 2018. ISSN 2510-2044. doi: 10.1007/s41781-018-0017-9. URL http://dx.doi.org/10.1007/s41781-018-0017-9.

[35] R. Dhamija, S. Sandilya, and A. Giri. Study of Lepton ID Fake Rate using D^* sample. July 2020. URL https://docs.belle2.org/record/1953/files/BELLE2-NOTE-PH-2020-034.pdf.

[36] Charu C. Aggarwal. *An Introduction to Neural Networks*, pages 1–52. Springer International Publishing, Cham, 2018. ISBN 978-3-319-94463-0. doi: 10.1007/978-3-319-94463-0_1. URL https://doi.org/10.1007/978-3-319-94463-0_1.

[37] Cristopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006. ISBN 978-0387-31073-2.

[38] Durán, J. Everything You Need to Know about Gradient Descent Applied to Neural Networks. https://medium.com/yottabytes/everything-you-need-to-know-about-gradient-descent-applied-to-neural-networks-d70f85e0cc14, September 2019.

[39] Krishnan, B. WHEN and WHY are batches used in machine learning? https://medium.com/analytics-vidhya/when-and-why-are-batches-used-in-machine-learning-acda4eb00763, November 2019.

[40] Pradeepta Mishra. *PyTorch Recipes: A Problem-Solution Approach*. Apress, USA, 1st edition, 2019. ISBN 1484242572.

[41] Andreas Josef Zöeller. *Artificial Neural Network Based Pulse-Shape Analysis for Cryogenic Detectors Operated in CRESST-II*. PhD thesis, Technische Universität München, July 2016.

[42] Sebastian Kupny. *Digital Pulse Shape Analysis with Neural Networks: Application to Flow Measurements in the ASY-EOS Experiment*. PhD thesis, Jagiellonian University, March 2016.

[43] J.L. Flores, I. Martel, R. Jiménez, J. Galán, and P. Salmerón. Application of neural networks to digital pulse shape analysis for an array of silicon strip detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 830:287–293, 2016. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2016.05.107. URL https://www.sciencedirect.com/science/article/pii/S0168900216305095.

[44] P. Cennini et al. A neural network approach for the TPC signal processing. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 356(2):507–513, 1995. ISSN 0168-9002. doi: https://doi.org/10.1016/0168-9002(94)01310-1. URL https://www.sciencedirect.com/science/article/pii/0168900294013101.

[45] NEOS II Collaboration. Pulse Shape Discrimination of Fast Neutron Background using Convolutional Neural Network for NEOS II. 2020. URL https://arxiv.org/abs/2009.13355.

[46] PyTorch. https://pytorch.org/, . Accessed: 2021-04-30.

[47] PyTorch ReLU. https://pytorch.org/docs/stable/generated/torch.nn.ReLU.html?highlight=relu#torch.nn.ReLU, . Accessed: 2021-04-30.

[48] PyTorch MSELoss. https://pytorch.org/docs/stable/generated/torch.nn.MSELoss.html?highlight=mse, . Accessed: 2021-04-30.

[49] PyTorch Optim Adam. https://pytorch.org/docs/stable/optim.html?highlight=adam#torch.optim.Adam, . Accessed: 2021-04-30.

[50] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, 2017. URL https://arxiv.org/abs/1412.6980.

[51] Cedric Ly. Shower shape conversion between electrons and photons in Belle II using Cycle Generative Adversarial Network. Master's thesis, Hamburg, University of hamburg, Hamburg, 2021. URL https://docs.belle2.org/record/2347.

Eidesstattliche Erklärung

Ich versichere, dass ich die beigefügte schriftliche Masterarbeit selbstständig angefertigt und keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle Stellen, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, habe ich in jedem einzelnen Fall unter genauer Angabe der Quelle deutlich als Entlehnung kenntlich gemacht. Dies gilt auch für alle Informationen, die dem Internet oder anderer elektronischer Datensammlungen entnommen wurden. Ich erkläre ferner, dass die von mir angefertigte Masterarbeit in gleicher oder ähnlicher Fassung noch nicht Bestandteil einer Studien- oder Prüfungsleistung im Rahmen meines Studiums war. Die von mir eingereichte schriftliche Fassung entspricht jener auf dem elektronischen Speichermedium.

Ich bin damit einverstanden, dass die Masterarbeit veröffentlicht wird.

Hamburg, den 13.06.2021

Unterschrift