Event classification for a Compton/Pair Telescope and simulation studies of the COMCUBE CubeSat project for gamma-ray Astronomy

Dissertation zur Erlangung des Grades
Doktor der Naturwissenschaften
vorgelegt dem
Fachbereich Physik, Mathematik und Informatik (FB 08)
der Johannes Gutenberg-Universität Mainz

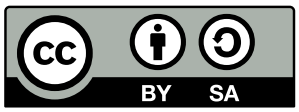
Jan Peter Lommler

geb. in Bingen am Rhein

Mainz, den 15.12.2025

1. Berichterstatter:
2. Berichterstatter:

Datum der mündlichen Prüfung: 26.03.2026
D77: Dissertation Johannes Gutenberg-Universität



Hiermit erkläre ich,

dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel (dazu zählen auch KI-basierte Anwendungen oder Werkzeuge) benutzt habe. Sämtliche wörtliche oder sinngemäße Übernahmen sind kenntlich gemacht und nachgewiesen.
Die vorgelegte Arbeit wurde nicht als Prüfungsarbeit oder Teil einer Prüfungsarbeit für eine andere Prüfung oder Dissertation an einer anderen Fakultät oder einem anderen Fachbereich eingereicht. Genutzte KI-Tools sind im Anhang *Nutzung von KI-Werkzeugen* dokumentiert. Mit Abgabe der vorliegenden Leistung übernehme ich die Verantwortung für das eingereichte Gesamtdokument und verantworte damit auch jegliche KI-generierte Inhalte in dieser Arbeit. Die Richtigkeit der übernommenen Aussagen und Inhalte wurden nach bestem Wissen und Gewissen überprüft.
Mir ist bekannt, dass ein Verstoß gegen die genannten Punkte prüfungsrechtliche Konsequenzen hat.

Mainz, den 15.12.2025,

Jan Peter Lommler
ETAP
Institut für Physik
Staudingerweg 7
Johannes Gutenberg-Universität D-55099 Mainz
lommler@uni-mainz.de

Abstract

Cosmic gamma-rays are our main observational window into the universe's most violent processes. The observable spectrum ranges from 100 keV to several PeV. Their observation gives insight into a wide variety of astrophysical processes. Cosmic nucleosynthesis in massive stars leaves their imprint in line energies. Their measurement advances our understanding of their evolution and the distribution of heavy elements in our galaxy. Cosmic acceleration sites leave their imprint in continuous emissions, giving insight into the acceleration processes as well as the composition of the accelerated particles. Photons emitted during Gamma Ray Bursts (GRBs) are the messengers of the most extreme events in the universe. Their gamma-rays allow us to investigate the source conditions of these supernovae and merger events. The range between 0.1 to 30 MeV constitutes the 'MeV-Gap'. No mission after the Compton Gamma Ray Observatory provided improved sensitivity in this range of the spectrum. Compton/Pair telescopes can cover this range and provide additional sensitivity up to several GeV in one instrument using a segmented silicon tracking detector on top of a solid state calorimeter or liquid noble gas Time Projection Chamber. This work is divided into two main parts. The first part focuses on presenting a novel deep learning driven classification algorithm capable of providing online classification of events directly at the satellite. The task description, evaluation methods, alternative classification approaches and results are presented and discussed. The second part describes the design of the COMCUBE proposal, a CubeSat GRB polarimeter developed for the EU AHEAD 2020 initiative. CubeSats provide a potential bridge to cover the lower end of the MeV-Gap using small and cheaper mini-satellites. This work discusses performance estimates for the initial 4U detector design proposal, as well as adaptations done for the first qualification flight of the prototype during the Strato Sat 2023 balloon campaign in Timmins (Ontario). Further trigger rates for the expected radiative environment and sensitivity estimates for the Crab-Nebula are presented. The measured trigger rates are compared to simulated data.

Zusammenfassung

Kosmische Gamma-Strahlung ist unser Hauptfenster zu den energiereichsten Ereignissen unseres Universums. Das beobachtbare Energiespektrum liegt zwischen 100 keV und einigen PeV. Ihre Beobachtung gibt Einblick in eine Vielzahl von astrophysikalischen Prozessen. Linienemissionen geben Aufschluss über die Entstehung und Verteilung schwerer Elemente, die durch Nukleosynthese in massereichen Sternen entstehen und erlauben Rückschlüsse über deren Entwicklung. Kosmische Beschleuniger hinterlassen ihre Spuren in kontinuierlichen Gamma-Emissionen und geben Einblick über Art der beschleunigten Teilchen und Stärke verschiedener Beschleunigungsprozesse.
Photonen aus Gammastrahlenblitzen (GRBs) sind die Hauptinformationsträger der gewaltigsten beobachtbaren Ereignisse. Sie erlauben Rückschlüsse auf die Bedingungen an der Quelle von unter anderem Kern-Kollaps Supernovae extrem massereicher Sterne und Neutronenstern-Kollisionen. Der Bereich zwischen 0.1 MeV und 30 MeV wird auch als „MeV-Lücke" bezeichnet. Seit dem Compton Gamma Ray Observatory hat keine Mission in diesem Energiebereich zu einer Verbesserung der Sensitivität in diesem Teil des Spektrums beigetragen. Compton/Paar Teleskope können den gesamten Bereich zwischen 0.1 MeV und einigen GeV abdecken. Ihr Design besteht üblicherweise aus einem Silizium-Spurdetektor, der auf einem Festkörper-Kalorimeter oder einer Flüssig-Edelgas Zeitprojektionskammer montiert ist. Diese Arbeit besteht aus zwei Hauptteilen. Im ersten Hauptteil wird ein neuartiger auf mehrschichtigem, tiefen machinelem Lernen basierender Klassifizierungsalgorithmus vorgestellt. Dieser Algorithmus kann Ereignisse *online* am Satelliten klassifizeren. Es werden die Klassifizierungsaufgabe, verwendete Datensätze, Evaluierungsmethoden, Modellarchitektur und alternative Ansätze diskutiert und bewertet. Der zweite Teil beschreibt die Entwicklung und Erprobung des COMCUBE CubeSat Designs im Rahmen der AHEAD 2020 Initiative. CubeSats sind eine interessante Brückentechnologie zur Abdeckung des niederenergetischen Teils der MeV-Lücke. COMCUBE ist ein GRB-Monitor, der auf Basis von Comptonstreuung zeitgleich die Polarisierung der einfallenden Photonen messen kann. Diese Arbeit präsentiert und diskutiert ermittelte Leistungsmaßstäbe des 4U Entwurfs sowie Modifikationen und Anpassungen des Prototyps für den Ballonflug während der Strato Sat 2023 Kampagne. Desweiteren werden Triggerraten und die Sensitivität für den Krebsnebel für die erwarteten Umweltbedingungen durch Simulationen ermittelt. Die vorhergesagten Raten werden mit gemessenen Daten verglichen.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The electromagnetic spectrum is our main observational window into the Universe, ranging from radio waves to highly energetic gamma-ray photons. Gamma-ray astronomy is the observation of high energy cosmic gamma-rays, covering the range of 100 keV using space-based detectors and up to several PeV measured with ground-based observatories. Gamma-ray photons are messengers of violent events such as cosmic-ray acceleration, core-collapse supernovae and neutron-star mergers. The study of gamma-ray line emissions gives insight into the creation of heavy elements via nucleosynthesis and the evolution of massive stars. Continuous emissions trace cosmic-ray interaction sites and further our understanding of of the origin of cosmic rays. Transient sources such as high flux Gamma Ray Bursts (GRBs) enable the investigation of the source conditions at the most extreme supernovae over a wide range of extragalactic distances. Current instrumentation only covers the range between 100 keV to 1 MeV and energies larger than 30 MeV. This distinct sensitivity gap is called the MeV-Gap. A Compton/Pair Telescope like e-ASTROGAM is designed to provide spectral-resolved imaging of cosmic photons over the range of sub-MeV to several GeV by using a tracking detector D1 and calorimeter D2. In general, the photon flux of cosmic gamma-ray sources is decreasing with emission energy. This sets high demands on data purity, background modelling and reconstruction efficiencies. A complete data pipeline involves classification of individual events and then subsequently reconstructing them by event type. Not all event topologies lead to reconstructible events. Therefore, the classification algorithm must determine the future handling of the event by determining whether to reconstruct the event and which reconstruction to apply. Errors at this stage lead to misreconstruction and information loss. Machine learning-driven algorithms have emerged as the leading approach for a variety of tasks in other fields. This thesis proposes a novel, deep learning driven event classification algorithm to tackle the first and nonetheless very important task of processing and tagging each event, so that it can be processed with the proper reconstruction algorithm. The study uses simulated data generated with the e-ASTROGAM design proposal. The chosen model implementation is geared toward online-classification at the detector in orbit. This approach frees up critical

resources such as storage buffers and transmission bandwidth, while also improving the overall data quality.
Since the Compton Gamma-Ray Observatory no mission offering complete coverage of the MeV-Gap was funded. A way to obtain sensitivity coverage in the lower part of the MeV-Gap is the use of CubeSats. CubeSats are mini-satellites designed to adhere to an industrial-standard. This allows the use of commercially available off-the-shelf components and therefore reduces design and launch costs. The use of standardized parts also enables the construction of multiple small satellites operating in a swarm, while keeping costly redesigns of components to a minimum. A possible application of a CubeSat swarm for gamma-ray observation are constellations of Gamma Ray Burst monitors. Compton scattering conserves the polarization of the scattered photon, allowing the construction of small GRB polarimeters with minimal localization capabilities. Multiple swarms would allow for a complete coverage of the sky with triangulation between the satellites as main localization method. The resulting design proposal was named COMCUBE. The thesis describes the design phase of the detector concept and the first qualification flight of the prototype in August 2023 from Timmins (Ontario).
The thesis is structured into three parts. The first three chapters form the introductory part. Chapter 2 gives an introduction into gamma-ray astronomy, the emission mechanisms and sources. Due to the intended science topic of COMCUBE, Gamma Ray Bursts are discussed in more detail. Chapter 3 discusses detection techniques, the measurement environment in space as well as past, present and future instrumentation. Chapter 4 is introducing key concepts of Machine Learning with a strong focus on Deep Learning and Convolutional Neural Networks. The second part is presenting results obtained for event classification using the e-ASTROGAM design proposal as example. The baseline performance assumes the MEGAlib implementation of ”classification during reconstruction”, which will be presented first in 5.2. A classical machine learning baseline derived from a Random Forest approach is discussed in section 5.6. Section 5.4 discusses the used data sets. A description of the model design, computational overhead and results is given in 5.7, including a comparison to baseline approaches. In the third part, the design phase and first qualification flight of the COMCUBE design proposal are described. Section 6.2 presents the initial design options, performance estimates and design decisions. Section 6.3.1 shows the prototype design and differences from the initial design proposal. Section 6.3.2, 6.3.3 and 6.3.4 discuss the expected background environment, trigger rate estimates and sensitivity estimations for the Crab Nebula. Section 6.3.5 presents the first qualification flight of COMCUBE and results derived from the data. Chapter 7 gives a summary of the thesis and an outlook of future topics to investigate.

Chapter 2

Gamma-Ray Astronomy

2.1 The Electromagnetic Spectrum

Electromagnetic radiation is the main astrophysical messenger. Its spectrum spans over many orders of magnitude from radio emissions, to optical photons to gamma ray emissions in the highest regime. The emission mechanisms and the resulting spectra are giving windows into very different astrophysical phenomena. Figure 2.1 shows the whole range of the electromagnetic spectrum used for observation.
Radio emissions are the lowest energetic photons. They are either of thermal nature, i.e. the emission distribution follows a Maxwell-Boltzmann distribution, or non-thermal and follow e.g. a power-law. Sources are distributed over the whole distance range. The closest sources are in our own solar system, where the sun is the most dominant. Thermal emission can be measured from close-by giant stars or active stars hidden in dust clouds. Here the emissions are generated by e.g. electrons that are in local thermal equilibrium in a hydrogen cloud interaction with HII ions. The resulting bremsstrahlung emission is then called free-free emission (the electron is unbound before and after the interaction). Non-thermal emissions in our galaxy are mostly the result of synchrotron radiation from accelerated electrons. Outside of our galaxy blazars and non-blazar active galactic nuclei are the main steady sources [MTF15].
With increasing frequency, the structure scale moves to smaller sizes that can be resolved. Radio emissions are in part created by electrons moving along magnetic fields, allowing to map the present field, which can span several light years, as in the case of the Crab nebula (see e.g [Con71]). Radio and infrared emission can be used to determine the presence of dust clouds and the extent of molecular clouds. Optical photons allow to measure the temperature and composition of the emitting object. While the former make use of transitions in the shell of the atoms, X-rays and especially gamma rays provide insights into the properties of decaying nuclei and highly energetic acceleration processes.
Observations are done either on Earth with ground based observatories, at high altitudes with stratospheric balloons or on satellites in space. This is warranted due to

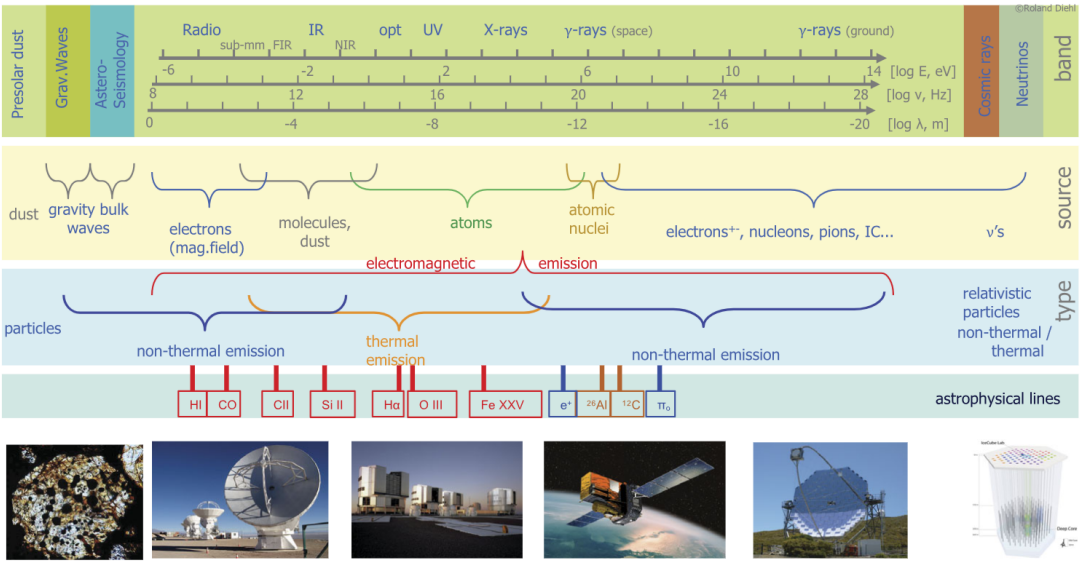


Figure 2.1: Energy range and emission bands for astrophysical sources. Given are the energy ranges, emission types and sources. From [Die21].

the wavelength dependent absorption of electromagnetic radiation. Figure 2.2 shows the altitude at which the atmosphere becomes transparent to electromagnetic radiation as function of frequency. Radio emissions can penetrate the atmosphere almost freely, allowing for large, ground based antenna arrays like the Effelsberg radio telescope [Wie71]. Cosmic infrared radiation gets either absorbed by dust and vapor or drowned in background generated by Earth's atmosphere, making the use of either airplanes (e.g. the now decommissioned Stratospheric Observatory for Infrared Astronomy (SOFIA) [Rei+18]) or satellites (e.g the Infrared Astronomical Satellite (IRAS) [Neu+84] or the James Webb Space Telescope (JWST) [Gar+23]) necessary. Only in the near infrared, ground based observations are possible from very dry mountain tops as for example with the the Visible and Infrared Survey Telescope for Astronomy (VISTA) at the Paranal Observatory [Sut+15]. Optical photons can pass the atmosphere relatively free, allowing for ground based observatories, usually in very dry and atmospherically stable locations like the European Southern Observatory (ESO) with their numerous instruments (e.g. the Very Large Telescope (VLT) [Vlt]). The atmosphere is opaque to UV and X-Ray photons, due to atomic absorption. Therefore space-bourne observatories are mandatory. The Hubble Space Tescope (HST) houses the Cosmic Origins Spectograph (COS [Gre+12], which operates in a wavelength range of 90-320 nm and is an example of an UV instrument. There have been numerous X-Ray missions over the course of the years, the most prominent ones being Chandra [Wei+00] and XMM-Newton [Sch+24].
The targeted photon band for Compton and Pair telescopes are γ-rays. Due to the density of the atmosphere, they are prohibited from reaching the ground, resulting in the need of observations with satellites (see later examples) or ground-based ob-

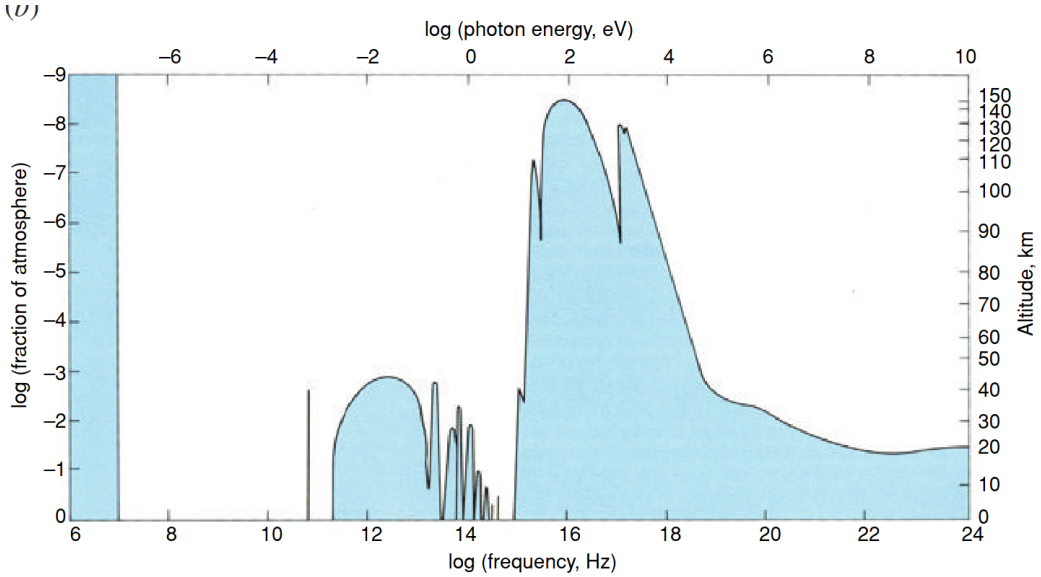


Figure 2.2: Transparency of the Earth atmosphere to electromagnetic radiation as function of frequency or energy. The solid line indicate the height above sea-level at which the the atmosphere becomes transparent for photons at the respective wavelength. From [Lon12].

servation of generated air-showers in the Earth's atmosphere (e.g. the Cherenkov Telescope Array (CTA) or the High Energy Stereoscopic System (H.E.S.S.) [CTA13; Aha+06]). They are ranging from the lower end at around 100 keV up to energies up to order of PeV. At the lower end, the emitted photons are the result of electrons of atoms subjected to strong magnetic fields. Once the scale of several keV up to order of MeV is reached, the source of the photon is the nucleus itself, being the result of nuclear deexcitation. At energies higher than 5 MeV, the emitted photons are either the direct result of accelerated particles (e.g. Bremsstrahlung of electrons, Synchroton emission), annihilating particles (mainly Pions) or photons that underwent Inverse Compton Scattering. The different spectral shapes allows the distinction of processes and conditions at the source's location. Nuclear deexcitation and direct annihilation of particle and anti-particle results in line spectra. Photons emitted by acceleration processes and Inverse Compton have a continuous spectral shape. The characteristics of these emissions will be discussed below.

2.2 Line Sources

Gamma-ray lines are the tracers of radioactive decays resulting from various astrophysical processes. They cover the lower end of the gamma-ray spectrum. The most prominent line is the 511 keV positron annihilation line. The first measured decay line was from the radioisotope ^{26}Al, which can be measured from the whole Milky-

Table 2.1: Gamma-ray lines of astrophysical relevance. Starred line energies denote not yet measured transition energies in an astrophysical context. Citations of first measurements are given, where applicable. Adapted from [Die01]

Isotope	Lifetime	Decay Chain	Energy[keV]	Measured	Source
^{7}Be	77d	$^{7}\mathrm{Be} \rightarrow {}^{7}\mathrm{Li}^{*}$	478*	None	Novae
^{56}Ni	111d	$^{56}\mathrm{Ni} \rightarrow {}^{56}\mathrm{Co}$ $\rightarrow {}^{56}\mathrm{Fe}^{*} + \mathrm{e}^{+}$	158, 812, 847, 1238	[Mat+88]	SN Ia
^{57}Ni	390d	$(^{57}\mathrm{Ni}\rightarrow)\ {}^{57}\mathrm{Co} \rightarrow {}^{57}\mathrm{Fe}^{*}$	122	[Kur+92]	SN II
^{22}Na	3.8y	$^{22}\mathrm{Na} \rightarrow {}^{22}\mathrm{Ne}^{*}+\mathrm{e}^{+}$	1275*	None	Novae
^{44}Ti	89y	$^{44}\mathrm{Ti}\rightarrow {}^{44}\mathrm{Sc}^{*}$ $\rightarrow {}^{44}\mathrm{Ca}^{*} + \mathrm{e}^{+}$	76, 68, 1157	[Iyu+94]	SN II (mass cut)
^{26}Al	$1.04\cdot 10^{6}$y	$^{26}\mathrm{Al}\rightarrow {}^{26}\mathrm{Mg}^{*}+\mathrm{e}^{+}$	1809	[Mah+82]	SN II, massive stars
^{60}Fe	$3.8\cdot 10^{6}$y	$^{60}\mathrm{Fe}\rightarrow {}^{60}\mathrm{Co}^{*} \rightarrow {}^{60}\mathrm{Ni}^{*}$	59*, 1173, 1332	[Smi04]	SN II
e^{+}	10^{5} - 10^{6}y	$\mathrm{e}^{+} + \mathrm{e}^{-} \rightarrow$ Positronium $\rightarrow \gamma\ \gamma\ (\gamma)$, inflight annihilation	<511, 511, >511	[JH73a], [Kno+25]	various

Way. A summary of the most interesting lines and their broader context is given in Table 2.1.
The line intensity is directly proportional to the activity $A = \frac{N}{t}$, the number of decaying isotopes in a given time interval. For individual sources like supernovae and their remnants, one can estimate the initial number of isotopes present by extrapolating from the number of measured decays via the exponential decay law:

$$N_{present}(t) = N_0 \cdot \exp\left(-\frac{t \cdot ln2}{T_{half}}\right) = N_0 \exp\left(-\frac{t}{\tau}\right), \tag{2.1}$$

where N_0 is the initial number of isotopes, $N_{present}$ is the number of isotopes present at time t and $T_{1/2}$ is the half-life of the given isotope and $\tau = \frac{T_{half}}{\ln 2}$ is the lifetime of the isotope respectively. For long lived isotopes,one can average over many sources that were produced at an average rate $\bar{P}$ over the last few lifetimes of the isotope:

$$A = \frac{N_{present}(t)}{\tau} = \bar{P} \tag{2.2}$$

As can be seen from Table 2.1, most observed lines have a connection to Supernovae. Positron annihilation radiation takes a unique role compared to the other lines. Due to the fact, that a large number of processes are capable of producing photons with sufficient energy, the potential sources of positrons are also exhibiting a large variety. The lifetime of most of these isotopes is very short. They can be used as tracers for very recent SN events. The longer lifetimes of ^{26}Al and ^{60}Fe make it feasible to derive

more global properties of our galaxy as discussed bellow. Further details about other astrophysical relevant isotopes can be found in [Die22]

2.2.1 ^{26}Aluminium

^{26}Al and ^{60}Fe are two isotopes generated by or prior to SN II events. ^{26}Al is created in proton-rich environments such as hot hydrogen burning via the reaction

$$^{25}Mg(p,\gamma)^{26}Al \tag{2.3}$$

and destroyed in neutron-rich environments such as, e.g. He-burning, via

$$^{26}Al(n,p)^{26}Mg, ^{26}Al(n,\alpha)^{23}Na \tag{2.4}$$

It can be traced via its β-decay ^{26}Al$(\beta^+)^{26}$Mg yielding a characteristic 1.809 MeV γ-photons from the excited ^{26}Mg* nucleus, as shown in Figure 2.3.
It is ejected into the interstellar medium by pre-supernovae processes, the core collapse supernova itself or stellar winds of Wolf-Rayet stars [Die+21a]. ^{26}Al in the in-

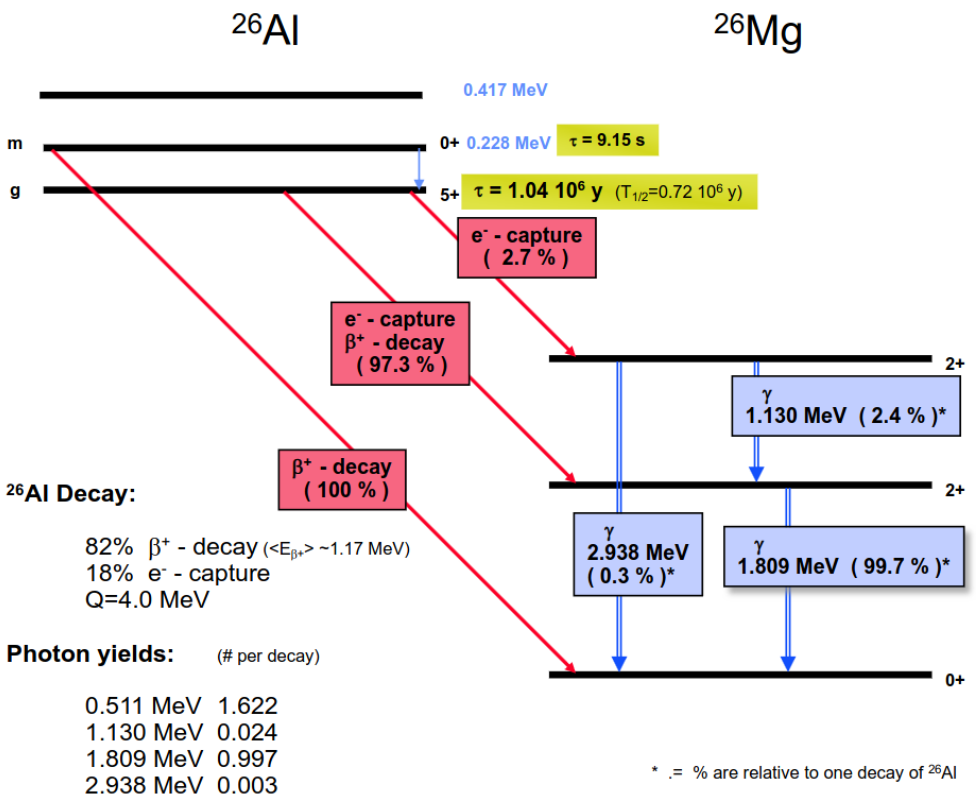


Figure 2.3: ^{26}Al decay scheme. From [Die22].

terstellar medium was the first measured nucleosynthesis line by HEAO-3 [Mah+82] with follow up measurements by COMPTEL on CGRO [Obe+96; Plu+01] and SPI on INTEGRAL [BJR15].
The maps in Figure 2.4 show diffuse emissions extending over the whole Galaxy, but also the presence of clusters. One example of a localized source is the region around the Vela Supernova Remnant (SNR). The diffuse emission flux can be used to derive an estimate of the star formation rate (SFR) in our galaxy. The center of the emission line is subject to Doppler shift. This allows to use the shape-deformation of the line to construct a rotation model of the galaxy [Kre+13], which can then be used for an estimate of the total amount of ^{26}Al. Diehl et al and Pleintinger [DHP18; Ple20] concluded from the present flux a mass of around 2 million solar masses. The value can be used to determine the average rate of core-collapse supernovas in the galaxy [Die+06]. The initial value 1.91 ± 1.1 per century was later refined to 1.4 ± 1.1 per century [DHP18; Ple20], equating to 1 average Supernova every 71 years. This value is in tension with the 294 observed SNRs in the Milky way [Gre19]. The rate would suggest that only Supernova remnants from the past 21000 years would be measurable. Estimates of the age of some of these SNR are exceeding this time scale significantly(see e.g. the Monogem Ring with $6.8 \cdot 10^4$ years [KSP18]). Implications to the estimated star formation rate in our Galaxy are discussed for example in [CP11] and illustrate the need for a better understanding of underlying model parameters.

2.2.2 ^{60}Iron

^{60}Fe is the product of successive neutron capture of stable Fe isotopes in neutron rich environments and has a half-life of 2.6 My [Rug+09]. It decays into ^{60}Co via a β^+ decay and from there to ^{60}Ni as shown in Figure 2.5, emitting two γ-photons with an energy of 1173 and 1332 keV. Both ^{60}Fe and ^{26}Al are produced by massive stars, but in different phases and locations of the star, as ^{60}Fe requires a neutron-rich environment in contrast to the requirement of a proton-rich environment for the production of ^{26}Al. Both isotopes can be synthesized in the inner parts of the star during Helium and Carbon shell burning or during explosive nucleosynthesis and are then ejected during the Supernova event. ^{26}Al however is also produced in the outer shells during H-burning, allowing for ejection via stellar winds before the explosion resulting in higher yields [Die+21a].
Since both isotopes share the same assumed source (massive stars), the ratio of the observed flux of ^{60}Fe to ^{26}Al can be directly compared to stellar yield ratios predicted by stellar evolution models for massive stars [Die+21b]. The measured flux ratio with 15 years of INTEGRAL data was determined to 0.184 ± 0.042, using exponential grid maps [Wan+20]. Two major uncertainties related to the observation of decaying ^{60}Fe nuclei are the number of generated ^{59}Fe nuclei as well as their decay rate. The stability of ^{59}Fe may vary in stellar environments compared to terrestrial observations due to decays from excited states. Gao et al measured the Gamow-Teller transition strengths and calculated the β-decay rates from excited states, observing an overall lower life-time than most model predictions, resulting in a decrease of 40% of the

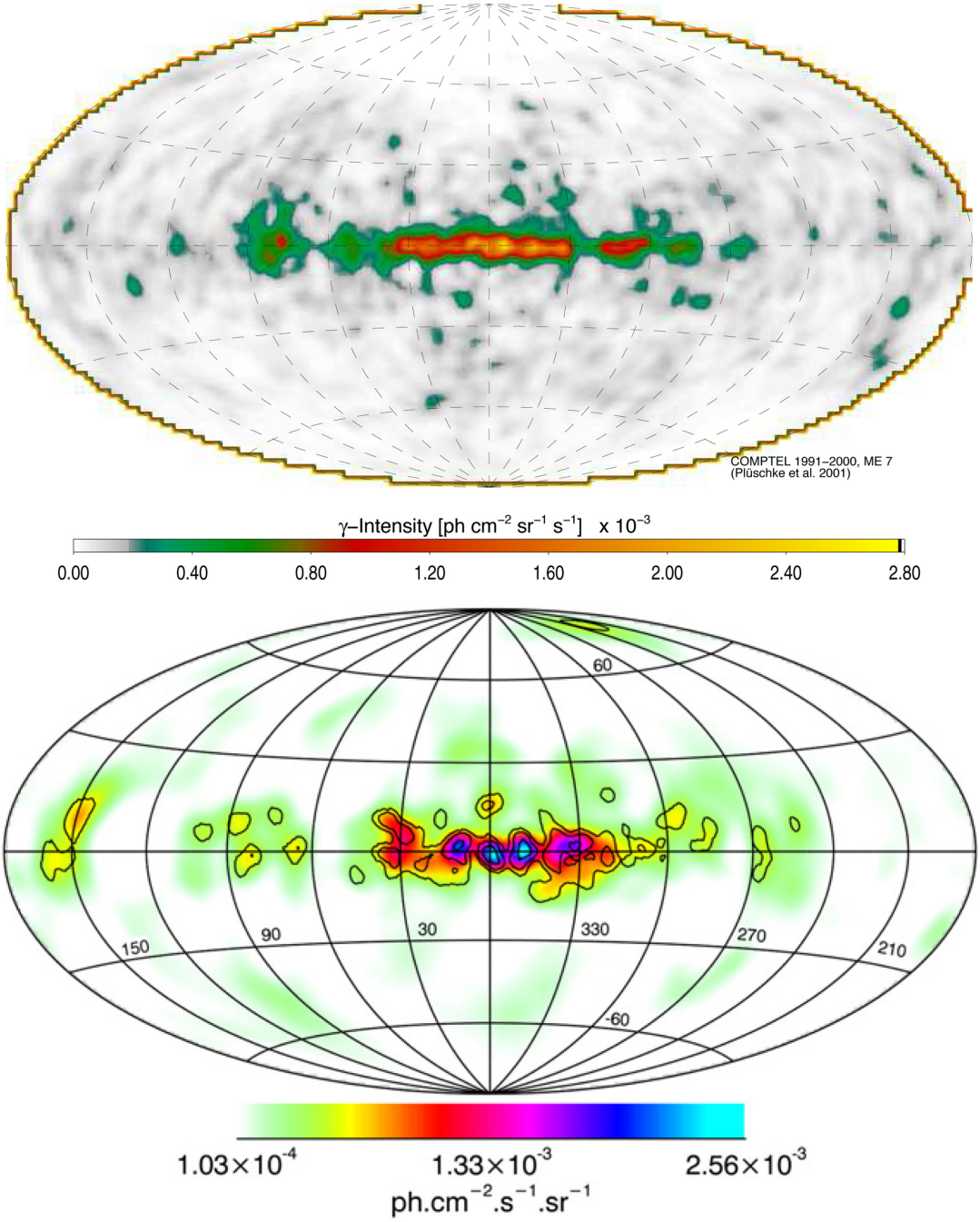


Figure 2.4: Flux maps of ^{26}Al measured by COMPTEL (top) and INTEGRAL(bottom). Both maps show the main flux originating from the center of the Galaxy and also contribution of local structures. From [Plu+01] and [BJR15].

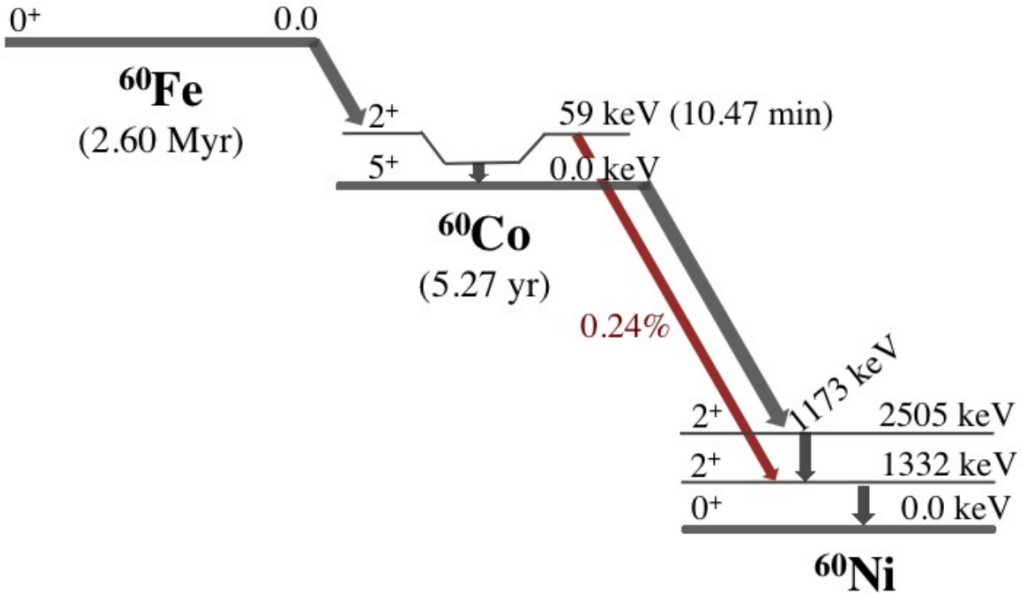


Figure 2.5: ^{60}Fe decay scheme. Given are the half-life of the states and decay energies. From [Hef+15].

previously modeled ^{60}Fe yields [Gao+21]. Also the neutron capture cross section of ^{59}Fe has been the subject of recent research. Spyro et al [Spy+24] found an increased cross section for the $^{59}\mathrm{Fe}(n,\gamma)^{60}\mathrm{Fe}$ reaction, indicating higher expected yield ratios. This points at the conditions to the source (mostly temperature in the producing shells and rotation) as a likely source for the observed discrepancies between models and observational data.

2.3 Positron Annihilation

The 511 keV annihilation line was first observed as diffuse gamma ray emission in the 1970s by a series of balloon experiments [JHH72; JH73b] and then later confirmed as positron annihilation by [LMS78; Alb+81]. The Compton Gamma Ray Observatory and INTEGRAL found the measured emission maps to be consistent with diffuse emissions, spanning from the center of the Milky Way (bulge region) to the outskirts of the galactic plane (disc region) [Pur+97; Chu+20]. Two example maps are shown in Figure 2.6. The majority of the measured 511 keV radiation can be attributed to the annihilation of positrons around the galactic center, as the flux in the bulge is larger than in the disc region. However, the attribution of the emitted photons to individual positrons sources is still an open question. One of the main challenges is the very low density of the interstellar medium, ranging from 0.01 to 10 particles per cm^3 [Cox05], resulting in lifetimes in the order of $10^5$to 10^7 years as found by e.g. [Jea+09] in a simulation study. This allows the positrons to travel long dis-

tances before finally annihilating at a position very far away from their origin. Since positrons also carry charge, their path can be deflected when passing strong electromagnetic fields, further obscuring their true place of origin. Possible hot spots in imaged annihilation photons are therefore not necessarily a hint for a possible positron source at this point, but can rather be the result of the local gas density distribution in that region.

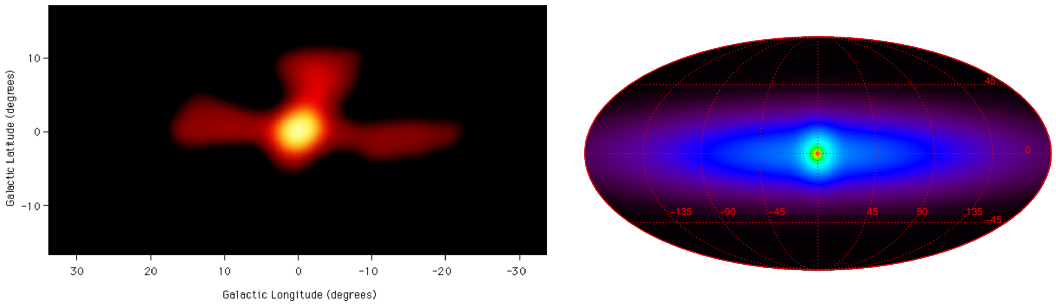


Figure 2.6: Maps of the measured 511 keV emission from OSSE (left, [Pur+97]) and the best fitting model of a bulge and disk component from INTEGRAL SPI data (right, [Sie+16]).

2.3.1 Characteristics of Positron Annihilation Radiation

The 511 keV radiation is the result of electron-positron pair annihilation in the intergalactic medium. Since the inter stellar medium is very thin, direct annihilation is suppressed and the positron will form an orbiting system with an electron called Positronium before annihilation. The total spin of the system can be either 1 or 0 resulting in either ortho-positronium or para-positronium. Para-positronium results in the emission of two annihilation photons at 511 keV gamma photons that give the characteristic line spectrum. Ortho-positronium on the other hand has to result in the emission of an odd number of photons due to spin-conservation. The available energy of 1022 keV is distributed into three photons in the most probable case [OP49]. The resulting spectrum is now a continuum spectrum peaking of at 511 keV as shown in Figure 2.7 on the left. A third mode is inflight annihilation, resulting in a Doppler-shifted two-photon continuum emission in the range of $m_e c^2/2$ to $E_p + m_e c^2 \cdot 3/2$, where $m_e c^2$ is the rest mass of the electron and E_p is the kinetic energy of the positrons [Kno+25]. The measured positron annihilation spectrum is the superposition of these annihilation modes and background components. Figure 2.7 shows the measured spectrum from SPI on Integral and the model components of the fit on the right. In addition to the para- and ortho-positronium component, the model takes into account the photon background and a broadened line contribution [Sie+19]. Note that inflight annihilation influences the spectrum of the line emission to only a very small extend.

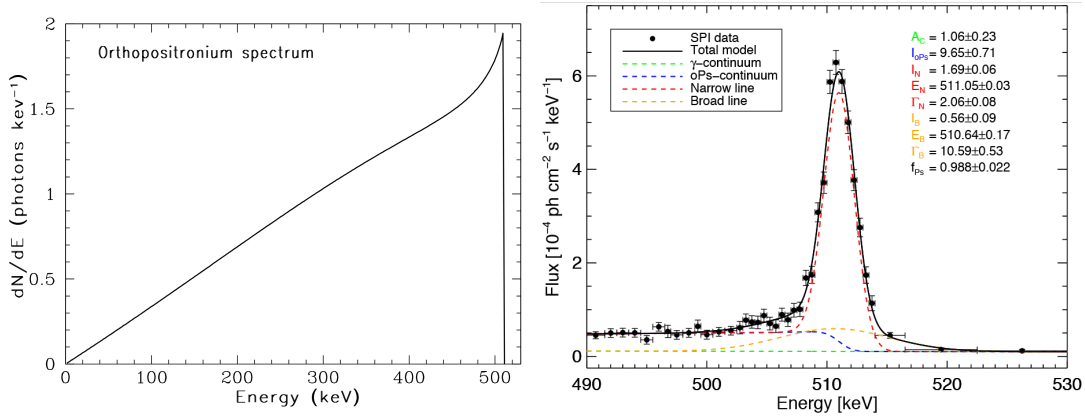


Figure 2.7: Left: Ortho-positronium spectrum for 3-photon decay. The peak energy is close to 511 keV. From [OP49]. Right: Positron annihilation spectrum of the Milky Way measured with Integral-SPI. Data points in black are taken from bins with more than 3 σ significance. The black line corresponds to the total fit, the model components are shown in the legend. From [Sie+19].

There are multiple production mechanisms, that generate positrons in one of the intermediate or final states. The most important ones are described below.

Nucleosynthesis

Positron emission can occur during radioactive decays of proton rich nuclei. Some examples with astrophysical relevance are the decays of ^{26}Al, ^{44}Ti, ^{22}Na and ^{56}Co from ^{56}Ni. The lifetime of the listed isotopes is ranging from 10^6 years to 10^2 days and they are created in very varying environments as can be seen in Table 2.1. For example ^{56}Ni decays into ^{56}Co and is created during Supernova explosions. Because of its lifetime of 114 d, the positron will likely be emitted into the rather dense and turbulent environment of the stellar ejecta at the late stage of the supernova event or the beginning of the remnant stage. This leads to an annihilation relatively close to the SNR-site for a majority of the injected positrons and only a small portion of the injected positrons can escape into the interstellar medium [CL93]. ^{26}Al on the other hand has a mean lifetime of 10^6 years and is released into the ISM by stellar winds, allowing the mother nucleus to travel a considerable distance away from the production site, leading to a diffuse emission pattern, which is further diluted by the distance that the positron can travel until annihilation [Die+06; Jea+09]. Simulations of the emission maps of positrons originating from these isotopes (see e.g. [Ale+14]) can be used to correlate the measured spatial distribution of the positron annihilation radiation to individual isotopes, a topic to be investigated with COSI (see Section 3.4.4).

Acceleration Processes

At cosmic acceleration sites, proton-proton collision at energies above the respective production threshold can trigger the production of mesons, where pions possess the lowest production threshold of around 280 MeV. Each positively charged pion will result in the emission of a positron [Nav+24] either in the dominant channel indirectly via:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \tag{2.5}$$

and

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu_\mu}. \tag{2.6}$$

or directly via

$$\pi^+ \rightarrow e^+ + \nu_e \tag{2.7}$$

Since the production takes place at a site with strong electromagnetic fields and at high energies, the produced mesons possess a large amount of kinetic energy.

Pair Production

Pair production can occur via the interaction of high energetic photons in either the electric field of a nucleus (and to a lesser extend the field of an electron) or by photon-photon interaction at energies above the pair threshold in the presence of a very strong magnetic field. The created pairs can then be either propagate away from their source and annihilate as ortho- or para-positronium, annihilate inflight or thermalize as in an accretion disk. There they can form an electron-proton-positron plasma. Studies by Beleberov et al show, that a part of this plasma can be carried away as a wind [Bel99] and the positrons therein will annihilate according to the plasma temperature [Sve82b; Sve82a; Sve83]. This will result in a Doppler boosted and broadened spectrum with no clear line shape [RM81].

2.3.2 Astrophysical Sources

Positrons are produced by a very broad variety of production mechanisms. The necessary source conditions are therefore also achievable by a wide variety of sources.
Massive stars are a source of the long-lived ^{26}Al, while core-collapse (SNII) and thermonuclear supernovae (SNIa) can account for β^+- decaying radioactive isotopes with lesser lifetime (see Section 2.2.1 and Table 2.1). The production rate can be estimated from the measured decays on the sky for the respective isotope for a given volumetric model of the Milky Way and the correct modelling of the absorption probability from the point of the decay to the observer. The most important isotopes to consider are ^{26}Al, ^{44}Ti and ^{56}Co [Sie23]. While these radionuclei may explain the disk component of the 511 keV emission, they can not explain the strong bulge emission [Mar+12; Ale+14]
Pulsars are candidates for positron production via pair creation in strong magnetic fields. The production rate can not be measured directly and has to be estimated

from theoretical modelling or observations of cosmic ray positrons in the Solar System. The positron production rate can be only estimated theoretically (see [ZC97; CHR86].
Low mass X-Ray Binaries are systems in which a compact object like a black hole or neutron star accretes mass from a smaller companion. Positrons are then created in the accretion disc via photon-photon-pair-production. Annihilation radiation from such objects has been measured inconclussively from the "Great Annihilator" [Bou+91], Nova Musca [Gol+92] and V404 Cygni [Sie+16b].
Models suggest, that the majority of produced Positrons are originating from LXRBs and Microquasars (52%). The remaining 48% are attributed mostly to Type 1a super novae (37%), massive stars (6%) and core-collapse super novae (4%). Only a sub-percent share is attributed to Pulsars and Dark Matter. The exact composition is still under debate, since the uncertainties of measurements and models are still large [Sie23]. Investigating the nature of positron annihilation in our galaxy will be one of the main research topics of COSI.

2.4 Continuous Emission

Continuous gamma emission can occur from thermal or non-thermal processes. For thermal radiation the spectral energy density follows

$$I(\nu) = \frac{8\pi h\nu^3}{c^3} \cdot \frac{1}{\exp h\nu/kT - 1} \tag{2.8}$$

and represents the spectral radiance of black bodies. In order for the spectrum to reach into the MeV regime, the temperature of the object in question would need to reach the order of 10^9 K [Die01]. Non-thermal processes are therefore more common sources for γ emissions at energies higher than ≈ 100 keV Gamma-rays from non-thermal sources are mainly generated from processes relating to strong fields accelerating charged particles.

2.4.1 Bremsstrahlung

Bremsstrahlung is the emission of electromagnetic radiation induced by deaccelaration of electrons in the field of other charged particles. Depending on the source conditions, the emission spectrum can be thermal or non-thermal. The low end for energies up to 30 MeV of the observed gamma-spectrum can be attributed to bremsstrahlung emission from electrons below 100 MeV [SS84]. The shape of the spectrum depends on the matter distribution where the photons are created, as well as on the energy of the accelerated electrons. In the case of electron-nucleus scattering in, e.g., a gas cloud, the change of trajectory of the electron results in a change of kinetic energy. The intensity per unit frequency is then given by [Die01]

$$I_\nu(E_e) = \frac{Z^2 e^6 n}{12\pi^3 \epsilon_0^3 c^3 m_e^2 v_e} \cdot ln\left(\frac{192 v_e}{Z^{1/3} c}\right) \tag{2.9}$$

Here Ze is the charge observed by the electron and n is the number density of matter. The resulting photon spectrum is flat up to

$$E_\gamma = (\gamma - 1)m_e c^2 \tag{2.10}$$

where γ is the Lorentzfactor $\sqrt{\frac{1}{1-\frac{v^2}{c^2}}}$ and drops to zero quickly as almost all kinetic energy of the electron has been transferred to the emitted photon. In hot plasma, the emission spectrum is of thermal nature and is a function of the plasma temperature. It takes the form [Die01]:

$$k_\nu = 5.443 \cdot 10^{-39} \cdot Z^2 \cdot g(\nu, T) \cdot exp\left(-E_\gamma / kT\right) \cdot \sqrt{T} \cdot n_e \cdot n_p \tag{2.11}$$

$g(\nu, T)$ is the Gaunt factor. It is a correction factor for the non-relativistic solution which can be approximated by for the gamma regime by $\sqrt{E_\gamma / kT}$. In the non-thermal case of e.g. Fermi-acceleration, the emission spectrum takes the shape of a powerlaw with index between -1 and -3.

2.4.2 Synchroton Emission

The emission of extreme relativistic charged particles due to the presence of a magnetic field is called *synchroton emission*. Following [Die01; Fun15] the shape of the synchroton spectrum peaks at

$$\nu_c = \frac{3}{2} \cdot \frac{eB}{2\pi m_e \cdot sin\Theta} \tag{2.12}$$

where B is the strength of the magnetic field and Θ is the angle between the trajectory of the accelerated particle and the orientation of the magnetic field. The energy of the emitted photons is then [AAK97]

$$E_{sync} = 0.2 \frac{B}{10\mu G} \left(\frac{E_e}{1TeV}\right)^2 eV \tag{2.13}$$

The energy of the photon is therefore proportional to the strength of the magnetic field and quadratic in the electron energy. Assuming a powerlaw distribution of the electrons α_e leads to a differential synchroton spectrum $\Gamma_{sync} = \frac{\alpha_e + 1}{2}$ [BG70].
The injected electrons loose energy proportional to E^{-1}. When the time scale of the energy loss is on a similar time scale as the age of the source, the electron spectrum has a break at

$$E_{e,br} = 1.2 \cdot 10^4 \left(\frac{B}{10\mu G}\right)^{-2} \cdot \left(\frac{t_{age}}{10^4 a}\right)^{-1} \tag{2.14}$$

from α_e to $\alpha_e + 1$, where t_{age} is the age of the source.

2.4.3 Inverse Compton Scattering

Relativistic electrons can scatter off photons. The transferred energy results in a blue-shift of the scattered photons into the high-energy gamma-ray regime.
This process is dominant in regions with high photon densities which form the target for the relativistic electrons. The spectral intensity distribution is then [Die01]

$$j_\nu(\nu_0) = \frac{3\sigma_T c}{16\gamma^4} \cdot \nu \cdot \left(2\nu \cdot ln\left(\frac{\nu}{4\gamma^2\nu_0}\right) + \nu + 4\gamma^2\nu_0 - \frac{\nu^2}{2\gamma^2\nu_0}\right) \tag{2.15}$$

where σ_T is the Thompson-cross section. For sufficiently high energies the Thompson cross section has to be replaced with the Klein-Nishina cross section. The energy of the up-scattered photons peaks at

$$E_{IC} = 1.3\left(\frac{E_e}{1TeV}\right)^2 \cdot \left(\frac{E_{ph}}{2 \cdot 10^{-4}}\right) GeV \tag{2.16}$$

with E_{ph} being the typical energy of the ambient photon-field. Since the electron energy enters quadratically, inverse Compton scattering is the dominant production mechanism of the highest energy gamma photons.

2.4.4 Pion Bump

Acceleration of protons at cosmic accelerator sites can lead to the creation of neutral current pions when the protons undergo strong force interactions with each other or other nuclei from ambient gas in collisions:

$$p + p \rightarrow p + p + \pi^0 \tag{2.17}$$

The dominant decay mode is the decay into two photons

$$\pi^0 \rightarrow \gamma\gamma \tag{2.18}$$

where each of the photons carries roughly half of the rest mass of the pion. The created pions carry some momentum from the highly energetic scattering, which leads to a broadening and Doppler-shifting of the measured photons.

2.5 Gamma-Ray Bursts

Gamma-Ray Bursts (GRBs) are transient high flux emissions of photons in the range of several keV to several GeV. The temporal emission profile can be divided into the prompt emission of gamma rays ("the burst") and afterglow emissions ranging up to weeks over the whole electromagnetic spectrum [Lev18]. GRBs were first measured in the 1960s by the Vela satellites designed to monitor Earth for tracers of thermonuclear explosions. An analysis of the data was only published much later by [KSO73] in 1973, confirming the extra-terrestrial origin of the measured photons, but immediately raising further questions:

1. Where is the origin of these sources?
2. What are the characteristics of these sources?
3. What are the processes that generate the necessary amount of energy?

These questions are central science goals of the COMCUBE-Project and are discussed in more depth.

2.5.1 Origin

The first large scale GRB-survey was done with BATSE (see Section 3.4.1) on the Compton Gamma-ray Observatory. It was able to record GRB triggers at a rate of roughly 300 per year. BATSE was also able to show, that the spatial distribution of the reconstructed locations in the sky showed no anisotropies, especially in the bulge and disc of the Milky Way [Bri+96]. The reconstructed locations in galactic coordinates are shown in Figure 2.8.
The extra-galactic nature of GRBs was confirmed with a multi-instrument measurement performed by BeppoSAX and the Keck observatory. During its live-time from 30 April 1996 to 30 April 2002 BeppoSAX [Fro19], an Italian and Dutch X-Ray satellite mission, detected GRBs with its own burst monitor (Gamma Ray Burst Monitor (GRBM)) and then observed and localized the X-Ray after-glow with their Wide Field Cameras (WFCs).
This allowed to pinpoint the origin of several GRBs down to order of arc-minutes, allowing further, more accurate measurements of the afterglow in the optical range. This was successfully done for the first time in 1997 with GRB970508 [Ama+99] and measurements of the Keck Low Resolution Imaging Spectrograph [Met+97]. They obtained a redshift value for Fe-lines measured of the source's afterglow-emissions at $z = 0.835$, confirming the extra-galactic nature of GRBs.
The approach to use multiple wavelengths to subsequently improve the localization as well as measuring the emission of a GRB over a large range of the electromagnetic spectrum was furthered in the design of the Neil Gehrels SWIFT-observatory [Bur+05]. Its scientific payload consists of three main instruments. The initial GRB detection is achieved with the Burst Alert Telescope (BAT), a coded mask telescope sensitive to energies up to 150 keV. The satellite bus is optimized to rotate into a new pointing

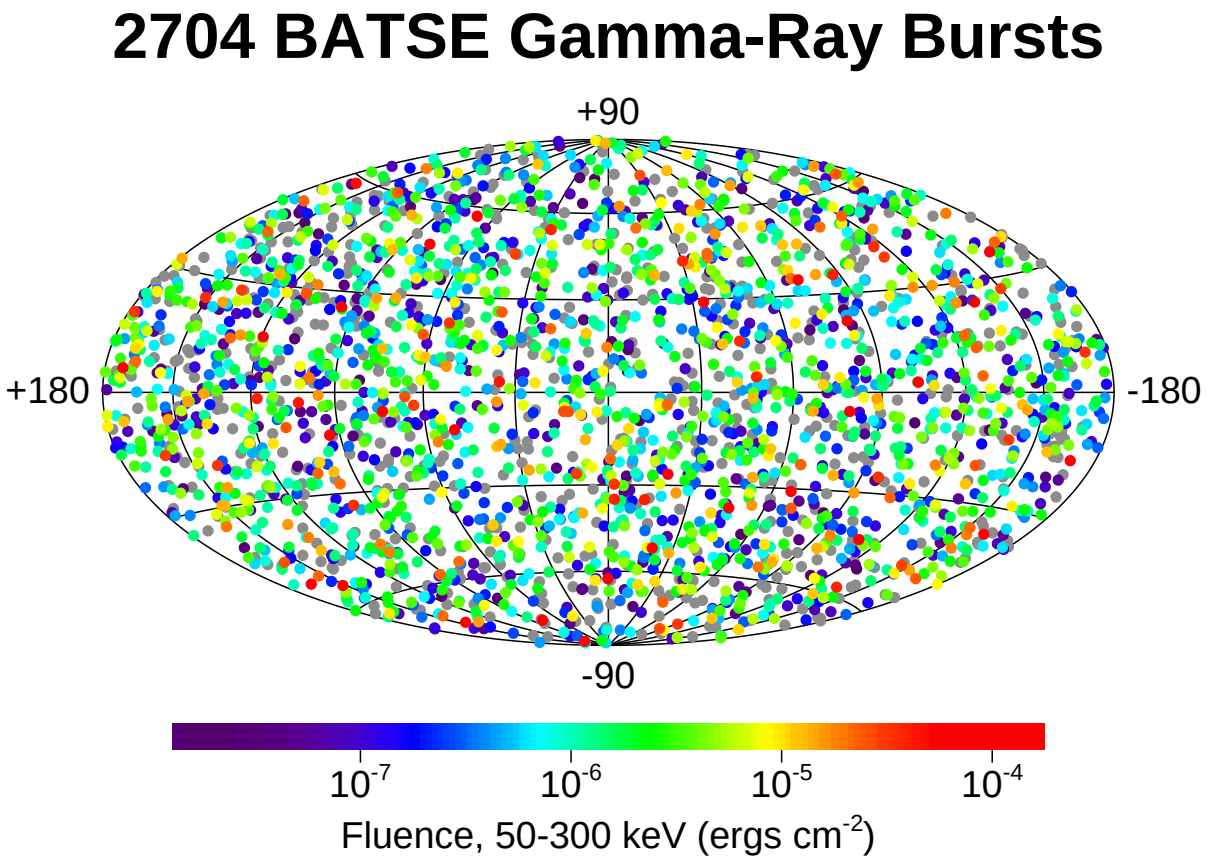


Figure 2.8: Reconstructed locations of 2704 GRBs in the BATSE4B catalogue. The color code gives the integrated flux of the GRB. Grey dots correspond to bursts with incomplete data. From [Mee+98].

fast, precisely and is a fully automatic way to allow measurements of the afterglow with the X-Ray Telescope (XRT) and the Ultraviolet/Optical Telescope (UVOT) covering an energy range 150 keV down to the optical range in one satellite. This allowed measurement of short GRBs, which live on a time-scale of up to 2 seconds, with a similar precision as for long GRBs.

2.5.2 Source Characteristics

Observations on the duration of the prompt emission on a large sample of GRBs with the Burst and Transient Source Experiment (BATSE) [Hor91] on the Compton

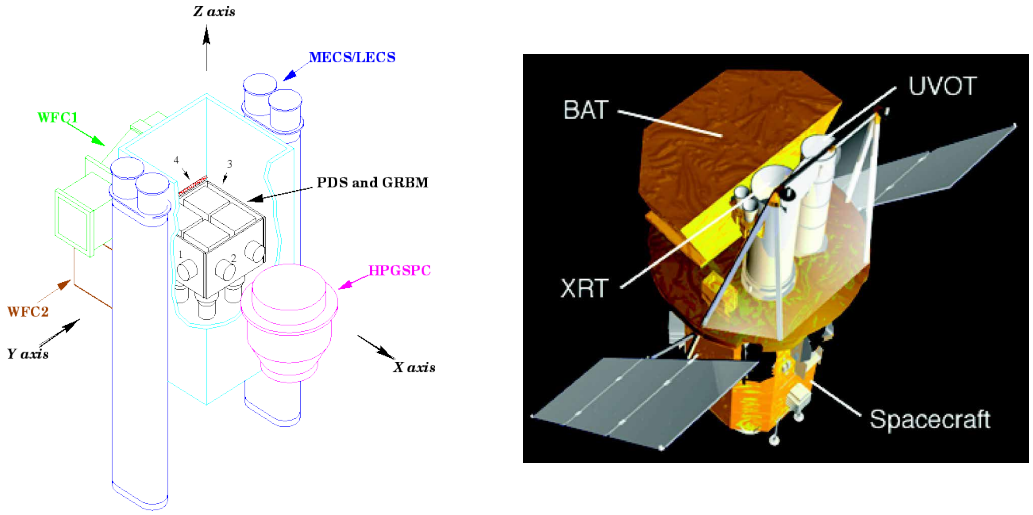


Figure 2.9: Left: Sketch of the instrumentation placement of BeppoSax. The two main instruments used for GRB detection and localization were GRBM, WFC1 and WFC2. The main problem with the setup was the perpendicular pointing of WFC with respect to GRBM, resulting in the necessity of a manual repointing of the instrument after a GRB detection. From [Fro19]. Right: Render of the Neil Gehrels SWIFT observatory. To reduce the time between detection of the GRB and follow up measurements, the Burst Alert Telescope is pointing along the same axis as the rest of the instrumentation. As a result, the pointing to the burst location can be performed automatically and very fast. From [Bur+05].

Gamma Ray Observatory (CGRO, described in Section 3.4.1) suggested at least two species of GRBs [Mee+98]. Figure 2.10 shows the distribution of the t_{90} durations (time between 5 and 95% of all measured background subtracted counts of a GRB trigger) of GRBs contained in the final BATSE catalogue. The data suggest at least a bimodal distribution, which led to the definition of a GRB being called short, if t_{90} is smaller than 2 seconds [Kou+93a] for BATSE recorded GRBs.
Since the exact underlying processes generating the prompt emission are not known very well, one can not rely on superposition of known emission spectra to fit the emission spectrum. The spectral shape of GRBs for energies up to several MeV can however be empirically well described with a Band function [Ban+93]. The flux can be calculated as

$$F_E = \begin{cases} E^{\alpha} \cdot \exp\left(-\frac{E}{E_0}\right), & E \leq (\alpha - \beta) \cdot E_0 \\ [(\alpha - \beta) \cdot E_0]^{(\alpha-\beta)} \cdot E^{\beta} \cdot \exp(\beta - \alpha), & E > (\alpha - \beta) \cdot E_0 \end{cases} \tag{2.19}$$

with a smoothly joined broken powerlaw. α is thereby the index of the low-energy part up to the break point E_0, while β is used to describe the shape for energies beyond. An example is given in Figure 2.11. A common interpretation of E_0, often also called break energy, is the change of low energy emission to high energy emission

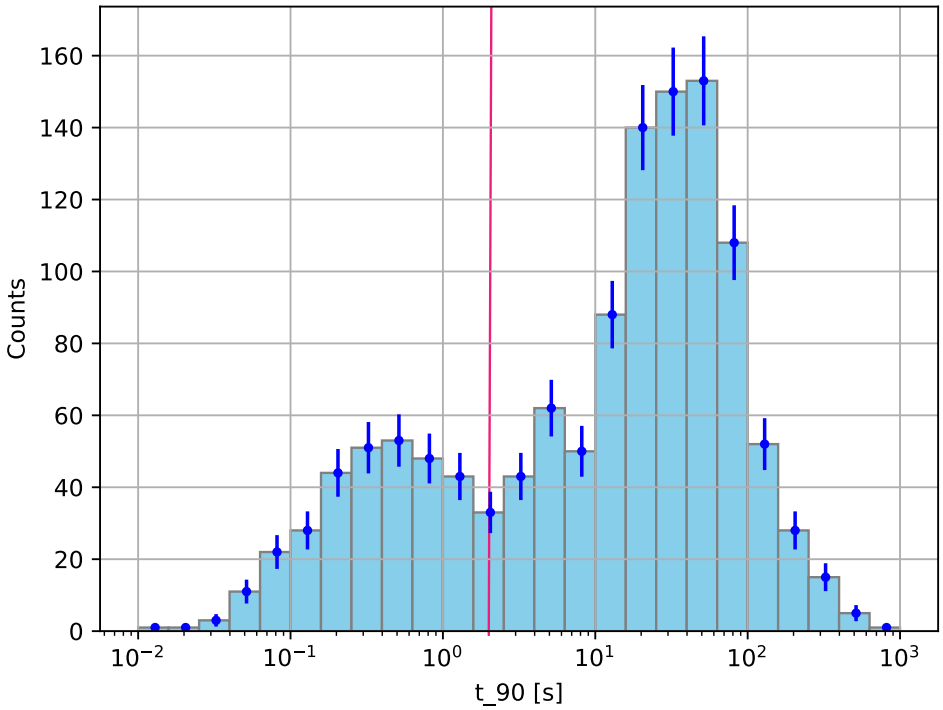


Figure 2.10: Burst durations for all recorded triggers in the BATSE 4B catalogue using the duration in which 90% of the detected emission was found (t_{90}). The red line marks the 2 s classification border between short and long GRBs. Generated from [Mee+98].

mechanism past the break point.
Measurements with Fermi-LAT [Atw+09] and GBM [Mee+09] have shown, that GRB emissions can extend into the GeV range. An example is given in Figure 2.12. The emissions in the range above 40 MeV still show similar behaviour to the count spectra in the lower range and can be fitted with the same model [Abd+09] with small modifications.
In 2019 MAGIC [Fer05] was able to measure photons of the burst's afterglow emission from GRB190114C at the TeV scale [Nat]. Analysis of the data from SWIFT, GBM, Fermi-LAT and Magic suggested the presence of synchroton self-Compton-radiation at higher energies [Acc+19]hinting at different processes responsible for prompt-emission and afterglow. A fit of the model is shown in Figure 2.13.
The energy spectra of short and long GRBs show similarities in the higher energy range but also differences in the hardness of the spectrum at the peak energy [Kou+93b]. The distribution for the BATSE 4B catalogue is shown in the histogram in Figure 2.14. The Hardness of the spectrum is defined as the ratio of the high energy fluence over

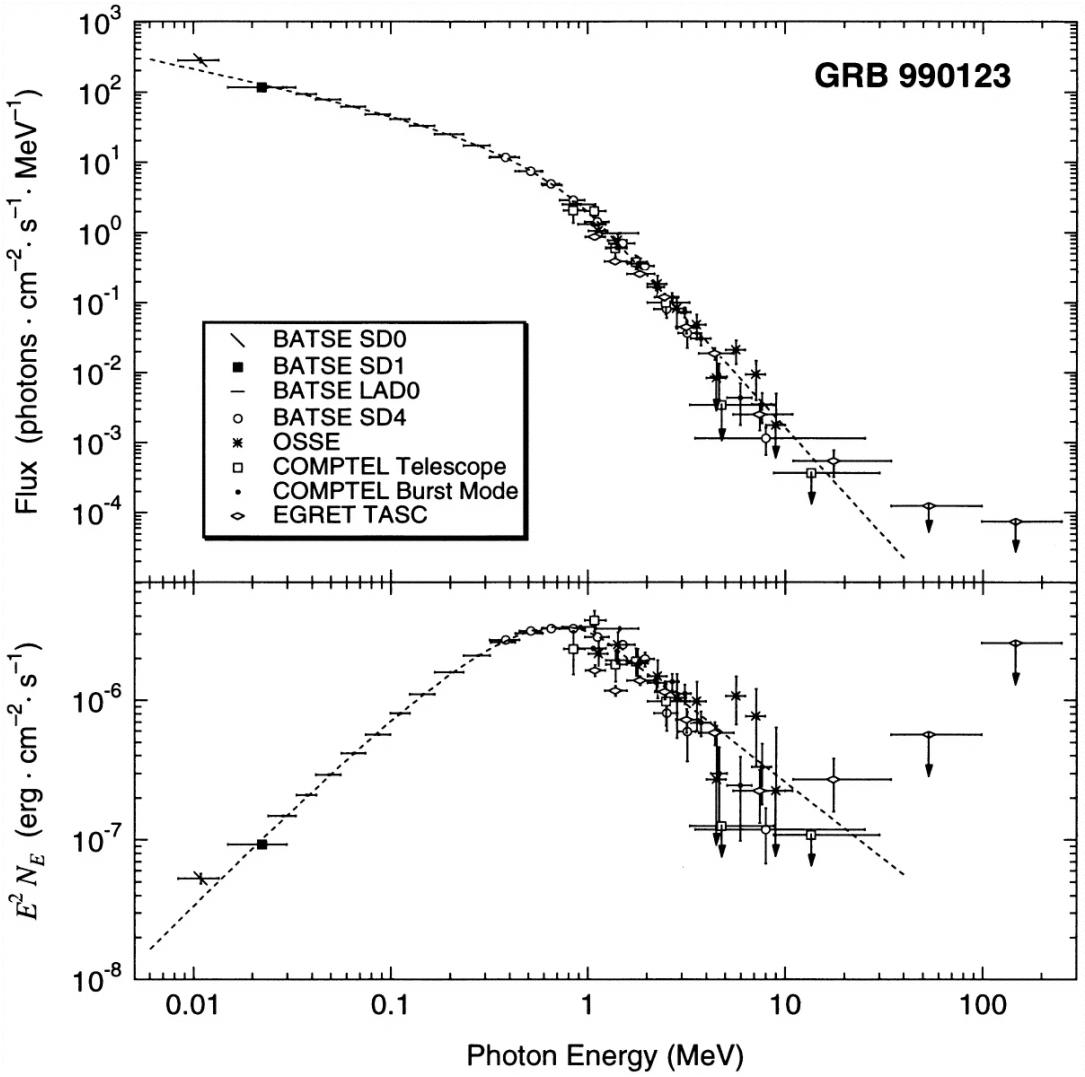


Figure 2.11: GRB990123 spectrum over all instruments on CGRO. Shown is the flux vs. energy and the Spectrum in energy space. The Band-function fits the data points very well over a large range of the data points. From [Bri+99].

the low energy fluence versus the t_{90} duration. GRBs with a shorter duration cluster at higher hardness values than longer GRBs. This led to the assumption, that short and long GRBs have different progenitors.

2.5.3 Progenitors

Long GRBs have been studied for an extended period of time. The longer duration of the burst allows for follow up measurements either on the detecting satellite itself or alerted instruments in space or on ground.

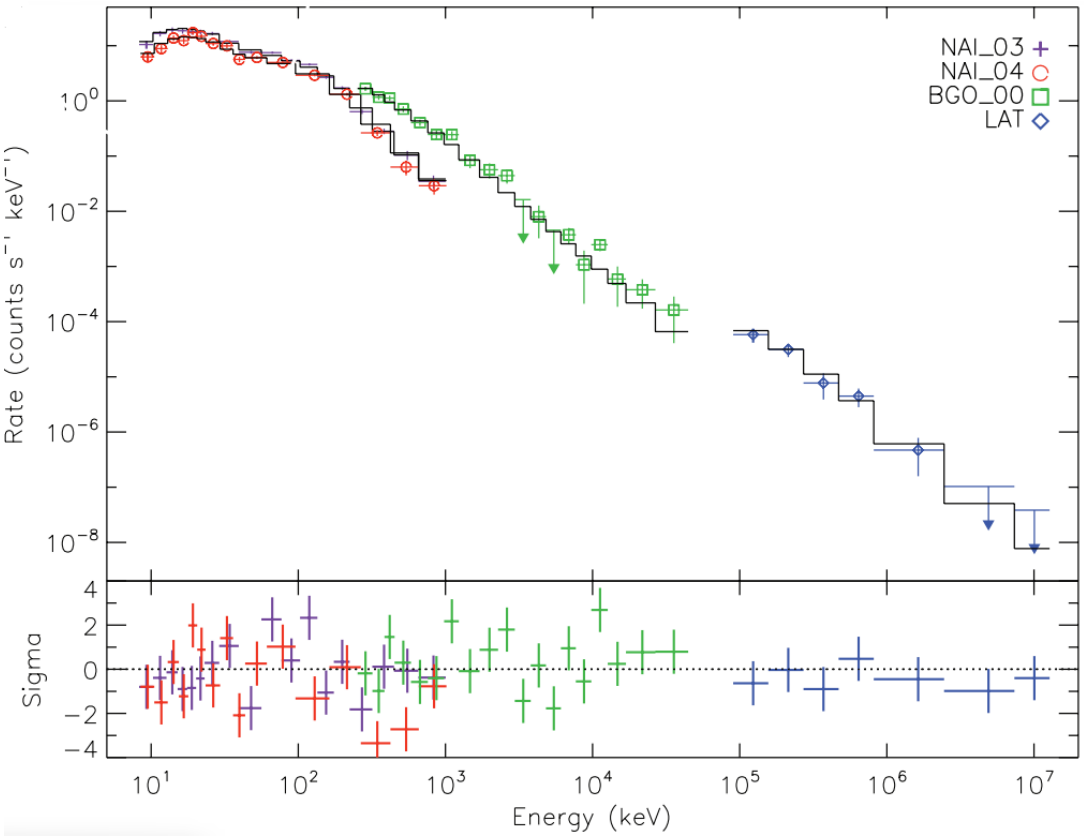


Figure 2.12: Count spectrum of GRB080916C. Given are the spectra measured by the individual detectors between 3.58 to 7.68 s after trigger. The data points have 1σ error bars, whereas upper limits are 2σ. The histograms show the number of counts obtained by folding the photon model through the instrument response models. From [Abd+09].

One example is a very bright long GRB from 1999 observed by BATSE [Bri+99] and BeppoSAX [Pir99] tagged as GRB990123. The Robotic Optical Transient Search Experiment (ROTSE) responded fast enough to take optical data *during* the prompt emission of the burst [Gis+99]. Keck as well as the Hubble Space Telescope measured the afterglow over an extended period of time. These measurements could then be used to determine the total measured flux over a very large energy range from the initial burst to several days after. The HST measurement was also used to characterize the potential host galaxy. Redshift measurements have located the source to a galaxy at $z = 1.6$, offset from the center of the host in a potentially star forming region [Fru+23].
This quickly established a correlation pointing to star forming regions of galaxies with a population of blue stars as host galaxies of long bursts and therefore link at least some long GRBs to supernovas [WB06; Pod13].

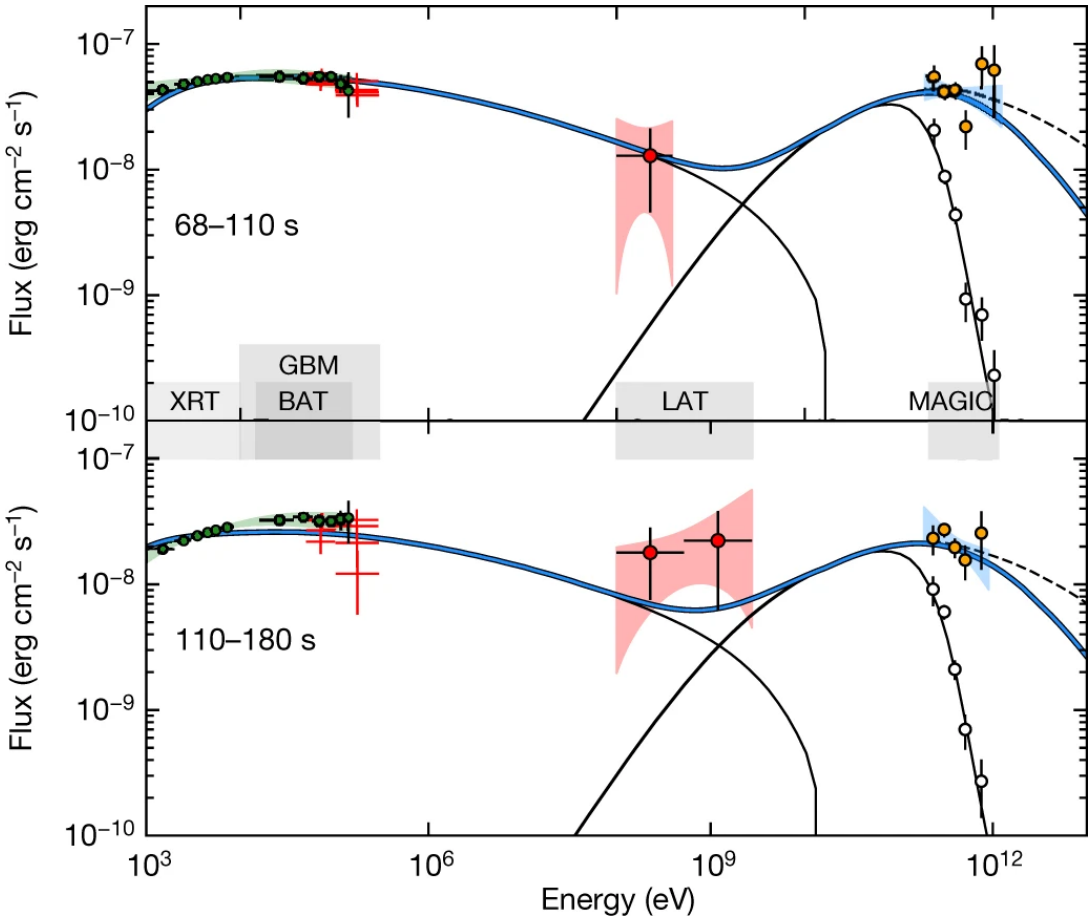


Figure 2.13: Modelling of the broadband spectra of GRB190114C in the time intervals 68–110 s and 110–180 s. The thick blue curve models the multi-band data in the synchrotron and SSC afterglow scenario. The thin solid lines correspond to synchrotron and synchroton self-Compton (SSC) emission (observed spectrum) components. The dashed lines show the SSC component when internal γ–γ opacity is neglected. Empty circles show the observed MAGIC spectrum uncorrected for attenuation caused by the Extragalactic Background Light (EBL). The yellow solid points correspond to the MAGIC measurement after correction. From [Acc+19].

The total amount of energy measured for GRB990123 was found to be in tension with estimates on emissions from spherical shells, which was further hinting at the jet-like character of GRB emissions [CT+99]. In order to model the processes at the source the emission model of GRBs needs to account for the jet-like nature of the emitted gamma-rays.
In order to drive the emission process beyond the initial prompt emission, a central engine providing the necessary energy is needed. The two main models are a black hole (collapsar) [Woo93] or neutron star (magnetar) [Whe+00] formed during a core-

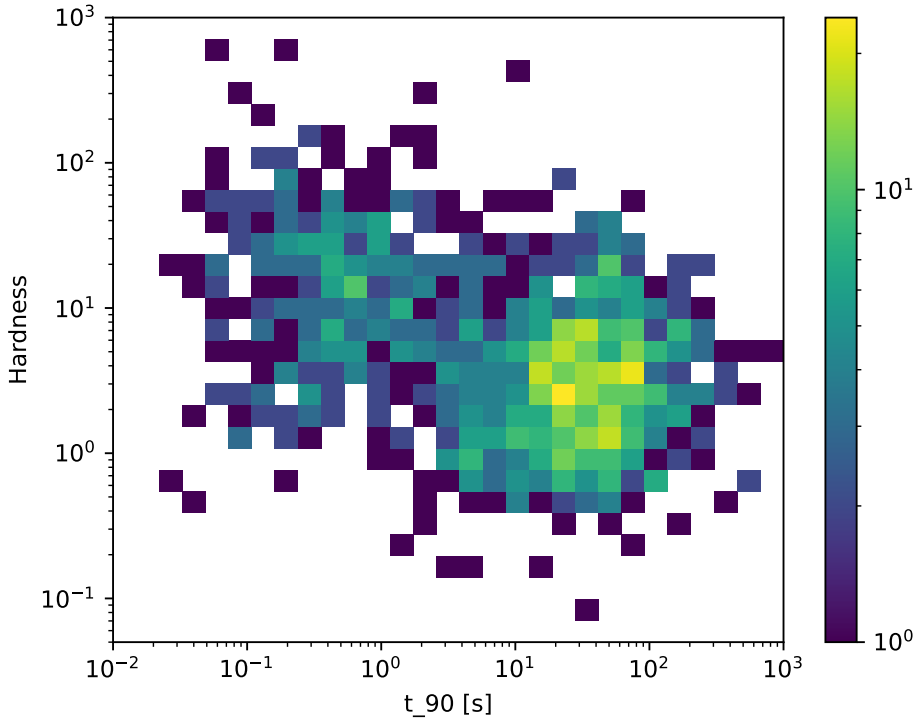


Figure 2.14: Hardness determined by the ratio of the high energetic fluency relative to the low energetic fluency vs. burst duration collected from the BATSE4B catalogue [Mee+98].

collapse supernova.
In the magnetar model, the originating body is a fast rotating newly formed neutron star. During the collapse a proto-neutron star is formed inside the collapsing star and triggers a super-nova explosion. After the explosion, the not yet relativistic magnetic wind forms a bipolar jet. When the neutron star cools down further the increase in magnetization causes the jets to become relativistic and break through the ejecta, creating the prompt GRB emission. Further cool-down of the neutron star and with this further increase of the jet-magnetization results in an exponential decay of the prompt GRB emission, as the acceleration and dissipation process becomes less and less effective.
The collapsar model is very well motivated by the link of long GRBs to regions of active star creation of their host galaxies. This hints at blue supergiants as possible progenitors. During the supernova event the remnants of these stars can form black holes. Some of the expelled stellar material could form an accretion disk with high enough angular momentum to not collapse into the black hole. The energy driving the GRB can be extracted from such a system in two different ways. One possible

energy extraction mechanism could be driven by neutrino-antineutrino annihilation from particles created in the hot disk. According to [Nag18] the annihilation rate has an angular dependency, resulting in jets originating close to the poles of the black hole. Another possibility to generate relativistic jets in this scenario is by tapping the rotational energy of a slowly rotating black hole via the Blandford-Znajek (BZ) effect [BZ77]. Here the charged particles in the accretion disk are accelerated and generate a Poynting-Flux originating in the ergosphere which is collimated towards the poles by a wind of charged particles generated from the rotating accretion disk.
The general consensus is that long GRBs originate from the collapse of massive stars with a central engine driving the emission process. However, there have been observations of plateau-shaped X-ray afterglows similar to observations on long GRBs, which hint at the presence of longer-lived engines similar to long GRBs (e.g., [Cio18]). Studies on short GRBs measured with the Neil Gehrels SWIFT Observatory have shown that the origin of a fraction of these bursts were elliptical galaxies. These galaxies are the result of galaxy-collisions. During the process the majority of the unbound gas is ejected, reducing star creation compared to spiral galaxies. Therefore, this type of galaxies consists in general out of older red stars with low metallicity ([FIH89; GF10]). In addition, the GRB spectrum was found to be less energetic on average [Nak07; Ber11]. This was consistent with models, that predicted short GRBs from massive binaries [NPP92] and neutron star mergers [Eic+89]. This link was finally established beyond reasonable doubt with the measurements of GRB170817, which was observed over all available wavelengths and gravitational waves [Aa17]. The analysts concluded that the observed signature is most consistent with the merger of a neutron star binary, followed by a short GRB emission, confirming the link between short GRBs and the merging of compact objects. Figure 2.15 shows the temporal evolution of the signal.

2.5.4 Emission Processes

The emission process that generates the prompt emission is still an open question. Emission models are constrained by lightcurve and spectral data, which still allow for very different generating processes. One possibility to distinguish GRB emission models is the amount of linear polarization measured from the source photons. Most GRB spectra have their highest flux in the range of 100 - 400 keV [Lev18]. The degree of linear polarization that can be measured is dependent on the field conditions present at the source [Lyu04] as well as the relative angle between polarization axis and line of sight under which the source is observed. Different emission models therefore only allow for distinguishable maximum polarization values. Some older examples are summarized in Table 2.2.
Cannonball models (see e.g. [DDR04]) try to explain GRB emissions from inverse Compton-scattering of ambient photons. The electrons needed to scatter the optical photons into the MeV range are contained in discrete plasmoids called "Cannonballs" generated from abruptly in-falling material of an assumed accretion disk. Each GRB pulse in the light curve would then correspond to one traversing Cannon-

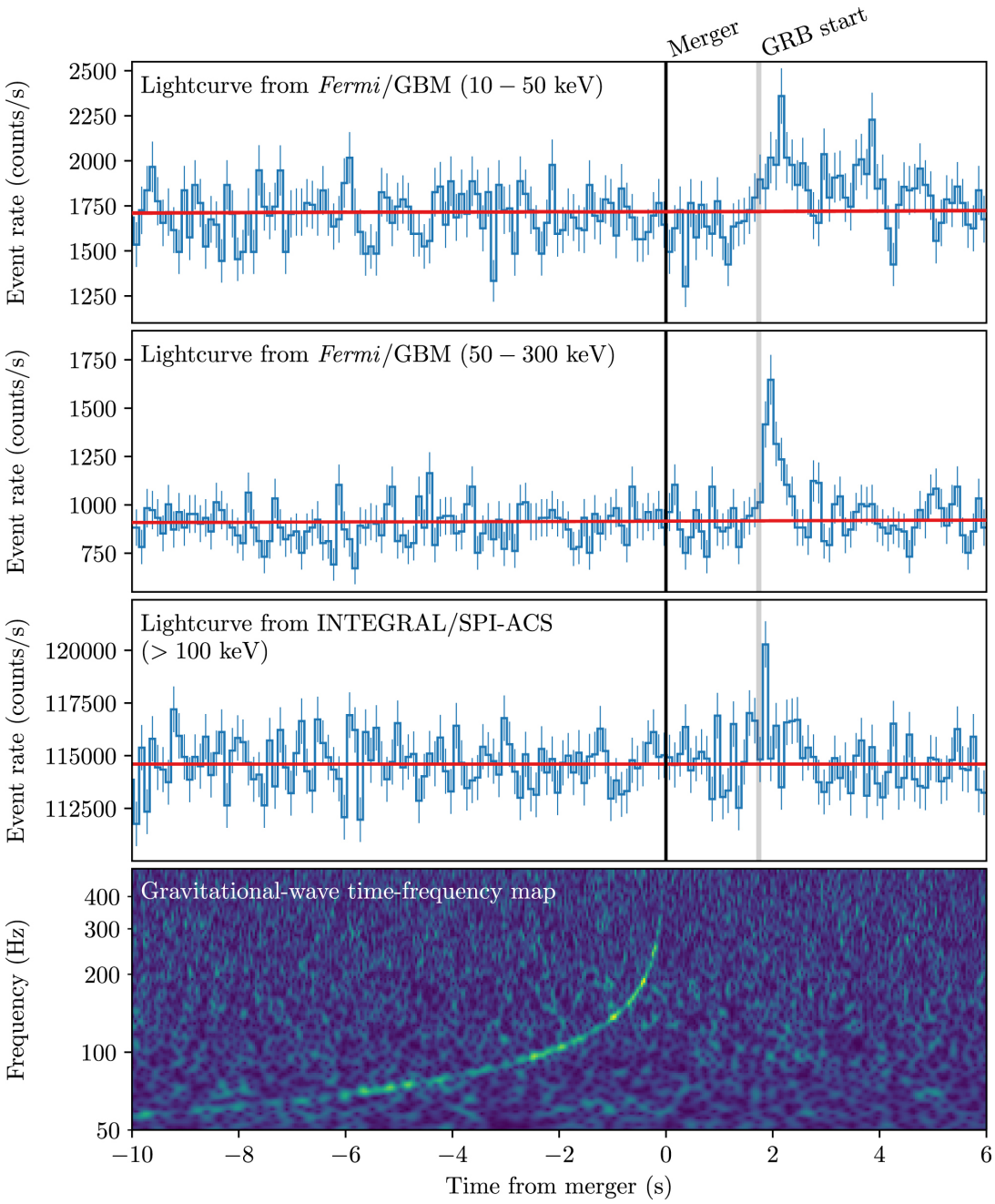


Figure 2.15: Temporal evolution of the Neutron-Star merger signal GRB170817. From [Aa17].

ball. One of the main features that set it apart from other models is the generation of up to 100% linear polarized photons [Lyu04; DDR04]. Figure 2.16 shows a sketch of the model.

Table 2.2: Different emission models for prompt GRB radiation and their maximum achievable polarization. From [Lyu04].

Emission Model	Max. Polarization
Fireball	20%
Electromagnetic	50%
Cannonball	100%

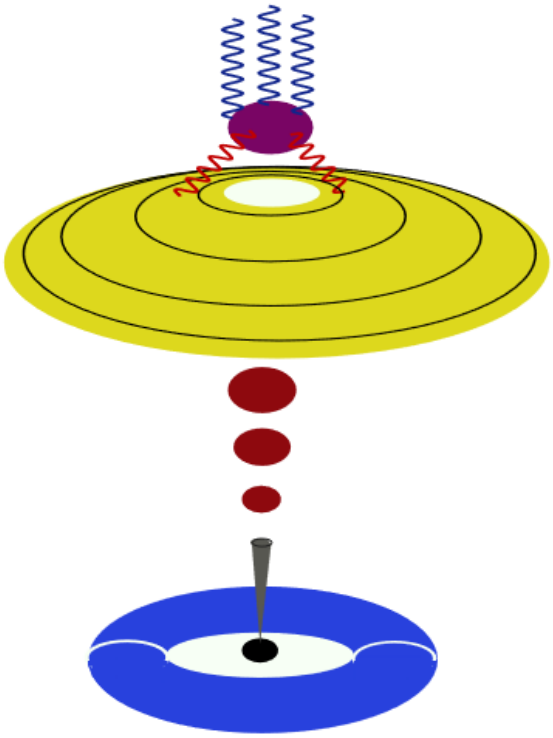


Figure 2.16: Sketch illustrating the Cannonball model. From [Dar05].

The classical Fireball model [Goo86; Pac86] assumes the creation of an opaque photon-lepton-plasma due to a sudden energy release (e.g. explosion) into a compact region of space. The plasma accelerates and transfers its energy to the baryons around it. The actual GRB is then generated by the collision with shells of material ejected from the exploding star (internal shocks). The afterglow is produced by collisions with the surrounding interstellar medium or wind caused by the progenitor (external shock). Observation of high degrees of polarization would require fine-tuned conditions with respect to the speed of the shells. Estimates for a single shell amount to an a maximum of 20% linear polarization [Lyu04]. A sketch is shown in Figure 2.17. Electromagnetic Models (see e.g. [LB02]) assume the presence of either a fast rotating black hole, a proto-neutron star inside a collapsing star or a millisecond mag-

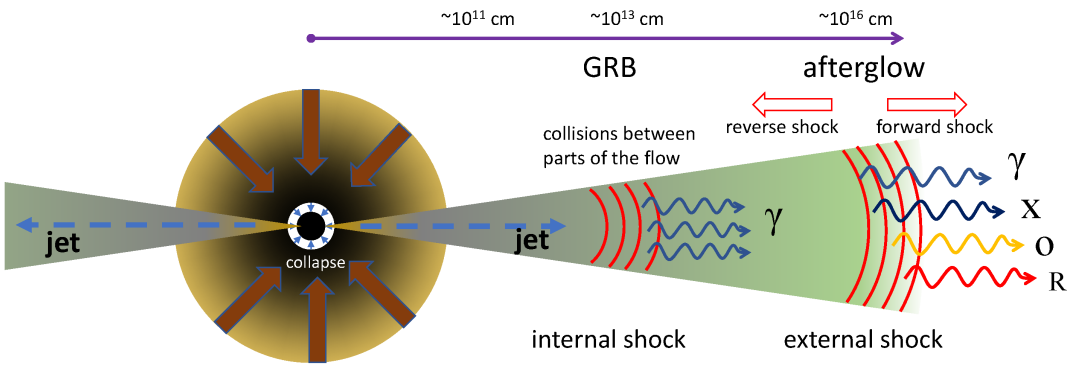


Figure 2.17: Sketch of the stages of the Fireball model. From [DDDR22].

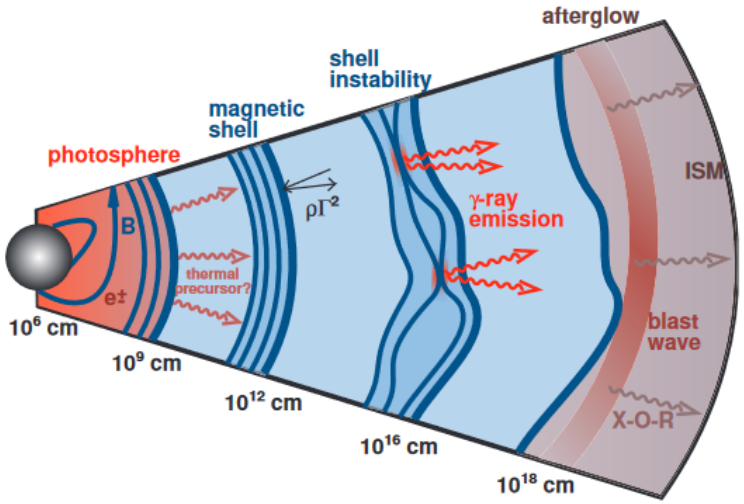


Figure 2.18: Sketch showing the workings of the Magnetohydrodynamic GRB model. From [Lyu04].

netar. The source creates an expanding EM-bubble which forms a twin funnel for the outflow, envisioned as a Poynting flux. Outside of the progenitor, the bubble expands relativistically until it forms an electromagnetic shell. Outside of this cavity, a magnetic shear develops and triggers the emission of Doppler-boosted synchroton emission and emission from accelerated pairs. After the break-out, the afterglow is modelled to originate from synchroton emission and inverse Compton scattering. Since the emission of the prompt GRB emission involves large, ordered fields with a toroidal symmetry, the maximum achievable degree of linear polarization should be about of 50% [Lyu04]. Figure 2.18 shows a sketch of the process.
Modern approaches try to model the properties of the emerging jet and the field conditions around it in more detail. A thorough collection of estimates for the most

prominent subcategories can be found in [GKG21].
Polarization associated with prompt GRB emissions is an under-investigated property of GRBs. It was first observed in 2002 with GRB021206 by [CB03], but doubted in re-analyses from [Wig+04] and [RF04]. In the following years between 2004 and 2016, 10 GRBs with polarization estimates were measured [McC17], equating to a detection rate of less than 1 per year. POLAR [Pro+18], a dedicated GRB polarimeter on the Chinese space station Tiangong2, operated from September 15 2016 to April 1 2017 and was able to add polarization estimates of 16 GRBs during this time [Kol+20]. As the discrimination power of this data-set is very limited, the number of possible emission models and scenarios is still very large, motivating the development of more missions and design concepts capable of measuring polarization in GRBs (see e.g.: POLAR2 [Pro+23], e-ASTROGAM [De +17a], AMEGO [McE+19]).

2.6 Fundamental Physics

Gamma-rays open up windows into related to fundamental symmetries of the universe. The main focus has been on dark matter searches. While there is strong evidence for dark matter on all larger scales in the universe from rotational curves of galaxies [RFT80] to galaxy clusters [Clo+06] to the cosmic microwave background [Pla+16], the constituents of dark matter are still unknown. Whereas the dark matter particles are not measurable with gamma-rays directly, their proposed decay or annihilation into standard model particles could produce a gamma-ray signature via processes like

$$\chi\chi \rightarrow \gamma\gamma \tag{2.20}$$

$$\chi\chi \rightarrow \pi^0\pi^0 \tag{2.21}$$

or

$$\chi\chi \rightarrow \phi\phi, \phi \rightarrow e^+e^- \tag{2.22}$$

to only show a few . The first channel would provide a potential "smoking gun" signature, as the spectrum of the gamma-rays would take a line shape with the centroid at the mass of dark matter particle. In the second case, the annihilation product is a pair of neutral pions, which would result in a box-shaped annihilation spectrum. The third cases describes the annihilation into a mediator particle ϕ that decays into an electron-positron pair. Here the annihilation spectrum will be boosted and posses a box shape similar to the pion case [BGW17]. The primary flux of the annihilation signal depends on the line-of-sight integrated dark matter density distribution J, the velocity averaged self-annihilation cross section $\langle\sigma v\rangle$ and the dark matter mass. The parameter a is, depending on the nature of the dark matter particle, either 1/2 (Majorana particle) or 1/4 (Dirac particle). It has the form of (see e.g. [Cir+11]:

$$\frac{d\Phi}{dEd\Omega} = \frac{a\langle\sigma v\rangle J}{4\pi m_\chi^2}\frac{dN_\gamma}{dE_\gamma} \tag{2.23}$$

Indirect searches are not limited to decay and annihilation signals of particle dark matter. Hawking radiation from evaporating primordial black holes would result in a temperature dependent emission spectrum [Car+10] that will be subject of surveys of the COSI mission. Indirect searches for anti-matter try to shed light into the matter-antimatter asymmetry at the beginning of the universe. They look for an excess of gamma-rays at around 100 MeV, which could be the fingerprint of proton-antiproton annihilation and then use the data to constrain possible models (see e.g. [Dol02]). Constraints on the anti-matter fraction can be derived from galaxy cluster collisions (see [Ste08]).

Chapter 3

Space-based Gamma-ray Astronomy

3.1 Interaction of Photons with Matter

Photons interact with matter via various energy dependent interaction channels. At low energies, the dominant interaction is photoelectric absorption. At high energies, the creation of electron-positron pairs is dominant. In between, Compton scattering is the main interaction channel. Figure 3.1 shows the regime of the three interactions with respect to the atomic number. Low-Z material have a larger energy range in which Compton scatter is effective, making these materials a prime candidate for scattering detectors in Compton telescopes, while high-Z materials cover more area for Photoelectric absorption as well as Pair creation. This makes them suitable as either conversion or absorption material in Pair telescopes or calorimeters in Compton telescopes.
Two detector materials found in recent proposals for Compton/Pair Telescopes are Silicon and Cesium-Iodine (see [De +17b], [McE+19]). The attenuation coefficients vs. energy are depicted in Figures 3.2
For Silicon (Z=14), the attenuation for Photo-effect and Compton scatter are equal at around 70 keV, while Pair creation starts to dominate for energies larger than 10 MeV as shown in the top Fig. 3.2, giving a range of 9.93 MeV in which the dominant interaction is Compton scatter.
In the case of CsI (effective Z of 54), Photo-effect and Compton scatter reach an equilibrium at 400 keV, while Pair creation already starts to dominate at 7 MeV (see Fig. 3.2 bottom), reducing the usable energy range for Compton scatter to 6.4 MeV.
In the following sections, the different interaction modes and their description are discussed in more detail.

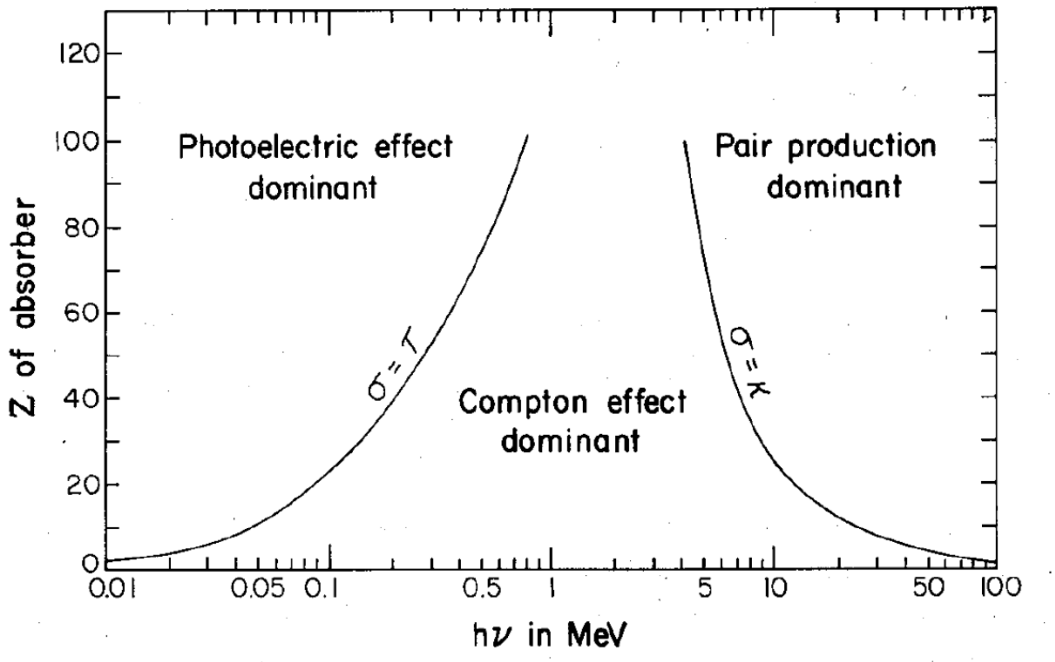


Figure 3.1: Equiprobability lines for Photoelectric effect, Compton Scatter and Pair Creation for various atomic numbers and energies. $\sigma = \tau$ denotes equal cross sections for Photoelectric absorption and Compton scattering, while $\sigma = \kappa$ marks equal cross section for Compton scatter and Pair creation. From [EE55].

3.1.1 Photo Absorption

The photoelectric effect describes the complete absorption of a photon by an atom. The photon energy is transferred to an electron situated usually in one of the inner shells. If the energy of the photon exceeds the binding energy of the electron the electron is kicked out of its shell with energy $E = E_\gamma - E_{bind}$. The now free electron starts to ionize surrounding atoms generating secondary electrons according to its kinetic energy. The cross section of photo absorption strongly depends on Z, and steeply drops with energy with distinct absorption edges at low energies corresponding to binding energies of the respective shell.

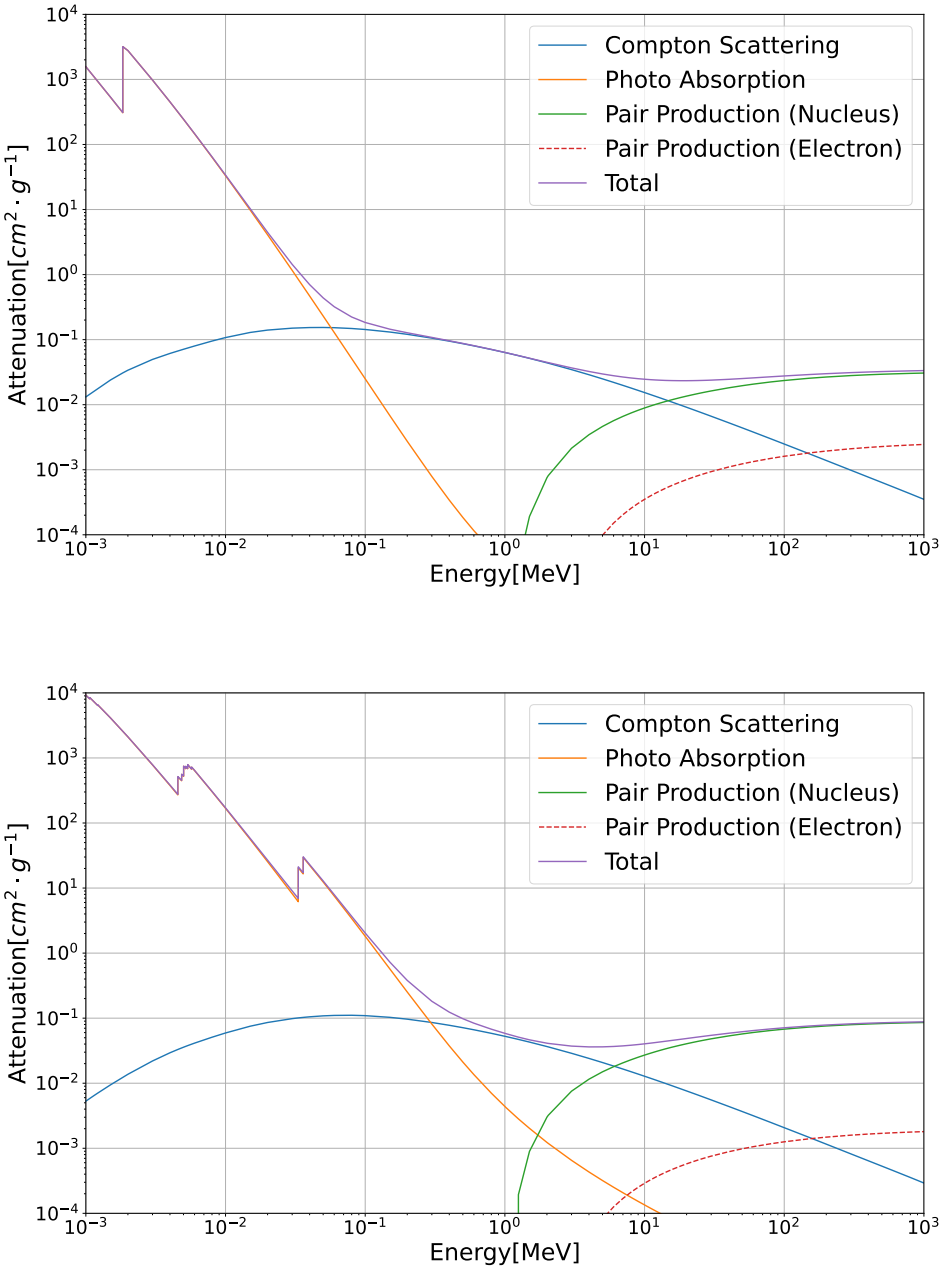


Figure 3.2: Attenuation coefficient vs. Energy for Silicon ($Z = 14$, top) and Cesium Iodine ($Z_{eff} = 54$, bottom). Generated with [Ber+09].

3.1.2 Compton Scattering

Free Electron Approximation

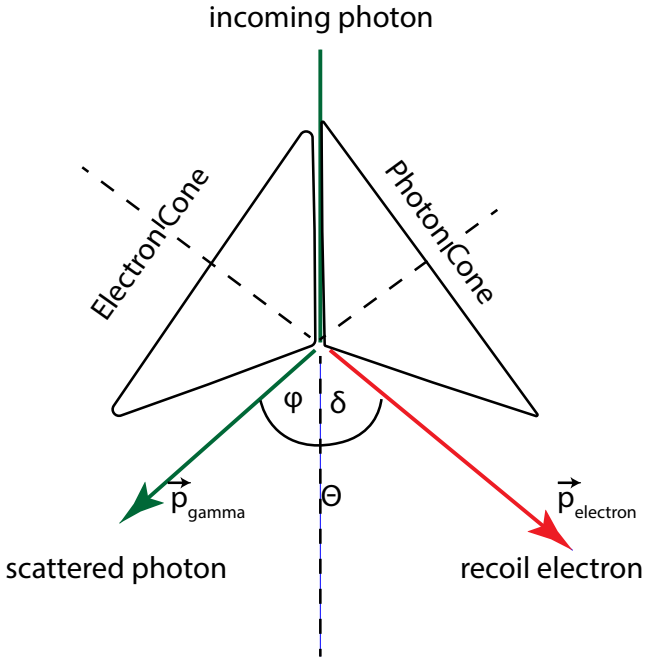


Figure 3.3: Compton-scattering with the used notation. The photon and electron scatter cones represent all possible original directions of the photon if the direction of the recoil electron respectively of the incident photon could not be measured.

At sufficiently high photon energies, the binding energy and the momentum of the electron can be neglected. One can consider it as a free electron at rest. Energy and momentum conservation can then be written as follows:

$$\begin{aligned} E_i &= E_\gamma + E_e^{rel} \\ \vec{p}_i &= \vec{p}_\gamma + \vec{p}_e \end{aligned} \tag{3.1}$$

Applying the relativistic energy-momentum relation together with $E_\gamma = p_\gamma \cdot c$ yields

$$\begin{aligned} \vec{p_i} = \frac{\sqrt{E_e^2 + 2 \cdot m_e c^2 \cdot E_e}\vec{p}_e + E_\gamma \vec{p_\gamma}}{E_e + E_\gamma} \\ E_i = E_e + E_\gamma \end{aligned} \tag{3.2}$$

for the initial direction and energy of the photon. For the photon scatter angle then follows:

$$\cos\varphi = 1 - \frac{m_e c^2}{E_\gamma} + \frac{m_e c^2}{E_\gamma + E_e} \tag{3.3}$$

which is also called Compton's formula. From this, one can constrain the energy range of the scattered gamma-photon to

$$\frac{m_e c^2 \cdot E_i}{2E_i + m_e c^2} < E_\gamma < E_i. \tag{3.4}$$

This implies for the energy transferred to the recoil electron:

$$0 < E_e < \frac{2E_i^2}{2E_i + m_e c^2}. \tag{3.5}$$

This is corresponding to back scattering (maximum transferred energy) and forward scattering (very little transferred energy).
For the recoil electron and the total scatter angle then follow

$$\cos\delta = \frac{E_e(E_i + m_e c^2)}{E_i\sqrt{E_e^2 + 2 \cdot E_e \cdot m_e c^2}} \tag{3.6}$$

$$\cos\theta = \frac{E_e(E_\gamma - m_e c^2)}{E_\gamma\sqrt{E_{e^2+2E_e \cdot m_e c^2}}} \tag{3.7}$$

from the geometry of the process.
The differential cross section was found to be

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2}\left(\frac{E_\gamma}{E_i}\right)^2 \cdot \left(\frac{E_\gamma}{E_i} + \frac{E_i}{E_\gamma} - 2 \cdot \sin\varphi^2 \cdot cos^2\eta\right) \tag{3.8}$$

by Klein and Nishina [KN29], where r_e is the classical electron radius and χ is the polar scatter angle. The cross section assumes an unbound electron. The cross section is thereby only an approximation since the electrons in, e.g. a calorimeter crystals are usually bound and moving. From Figure 3.4 one can see that the average Compton scatter angle decreases for higher energies, the path of the photon should thereby show more and more resemblance to an high energy electron track. For randomly polarized photons, averaging over the $\cos\chi^2$ yields 1. For linear polarization, the cross-section is maximal for φ perpendicular to χ.

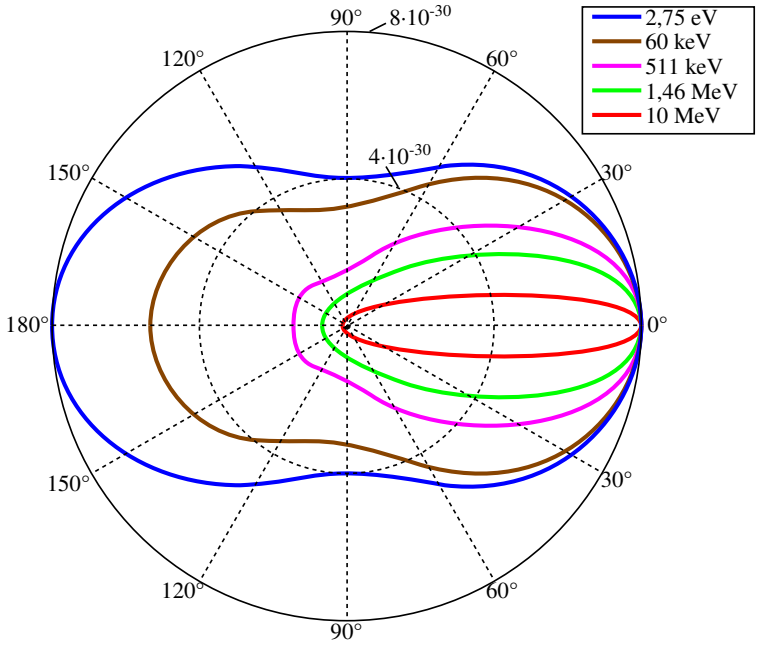


Figure 3.4: Klein-Nishina cross section as function of the photon scatter angle φ. For higher energies the area of possible scatter angles shrinks. From [Com23].

As a result of this, the polarization information survives in the distribution of the polar projection of the scatter angle. An example polarigram is given in Fig. 3.5. P_{min} is the minimum of the fitted oscillation pattern where P_{max} is the maximum. P_0 is the offset from the baseline and A is the amplitude of the fitted cosine oscillation.
The capability to measure the polarization of incident photons is given by the response to 100% linearly polarized photons:

$$\mu_{100} = \frac{P_{max} - P_{min}}{P_{max} + P_{min}} = \frac{A}{b} \tag{3.9}$$

Compton scattering of bound electrons

The previous discussion assumes scattering of a photon on an electron at rest. Since electrons in a detector medium are in general bound in shells and not in rest, the measured Compton scatters in the detector will deviate from the ideal modelling. The measured broadening of a spectrum obtained via Compton scattering was first interpreted in 1929 by [DM29] as Doppler broadening induced by the electron's velocity. The momentum distribution of the electron also leads to the necessity of modifications to the cross section. A formulation of the modified cross section was proposed by Ribberfors and Berggren in 1981 [RB82]:

$$\left(\frac{d\varsigma}{d\Omega}\right)_{\text{incoh}} = \left(\frac{d\varsigma}{d\Omega}\right)_{KN} \cdot S(\omega, \theta, Z) \tag{3.10}$$

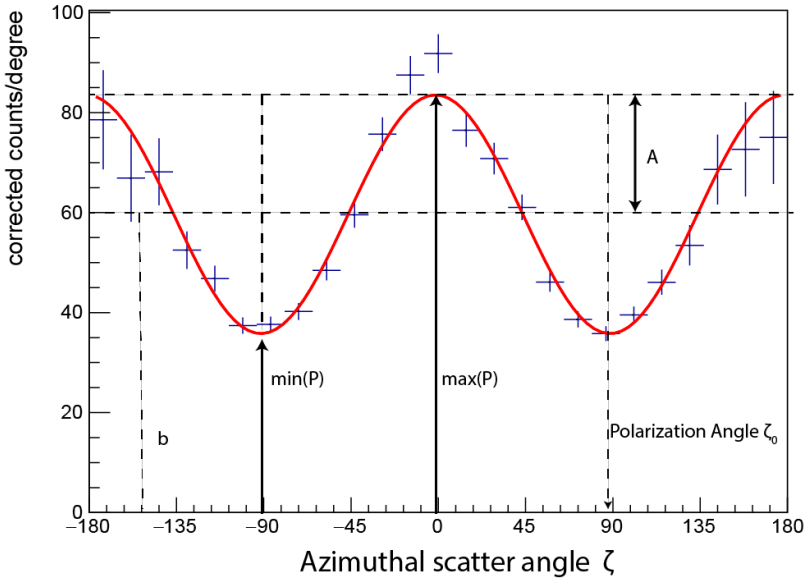


Figure 3.5: Simulated polarization signature for a 100% polarized photon beam. The error bars reflect the bin range in x and statistical uncertainties in y. The distribution has the form $P(\zeta) = b + A\cos(2(\zeta - \zeta_0 + \pi/2))$. Generated with [ZAS06].

Here $S(\omega, \theta, Z)$ is the incoherent-scattering function, which depends on the scatter angle θ, the energy of the photon ω and the atomic number Z. Applying Waller-Hartree's theory and the Hartree-Fock approximation for closed shells the authors give

$$S(\omega, \theta, Z) = Z - \sum_{m,n} |\langle \psi_m | exp(i\vec{q} \cdot \vec{r}) | \psi_n \rangle|^2 \tag{3.11}$$

as the incoherent scattering function, with ψ_m referring to the occupied orbit. In the relevant relativistic case, S can be expressed as

$$S(\omega, \theta, Z) = \sum_{i}^{occ} n_i(P_{i,max}) \tag{3.12}$$

with

$$n_i(P_{i,max}) = \int_{-\infty}^{p_{i,max}} dp_z J_i(p_z) \tag{3.13}$$

The function $J(p_z)$ is called the Compton-profile and is proportional to the z-projection of the electron momentum density of the scattered electron. n_i therefore gives the electron momentum density of the i-th electron in the shell.
This leads to shifts in scattering probability (the cross section for lower energies is increased) as well as changes in the scatter angle distribution (small and large scatter angles are more suppressed at lower energies). Since the electron possesses a

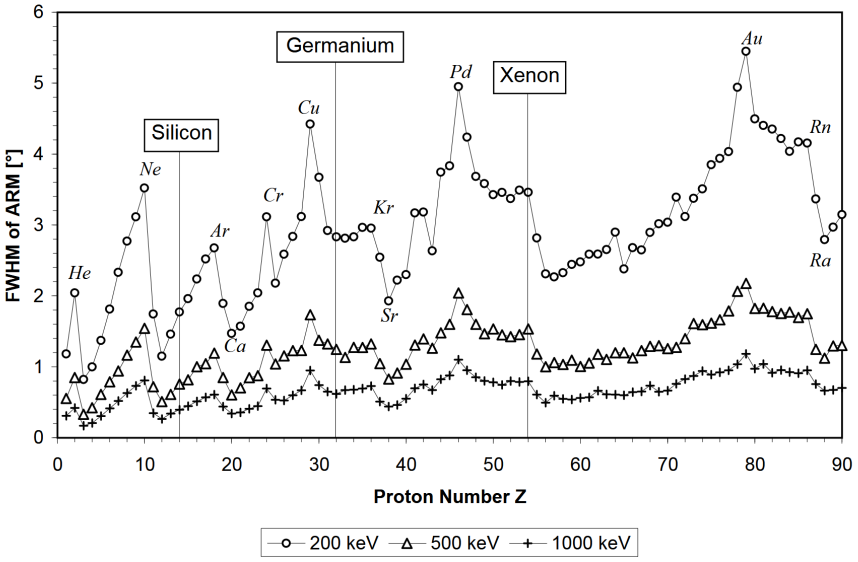


Figure 3.6: Best angular resolution vs. Z and energy assuming a perfect position and energy resolution. From [ZK03].

momentum, the measured scatter angle and consequently the reconstructed energy deposited by the photon differ from the ones of the unbound case (Doppler broadening). The latter has direct consequences for the angular resolution which introduces a fundamental limit on top of the position resolution in the detector and the capability to measure the energy after scattering correctly. Zoglauer and Kanbach discuss in [ZK03] the effect of Doppler broadening in different materials on the angular resolution. They define the angular resolution measure (ARM) as the difference between kinematic scatter angle $\bar{\varphi}$ and geometric scatter angle obtained from the positions of the hits φ_{geo} and the given source direction

$$ARM = \bar{\varphi} - \varphi_{geo} \tag{3.14}$$

One Full Width at Half Maximum (FWHM) is defined as the angular resolution at the given energy. Figure 3.6 shows the best achievable angular resolution as function of atomic number Z and energy.
The best angular resolution varies for Z, showing a general tendency to better resolution at lower Z, but also showing a correlation between the electron configuration of the respective valence shell and angular resolution. The more populated the valence shell gets, the worse the resolution becomes. The choice of detector material, therefore, has a profound impact on the performance the instrument can achieve even under ideal circumstances. Doppler broadening is most pronounced at energies below 500 keV, where the energy resolution otherwise limits the angular resolution. At high energies, Doppler broadening limits the angular resolution even for detectors with excellent position resolution.

3.1.3 Pair Creation

For energies larger than $2m_ec^2 = 1022$ keV the incident photon has a probability to convert into an electron-positron pair in the presence of a nucleus' field which has to absorb a portion of the photon's momentum. If the photon energy exceeds $4m_ec^2 = 2044$ keV the conversion can also happen in the electron shell. The nuclear recoil is not measurable, but the recoiled electron can leave a track in the detector. One could therefore observe three tracks from one pair creation event("triplet production").

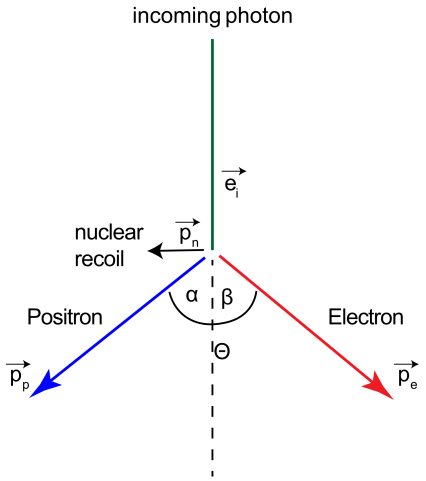


Figure 3.7: Schematic of a pair creation event. The photon converts into an electron-positron pair. The positron later annihilates with an electron into two photons with $E_\gamma = 511$ keV. α, β denote the positron and electron portion of the opening angle Θ.

The processes can be described by

$$\begin{aligned} E_{initial} &= E_{electron} + E_{positron} + E_{recoil} + 2 \cdot m_e c^2 \\ \vec{p}_{initial} &= \vec{p}_{electron} + \vec{p}_{positron} + \vec{p}_{recoil} \end{aligned} \tag{3.15}$$

The energy transferred to the nucleus is very low as the transferred momentum $\vec{p_n}$ to the nucleus is around 511 keV/c. Neglecting the transferred momentum by setting it to zero simplifies the geometry. The initial unit direction of the incoming photon $\vec{e_i}$ becomes then

$$\vec{e_i} \approx \frac{E_e\vec{e_e} + E_p\vec{e_p}}{E_e + E_p}, \tag{3.16}$$

where $\vec{e_e}$ and $\vec{e_p}$ are the unit momentum vectors of electron and positron. If one is not able to measure the energy of the electron and positron one can still try to estimate the origin of the photon via

$$\vec{e_i} \approx \vec{e}_{electron} + \vec{e}_{positron}. \tag{3.17}$$

3.2 Low- and Medium-energy Gamma-ray Telescopes

The purpose of a telescope in the scientific sense is to

1. localize the origin of a source (imaging)
2. reconstruct the energy of the photon (spectroscopy)
3. track the temporal evolution of the flux of a source

For optical photons this can be achieved by fixing the pointing of the instrument while recording photons and matching the measured photons to the observed region of the sky, reconstructing the wavelength (color) and with that the energy of the collected source photons and control timestamps on individual exposures in order to calculate light curves. In general, the same holds true for gamma-ray telescopes. In order to localize the source, the pointing of the instrument has to be known with good accuracy. The satellite is moving in an orbit around Earth, necessitating regular updates of the pointing. The quality of the image, spectral and temporal information obtained from an observation is strongly depending on the design of the telescope. Each design has its own set of advantages and disadvantages which are discussed below.

3.2.1 Coded Mask Telescopes

Coded mask telescopes use a passive occulter divided into opaque and transparent segments. Photons emitted from a point-source on the sky reach the mask as parallel beam and create an incident angle dependent shadow pattern. Position sensitive detectors measure the resulting intensity pattern (shadowgram). A second source would create a different pattern, that can be disentangled (Figure 3.8 left). Background photons from activation and diffuse emission leave a constant pattern in the shadowgram. To remove its imprint in the raw data, one can take shadowgrams of the source at slight angles. The source pattern will differ in each raw pattern, but the underlying background pattern will stay constant, enabling the background to be subtracted. The position of the point source can be reconstructed by applying deconvolution algorithms (see e.g. [CGG16] and [Car+87] for details). An example measurement is depicted in Figure 3.8 right.
Random patterns create noise in the reconstructed image due to small terms arising from Fourier-transform. Possible solutions are Uniformly Redundant Arrays (URA) [FC78]. Coded mask telescopes can achieve a very good spatial resolution. The achievable resolution depends on the geometry of the detector. For an on-axis source, one can obtain an Angular Resolution(AR) of approximately

$$AR \approx arctan\left(\frac{\sqrt{m^2 + d^2}}{H}\right), \tag{3.18}$$

where m is the size of the individual mask element, d is the pixel size of the detector and H is the distance between mask and detector array [GG22]. The pixel size is

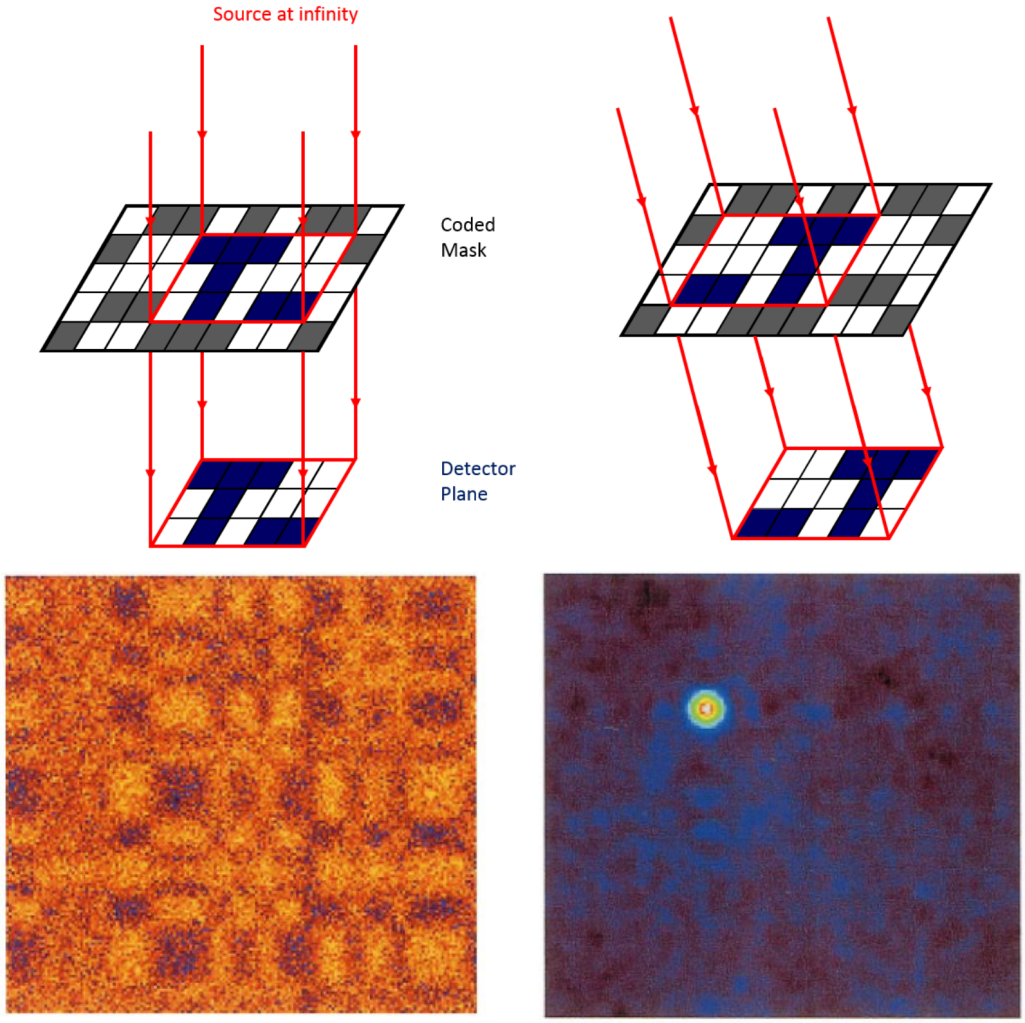


Figure 3.8: Working principle of a coded mask. Top: two sources creates shadow patterns on the pixel detector underneath the mask. Due to their different incident angles, the resulting pattern differs. From [GG22]. Bottom: an example of a measurement of Nova Persei with SIGMA/GRANAT in the 75 to 120 keV band. The left panel shows the measured shadowgram after pixel correction. The right panel shows the reconstructed image after deconvolution. From [Bou+01].

depending on the energy of the of the photons. At energies larger than a few hundred keV, the photo-effect is not sufficient to absorb the photon in one pixel. In order to contain photons that are undergoing Compton scatter, or potentially pair creation, in one pixel, the detector mass and with that the pixel size has to be increased. This has a negative impact on the potential resolution.
A major drawback of the coded mask approach is its usability up to fixed energies. In order to create shadow patterns, the mask must contain transparent and opaque

parts to the incoming γ−rays. The mean free path in matter is depending on the energy of the particle, resulting in the necessity to build a sufficiently massive masks in order to provide enough stopping power. Another problem arises from the low fluxes of high energetic astrophysical gamma-sources. In order for a shadow pattern to emerge, a sufficiently large number of source photons on top of background is necessary. This results in very long exposure times at higher energies due to photon starvation. Examples of space born gamma-ray detectors using coded mask instruments are the Burst-Alert-Telescope (BAT) on the Neil Gehrels SWIFT observatory [Bur+05] and IBIS and SPI on INTEGRAL[Win+03a]. They are also proposed for the X-Gamma rays Imaging Spectrometer (XGIS) on the THESEUS mission proposal [Ama+18].

3.2.2 Compton Telescopes

Compton telescopes make use of Compton-scattering to derive imaging and spectral information. There are, in general, three design options with different implications on the reconstruction approach. In order to reconstruct the origin of a given photon, one has to reconstruct the sequence in which the set of hits of the incident photon (event) was generated. This can be done in two ways. One can either measure the sequence directly by assigning each hit a time-stamp with sufficient accuracy (Time-of-Flight method, short: ToF) [SGD82; Sch+93] or exploit the kinematic properties of Compton scattering in order to deduce the correct sequence [Obe+00; BJ00]. The ToF method has very high demands on the read-out electronics and puts strong constraints on the choice of detector material. The detector material has to possess sufficiently fast rise and decay times, while the read-out has to be able to log the data with accurate and sufficiently precise time stamps. The most common approach uses a scattering detector D1 made of a material with low atomic number and an absorbing detector D2 made of high Z material placed with sufficient spatial separation to obtain distinguishable time stamps between the hits. This allows for a straight forward sequence reconstruction and good suppression of upward moving particles at the cost of a large distance between the individual detectors and with that a narrow field of view and lower effective area per mass. The first Compton telescope COMPTEL in space (see Section 3.4.1) used the ToF method. The kinematic approach either uses a segmented scattering detector or a monolithic detector volume with resolving power in x, y and z. Examples for segmented detectors are stacks of Germanium strip detectors, as well as double sided silicon strip detectors, or imaging solid state calorimeters (see e.g. IMCAL in [Orl+22]). Examples for monolithic detectors are liquid Xenon TPCs (see e.g. [Apr+08]). Examples for two-site and multi-site topologies are given in Figure 3.9.
Since the order of the hits is unknown, the reconstruction has to test all possible combinations to determine the correct sequence, a task scaling with $N!$, where N is the number of hits in the given event. It also allows for more reconstruction errors, because of the abundance of possible sequences. Segmenting D1 as well as a monolithic detector design lead to a more compact design with better field of view and

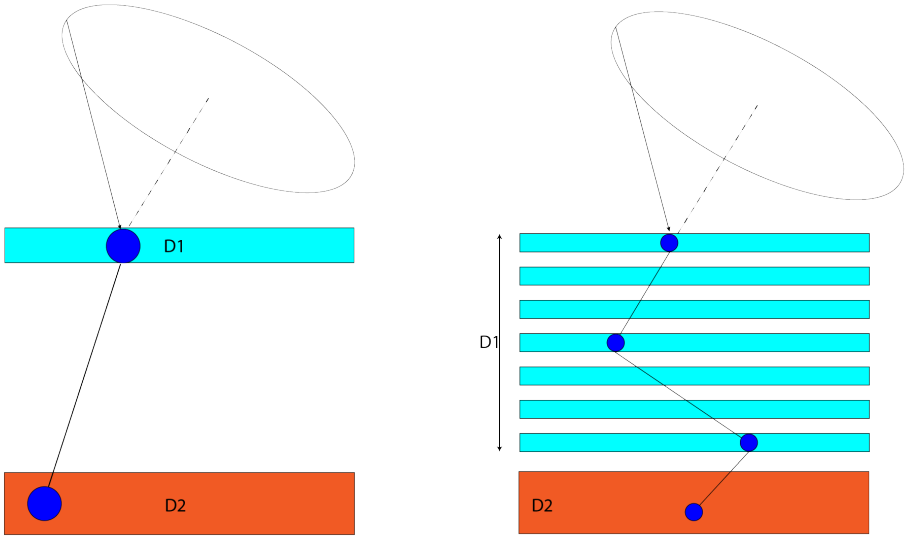


Figure 3.9: Untracked Compton events in two possible configurations of Compton telescopes. The sketch on the left illustrates a scatterer-absorber approach that uses the time of flight between 2-site events for reconstruction. The distance between D1 and D2 is dictated by the choice of detector material. The design on the right uses a segmented scatter detector to allow for more hits in a smaller volume, but has to rely on the kinematics to reconstruct the sequence. Both designs are only able to reconstruct the origin up to an event circle.

mass efficiency than designs that use the ToF method. However, the kinematic approach does not offer direct access to the sequence. One therefore needs to apply a computationally expensive reconstruction on the event data. With no direct access to the sequence, one has to also rely on an accurate sequence reconstruction to suppress upward moving particles originating from background sources.
Imaging makes use of the first two interactions inside of the event. Since the kinematic scatter angle is only determined to an arc (see Equation 3.3) one needs multiple events to determine the origin of the photon. A direct approach is back-projection of the scatter cones to a flat plane called image space illustrated in Figure 3.10. The more events from a source are correctly reconstructed, the more the source position is amplified as the crossing point of the cones. A major problem of this approach is the presence of incorrectly reconstructed events as well as incompletely absorbed photons which can lead to source confusion.
Another possibility is to transform the available information into the Compton Data Space (CDS) [Die+92; Sch+93]. The three important properties are the measured energy E_m, the photon scatter angle $\bar{\varphi}$ derived from the kinematics and the direction of the scattered photon. Note that ψ and χ are calculated in the detector frame and therefore encode the geometrically derived scatter angle φ_{geo} and E_m . One can therefore define a data space with dimensions $\bar{\varphi}$, and the polar as well as the azimuthal

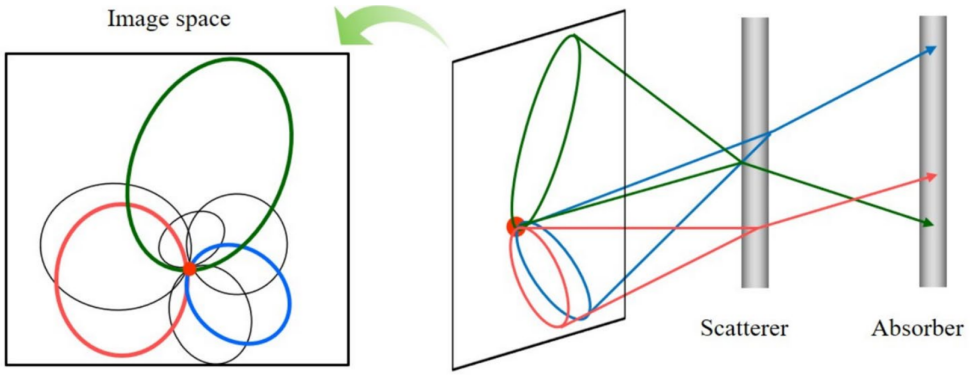


Figure 3.10: The source of a photon from a Compton event can only restricted to a ring touching the source position. When overlaying the projection of multiple Compton cones from the same source into the image space, the source position can be estimated from the crossing point of the event circles. From [KL24].

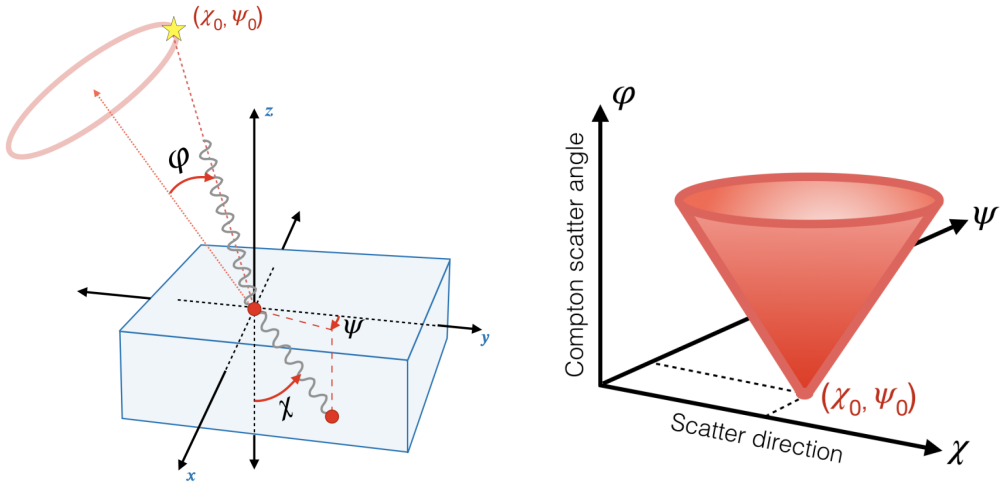


Figure 3.11: Left: Schematic of a two hit Compton event for a source located at (χ_0, ψ_0). The Compton scatter angle φ is calculated from the kinematics using Compton's formula. The polar and azimuthal angles of the scattered gamma ray direction χ and ψ are determined in the detector coordinate system. Right: The PSF in the resulting CDS is a cone with apex at (χ_0, ψ_0) with an opening angle of 90 degrees. From [KTK24].

angle of the scatter direction, usually named χ and ψ, respectively. For every energy, one can therefore define a cone with base at the point source location (χ_0, ψ_0) and opening angle of 90°, which represents the Point-Spread-Function (PSF)[1] of the tele-

[1]The point spread function describes the energy-dependant response of the instrument to a point source. The imaged point source appears as an extended Airy-disk in the image of an optical telescope

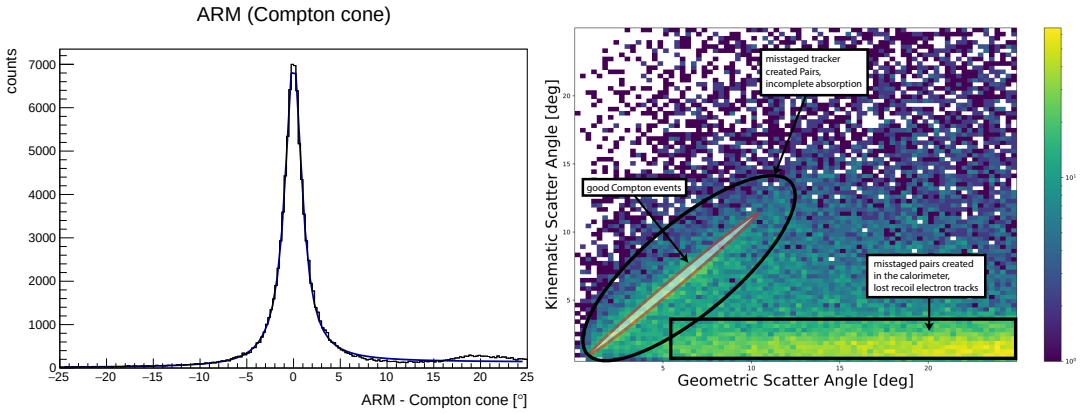


Figure 3.12: Left: ARM distribution for a 1 MeV on-axis source using the e-ASTROGAM mass model. The correctly reconstructed events form a peak at zero with very little features in the tails. Right: $\bar{\varphi}$ vs φ_{geo} 2D histogram for a 10 MeV on-axis source. The correctly reconstructed events form a diagonal, halo events hint at missclassified events and events with incomplete absorption. The band at low kinematic scatter angles is populated by pair events created in the absorber detector and misreconstructed events.

scope at the given energy. Properly reconstructed events from the same point source with the same energy are single points in the CDS that align along their respective PSF-Cone. Sufficiently separated point sources form individual PSF-Cones, allowing for source separation. The CDS is connected to the event circles in image space via the probability of a specific region to contribute to a bin in CDS. Each bin in CDS is represented by a back-projected circle. The response density along φ slices is correlated to the Klein-Nishina cross section. Since highly energetic photons generate smaller scattering angles, they preferentially populate the lower φ values and vice versa. The photon polarization is also encoded in the CDS cone. The modulation defined in equation 3.8 can be expressed in detector coordinates and leaves an imprint in (χ, ψ).

The angular resolution of a Compton Telescope is commonly given by the Angular Resolution Measure (ARM). It is defined as the difference between geometrical scatter angle derived from the hit locations in the detector system and the kinematic scatter angle calculated from Compton's formula (see eq. 3.14) for a given source location. The ARM describes the width of the PSF-cone and can be described as function of $\bar{\varphi}$. It is usually integrated over a range in $\bar{\varphi}$ between $[\bar{\varphi}_{min}, \bar{\varphi}_{max}]$. An example for a simulated on-axis source is shown in Figure 3.12 on the left.

The quality of the reconstruction can be assessed by displaying the kinematic and geometric scatter angle in a 2-dimensional histogram as shown in Figure 3.12 on the right. The plot shows the scatter angle distribution of Compton events for a

and as a cone of finite ARM width in the CDS of a Compton telescope.

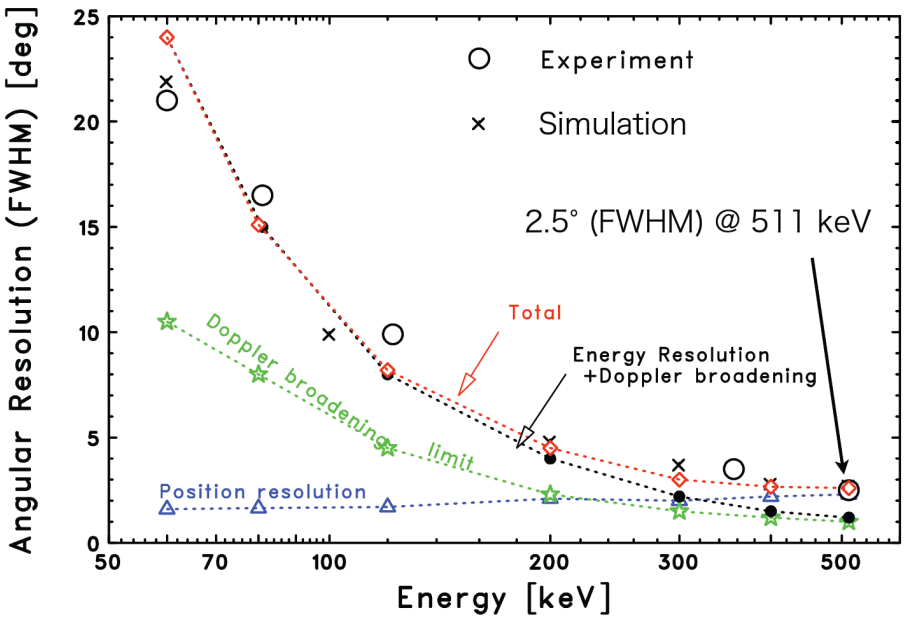


Figure 3.13: Example for an angular resolution of a silicon and cadmium telluride Compton telescope. Given are the values derived from simulations with the contributions of position resolution (blue), Doppler broadening with and without energy resolution (black and green) and the total expected angular resolution (red). The measured performance (open circles) aligns with the theoretical expectation. From [Tak09].

simulated on-axis source at 10 MeV using a Compton/Pair Telescope. All events with a correctly reconstructed hit 1 and hit 2 align on the diagonal of the data space. The width of this diagonal is a measure for the angular resolution for this specific source. Cluster away from the diagonal hint at faulty reconstruction or pollution from non-Compton-events.
The resulting Angular resolution of a Compton Telescope shows a strong dependence on the incident energy of the measured photons as shown in Figure 3.13. At low energies, the performance is limited by the accuracy of the energy measurement as this has a direct impact on the uncertainty of the calculated kinematic scatter angle. At higher energies the error on the energy measurement becomes sub-dominant and the angular resolution is capped by either the position resolution of the detector or the detector material via Doppler-Broadening (see Figure 3.6).

3.2.3 Pair Telescopes

Pair telescope exploit the shape of the electron and positron tracks to derive the origin as well as momentum and energy of the converted photon. While the pair creation probability becomes flat after several 10s of MeV, the actual magnitude is ma-

terial dependent (see e.g. Fig. 3.2). The imaging capabilities are a trade-off between conversion efficiency (thick, high Z material) and minimizing Moliere scattering of electron and positron (thin, low Z material) to retain directional information. Conversion and initial tracking was classically done with sparc chambers (see EGRET 3.4.1) or more recently with silicon strip trackers (see Fermi-LAT 3.4.3) interweaved with thin (high-Z) tungsten foil. For energy measurements, solid state calorimeters made of e.g. NaI or CsI emerged as the preferred approach. In order to reconstruct the origin one needs to identify the creation vertex and the individual tracks of the created e^+ and e^-. For determining the energy of the photon one can use the opening angle of the created particles [Ols63]. They give a correlation between opening angle and energy of the photon of

$$E_{MeV} \propto \frac{1.6}{\theta} \tag{3.19}$$

where θ is the opening angle in radian between electron and positron in the absence of multiple scattering. For sufficiently high energies the tracker is not able to resolve the opening angle any more and one has to resort to measure the energy in the pair either directly with e.g. a solid state calorimeter, creating the need for a very massive detector, that can contain the energy. The approach of Fermi-LAT refined for Pass 8 uses the profile of the calorimeter shower generated by the pair to estimate the energy and the incoming direction of the electron and positron. With that, they identify the tracks and the origin of the photon. This has the advantage of not needing to contain the whole energy of the pair in the calorimeter and therefore saves mass [Atw+13]. At lower energies, Molière scattering of the Lepton pair in the tracker layers washes out the opening angle,worsening the angular resolution at low pair energies. For higher energies the angular resolution is limited by the positional resolution of the tracker [TM22]. Pair telescope concepts with very good position resolution can reach an angular resolution in the order of up to tens of arc minutes. An example is the Advanced Energetic Pair Telescope (AdEPT), which can reach 0.6° at 70 MeV incident energy and in the order of 20 arc minutes at 200 MeV [Hun+14].

3.2.4 Compton/Pair Telescopes

Compton/Pair telescopes try to combine the capabilities of Compton and Pair telescopes in order to cover as much of the respective energy ranges as possible in one instrument. As a result, the performance is a compromise between the necessities of both interaction channels. The scattering detector has to be of low Z material in order to allow for Compton scattering to occur over a large energy range. On the other hand, the thickness of the individual parts of the scattering detector needs to be small enough to minimize Moliere scattering in the pair regime. This leads to a segmented tracker design, usually consisting of several layers of DSSDs or Silicon Pixel detectors as in MEGA [Blo+06] or ComPair-2 [Cap+24]. This enables charged particle tracking inside of D1. Due to the thinner layers, the tracker is able to track recoil electrons from Compton interactions, if the momentum transferred to the recoil electron is sufficient to allow the electron to leave the respective layer. The measured recoil

electron scattering angle δ as shown in Figure 3.3 can be used to define the Scatter Plane Deviation (SPD). It is the angle between the true scatter plane spanned by the direction of momentum of the initial and scattered photon and the measured plane spanned by the direction of momentum of the scattered photon and the scattered recoil electron :

$$SPD = arccos\left(\left(\frac{\vec{p_{incident}}}{\left\|\vec{p_{incident}}\right\|} \times \frac{\vec{p_{gamma}}}{\left\|\vec{p_{gamma}}\right\|}\right) \cdot \left(\frac{\vec{p_{gamma}}}{\left\|\vec{p_{gamma}}\right\|} \times \frac{\vec{p_{electron}}}{\left\|\vec{p_{electron}}\right\|}\right)\right) \tag{3.20}$$

When reconstructed correctly, the recoil electron track can therefore be used to constrain the Compton cone to a Compton arc and break up ordering ambiguities.
Since the number of electronic readout channels is usually constrained, thinner layers needed for the pair tracking result in less mass allocated to the scatter detector, which reduces the efficiency in the Compton regime. In order to retain a good Compton performance it is also imperative to minimize the amount of passive material in the tracker. One has therefore to refrain from high-Z material in between the layers as commonly used in pair telescopes to increase the pair conversion efficiency. This can have benefits at lower energies, as the electrons and positrons have less interactions on their path to the calorimeter, but leads to a smaller conversion efficiency in the tracker which results in the majority of photons to undergo pair creation inside the calorimeter and not in the tracker as will be discussed later.
Since the instrument uses both Compton scattering and pair creation as detection channels, the overall performance shows a bump in the transition region, where Compton scatter becomes less effective and the photon measurement via pair creation is still hampered by secondary scatterings of the electron and positron.
Despite these shortcomings, the end result is a design capable of measuring gamma photons in the range of 0.1 MeV to potentially several GeV in one compact instrument with large field and sufficient sensitivity suitable for surveys of the gamma-ray sky [De +21].

3.3 Space Environment in Low Earth Orbit

The European Space Agency defines numerous orbits by their inclination and perigee and apogee altitude (see e.g. [SLL24]). Among the most widely known are Geostationary Orbits (inclination between 0 and 25 deg, perigee and apogee altitudes between 35586 km and 35986 km) and Low Earth orbits (LEO) (perigee and apogee up to 2000 km). The majority of payload launches in 2023 targeted LEO orbits (2835 in total), while all other orbits combined for 58. Almost 98% of all payload launches are placed into LEO [SLL24]. The reason for that are launch costs (the higher the orbit, the more fuel and thrust has to be used to place the payload into orbit) as well as more favourable radiation conditions. Most LEO are still below the major radiation belts, shielding the payload to some degree from solar winds and cosmic radiation, but expose the spacecraft to trapped highly energetic particles [LSY19]. The targeted orbit for many scientific satellites is the LEO, whose main characteristics of the radiation environment are summarized in the following.

3.3.1 Cosmic Diffuse γ-ray background

The diffuse γ-ray background consists of γ-ray emissions that cannot be linked to a source. Their origin are cosmic ray interactions with the galactic interstellar medium, extragalactic background and unresolved point sources. The cosmic ray protons can interact with the interstellar medium and produce π^0 which then decay into γ-rays. Cosmic electrons can produce gamma-photons via inverse Compton scattering and bremsstrahlung [OS13].
The contribution of these two inner galactic components can be used together with particle transport models to probe the density and distribution of the interstellar medium [Hun+97]. The extra galactic background was measured to be highly isotropic and to follow a cut-off power law [Ack+15]. Its origin may be blazars, star-forming galaxies and radio galaxies [Aje+15]. Anisotropies (or lack thereof) in the extra galactic background can be used to constrain annihilation cross sections [And+07]. The cosmic diffuse γ-ray background is therefore not primarily a background source, bút also an active science object of INTEGRAL and the upcoming COSI-SMEX mission [Sie+22].
A more recent modelling from [Cum+19] over the range of hard X-rays to the GeV range combines the measured spectra from different instruments and is depicted in Figure 3.14.

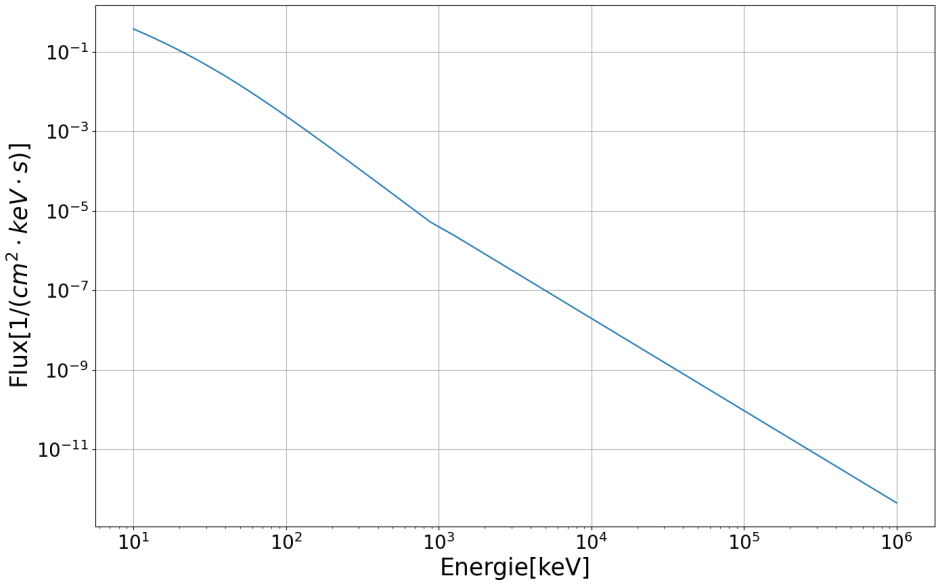


Figure 3.14: Cosmic diffuse γ-emissions modelled in the range up to 890 keV after [Tue+10] and [Ack+15] for higher energies. Generated with [Cum+19].

3.3.2 Cosmic Rays

Cosmic rays consist of stable charged particles and nuclei with life time of 10^6 years or longer. Their origin is from outside the solar system [Dru12]. The measured flux of particles below 10 GeV is anti-correlated with the solar activity [TBF22]. The flux of cosmic rays encountered in orbit is dependent on the strength of Earth's magnetic field, parametrized by the vertical cut-off-rigidity along the orbit. Assuming a dipole one can obtain [Sto55]:

$$R_{cut} = \frac{M \cdot \cos^4 \lambda}{r^2(1 + (1 - \sin\epsilon \cdot \sin\zeta \cdot \cos^3 \lambda)^{\frac{1}{2}})^2}, \tag{3.21}$$

where M is the dipole magnitude, r is the distance from the dipole center, λ the geomagnetic latitude, ϵ is the zenith and ζ is the azimuthal angle The orbit-averaged flux is therefore a function of mean altitude, inclination and solar cycle. Cosmic rays are measured to be highly isotropic.
Figure 3.15 shows the measured spectrum from air-shower measurements.
The spectrum can be divided into a 'knee', a second 'knee' and an 'ankle'. Cosmic rays with an energy of less than 10^9 GeV are assumed to be of inner galactic origin. A possible interpretation of the 'knee' would be that, at this energy, most galactic accelerators reach their maximum energy. The second 'knee' would correspond to a transition to heavier primary particles. The 'ankle' could mark the beginning of the

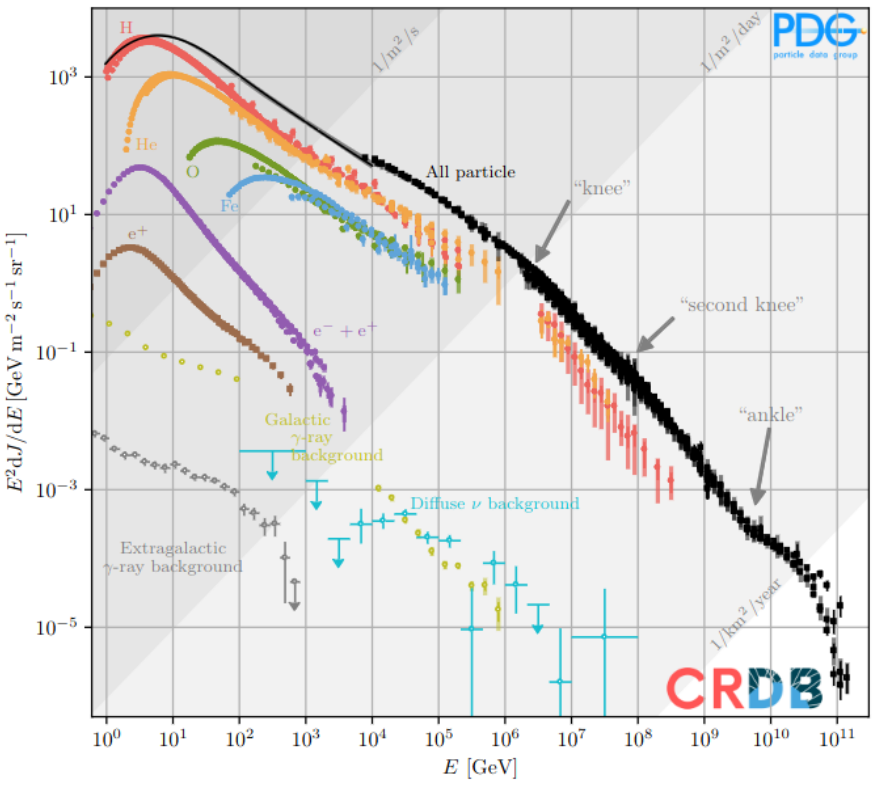


Figure 3.15: Cosmic ray all-particle spectrum from air shower measurements. From [Nav+24].

extra galactic component's dominance [Nav+24].
Cosmic rays are charged particles and can generate false signals in the telescope. To counteract this, one usually implements an anti-coincidence system that vetoes any event when it registers a particle passage. The immediate effect of cosmic rays should therefore be controllable [Sch01]. However highly energetic nuclei are capable of activating detector material, turning the detector itself into a background source. Cosmic rays have been found to be a primary source of detector activation in a low earth orbit [Dea+03].

3.3.3 Earth Albedo

Albedo particles are generated during atmospheric showers induced by cosmic rays. During intranuclear cascades the excited atmospheric nucleus is generating electrons, positrons, protons, neutrons and photons over a wide range of energy. Since Earth's magnetic field shields the atmosphere from the less energetic cosmic rays, it is the high energy component that ultimately leads to the emission of photons, protons, neutrons and electron-positron pairs. Figure 3.16 shows a schematic of the

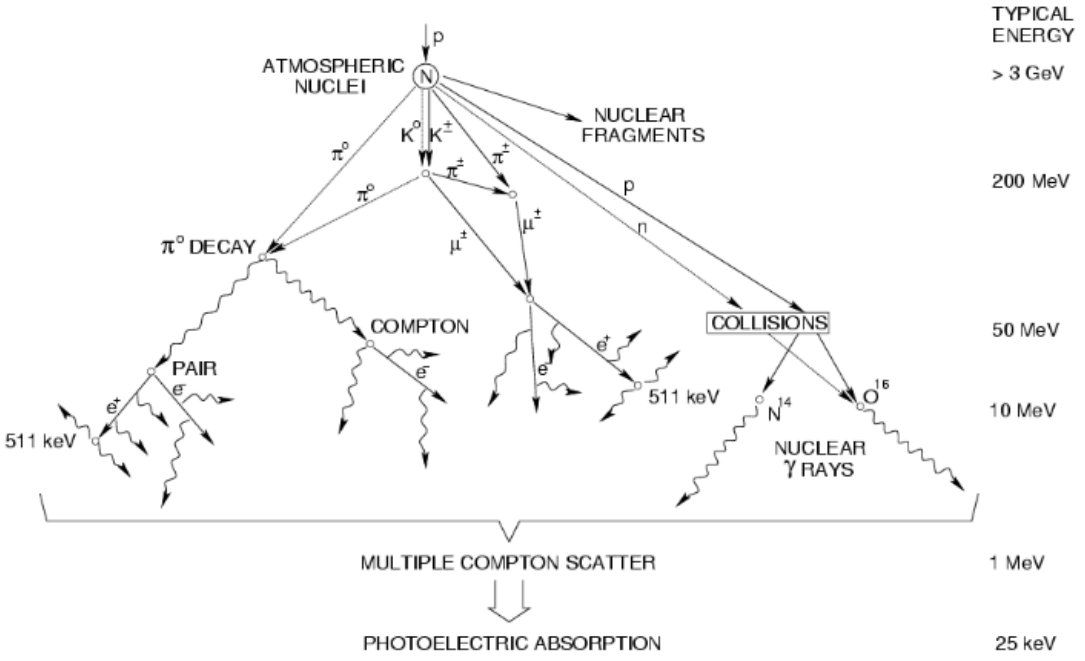


Figure 3.16: Schematic of the involved production processes for albedo radiation. From [Zom07].

production mechanisms.

Albedo Photons

The processes that generate albedo protons and neutrons lead also to the emission of photons with energies in the range of 10 keV to several GeV [Abd+09]. In fact the Earth atmosphere is the brightest γ-source a satellite born instrument could measure if sensitive enough in the lower energy ranges [Dea+03]. It consists of a continuous component stemming from particle creation and decay as well as an annihilation component [HSL03]. This 511 keV line is from annihilating positrons that get created by the cascading events. Both components are shown in Figure 3.17 and 3.18.
Since the incident direction of the particles is usually from below[2] a good way to reduce the amount of background is to determine the direction of the photon at trigger level as realized with COMPTEL by applying the time of flight method. Other options include active shielding or online directional reconstruction.

[2]It is assumed that the telescope is pointing towards the sky. This is not necessarily true for all observatories.

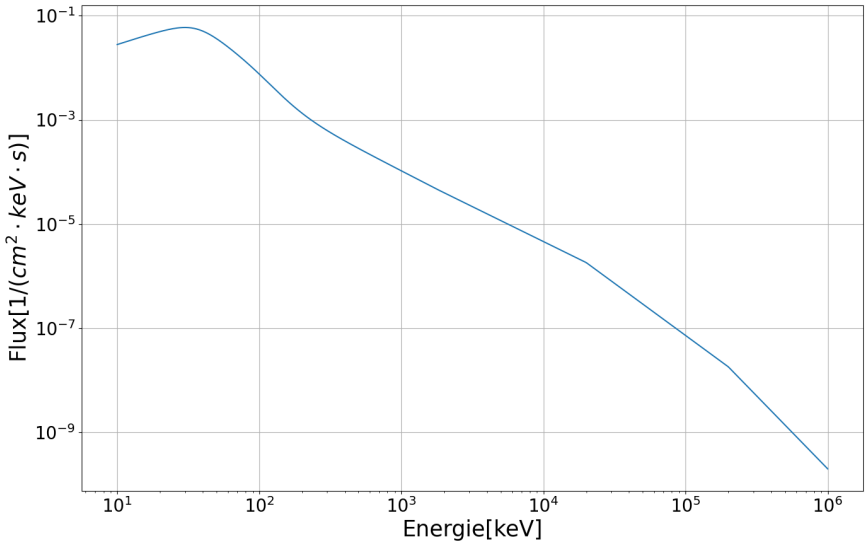


Figure 3.17: Contonous part of the Albedo Photon spectrum for a = 550 km and i = 2.5° modelled after [Chu+07], [Saz+07] and [Tue+10] for energies below 1850 keV and [Miz+04] for energies up to 200 MeV. Higher energies are modelled with [Abd+09] from Fermi-LAT measurements. Generated with [Cum+19].

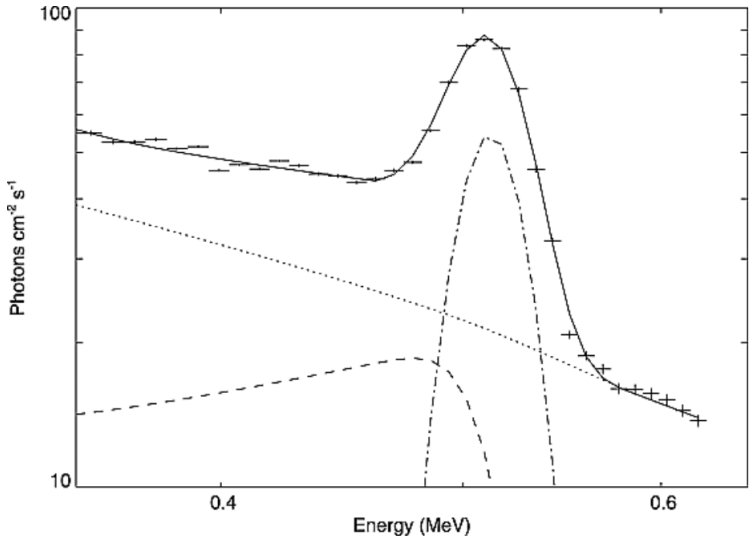


Figure 3.18: Count spectrum observed by the SMM GRS from Earth's atmosphere accumulated at cutoff rigidities > 11 GV during the 48 d interval 6 December 1986 to 19 January 1987, compared with a fitted model spectrum. Data points, measurements; dotted line, power law continuum component; dashed line, Compton scattered component arising from 0.511 MeV line; dot-dashed line, 0.511 MeV line component; full line, sum of three components. From [HSL03].

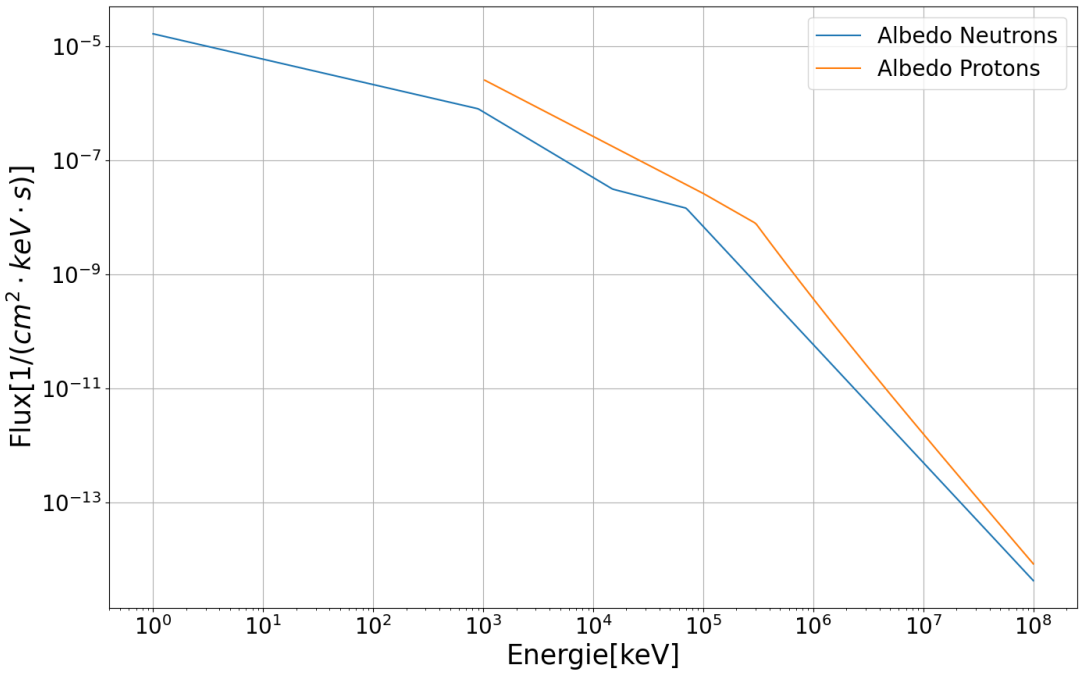


Figure 3.19: Albedo neutron and proton spectrum. Neutrons are modelled with [KPM15], protons use [Miz+04] for a orbit with a=550 km and cut-off rigidity of 4.5 GV. A thermal neutron component extends down to 10^{-2} keV (not shown). Generated with [Cum+19].

Albedo Neutrons and Protons

Neutrons are capable of activating the active and passive material in a detector, generating background. Since they don't possess a charge they are hardly distinguishable from photons for most detectors. Their spectrum ranges from around 100 keV to several GeV. The majority of the produced neutrons will have lesser energies. They are produced in hadronic air showers resulting from collisions of cosmic rays with nuclei of the atmosphere. Due to the large momentum transfer an a collision an atmospheric nucleus undergoes an intranuclear cascade, resulting in the ejection of neutrons with energies from 10 to several hundred MeV. Most of the generated neutrons will head toward Earth [KPM15].
After the intranuclear cascade, the remnant nucleus deexcites towards its ground state by emitting further hadrons. The energy of the neutrons generated in this process is around one MeV. After production the neutrons will continue to loose energy due to interactions with the Earth atmosphere until an equilibrium with the atmospheric nuclei is reached. The shape of the spectrum depends on the magnetic latitude and the solar activity. Earth's magnetic field is a shield that protects ,for example, the region of the equator from charged cosmic rays with a rigidity below 15 GV while the polar regions are only protected up to 1 GV. Therefore the fluxes of neu-

trons are highest at the poles and lowest at the equator. Also the solar magnetic field shields Earth in addition to its own magnetic field [KPM15]. The two magnetic fields have their biggest impact on the low energy cosmic ray particles. Consequently, the effects of variation in these fields are largest at the poles. For high-energy protons the interaction rate varies little with magnetic latitude and solar activity [KPM15]. The energy spectrum of of albedo protons generated during these cascades was found to be limited to energies up to 5 GeV[Adr+15]. Figure 3.19 shows modelled spectra for an orbit with altitude of 550 km and $R_{cut-off} = 4.5$ GV.

3.3.4 Trapped Particles

Earth's magnetic field can be described as a first approximation by a dipole. The poles are connected by field lines. Cosmic ray particle are charged and can therefore be either deflected or trapped by field lines. In the latter case the particle is forced to travel along a field line from pole to pole. These regions of trapped particles are also called Van-Allen belts. The usual altitude of the trapped particles is high enough to be of no concern. However, the magnetic field strength is not uniform around the globe. At the weakest point, the trapped particles dip down to altitudes very well in reach of a satellite in a low earth orbit. This region is called the South Atlantic Anomaly (SAA). The energy of the trapped protons ranges from a couple MeV to several hundred MeV. The flux rates inside the SAA can quickly reach dangerous levels and cause damage to the electronics [MSIV77]. A feasible approach is to shut down the satellite for the duration of passage. Figure 3.20 shows the extent of the SAA in 2015 determined by [ARB18] at an altitude of 800 km. They estimated the shape of the SAA by counting triggers over a threshold rate from Precipitating Electron/Ion Spectrometers ("Special Sensor J") within the satellites of the Defense Meteorological Satellite Program (DMSP). From this map it is clear, that every spacecraft in low earth orbit has to cross the region eventually, with equatorial orbits being preferred. The shape and intensity of the spectrum is depending on altitude and inclination with respect to the equator. Two example spectra for a 550 km orbit with 0 and 40 degrees of inclination are shown in Figure 3.21 for Leptons and Protons. The shape of the spectrum depends on the orbit parameters, resulting in varying contribution to activation by the SAA.

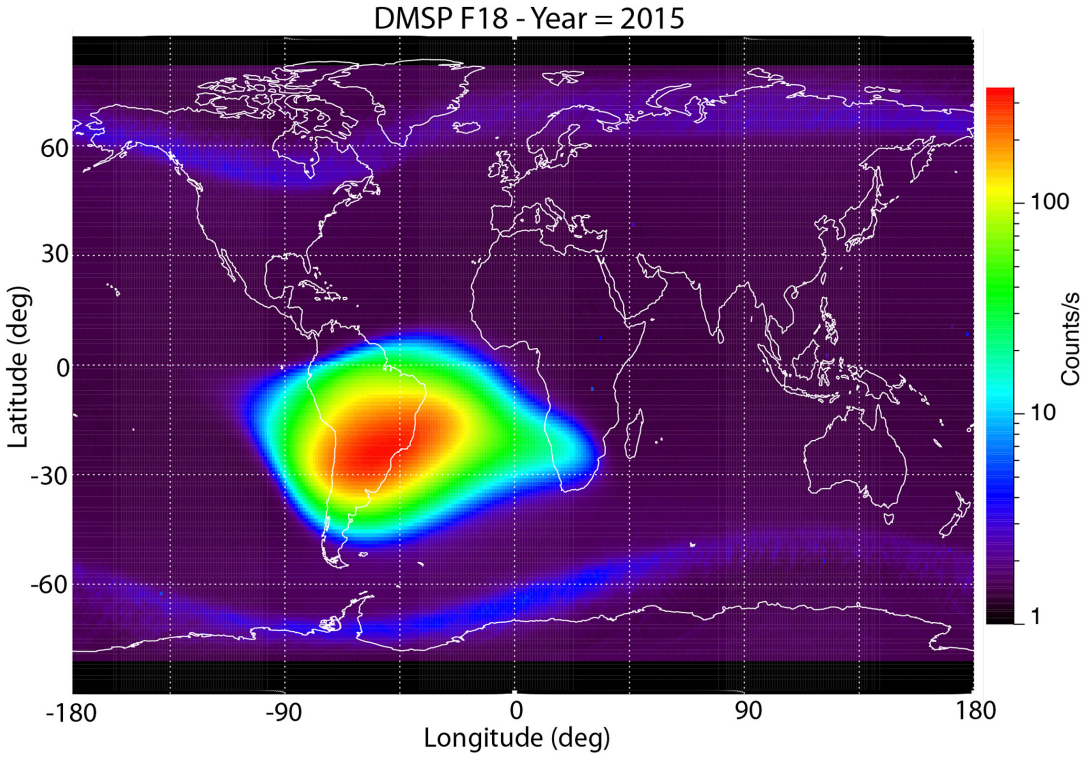


Figure 3.20: Count rate above threshold as function of geographic coordinates measured on board the Defense Meteorological Satellite Program (DMSP) spacecrafts at an altitude of 800 km. The count rate in Special Sensor J is correlated to the magnitude of the dip in magnetic field strength and can therefore be used to map the SAA. The faint bands close to the north and south pole correspond to radiation belts. From [ARB18].

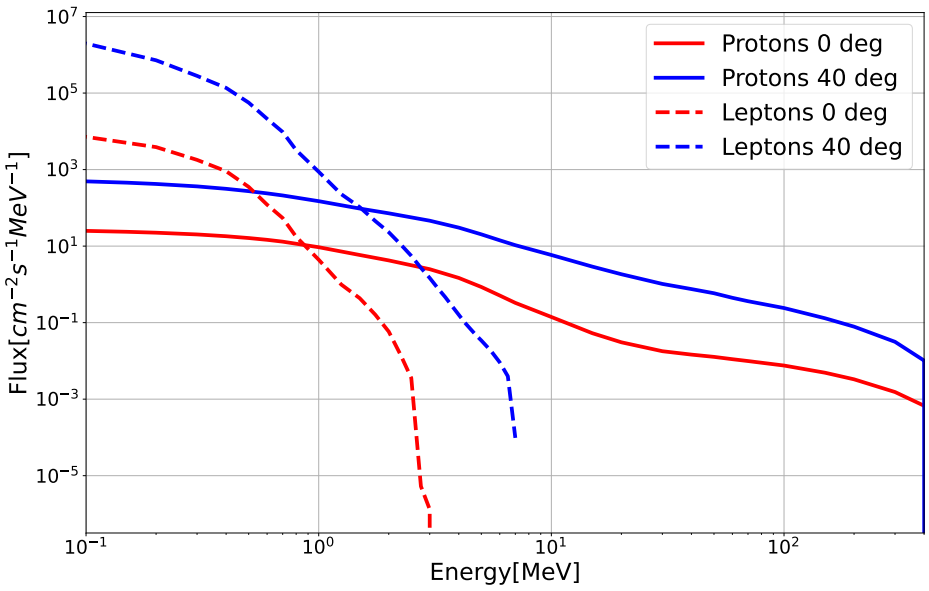


Figure 3.21: Simulated SAA spectra for two orbits at 550 km with 0 and 40 degrees of inclination using the AP9 and AE9 integration of SPENVIS. Generated with [Spe].

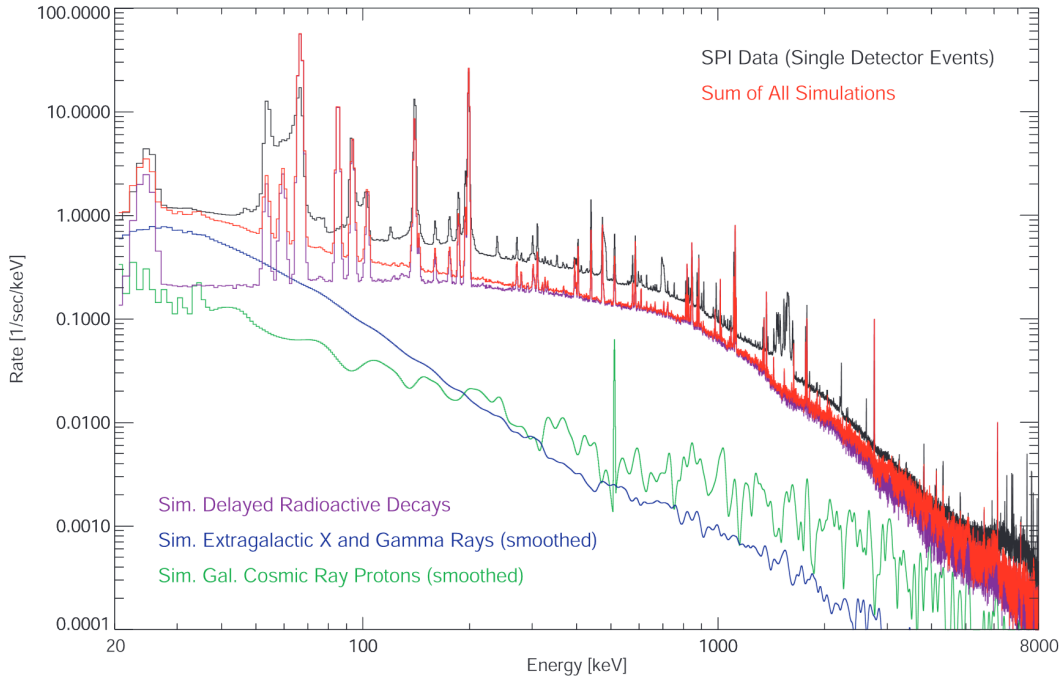


Figure 3.22: A comparison of simulated SPI single detector events with actual flight data. The broad spikes in the data in 1.4–1.6 MeV are electronic noise. The black curve denotes the measurement, the sum of all simulated components is given in red. From [Wei+03].

3.3.5 Internal Activation

The background in space does not only consist of the particles impinging on the detector. Over time, the satellite itself becomes a background source due to the activation of the spacecraft by nuclear capture processes. Figure 3.22 shows the single site events measured in SPI on INTEGRAL in March 2003. Weidenspointer et al [Wei+03] were able to measure 303 lines and to identify over 200 lines resulting from activation of the satellite. These events form an additional measurement background, which has to be modelled .

Since the detector is continuously exposed to activating radiation, the intensity of the activation spectrum varies with time. As can be seen from activation analysis in COMPTEL, individual components can reach an equilibrium or not, depending on the lifetimes of the created isotopes. Figure 3.23 shows the time evolution of individual activation rate spectra. As can be seen, most of the isotopes plateau out after some time. Changes can occur, when orbit parameters change, as shown in Figure 3.24. For the initial orbit, all components evened out. After the first orbit correction at 9280 s to 450 km most components initially rise, but reach a (in some cases higher) equilibrium between produced and decaying isotopes. For ^{22}Na, this was no longer the case. The rate grew constantly over the remaining duration of the

mission. The second reboost at 10560 s increased the production rate of ^{22}Na even further [Wei+01]. This was rooted in its lifetime of 3.7 years, compared to e.g. ^{24}Na with a lifetime of 21.6 h. Internal activation is therefore an orbit dependent and also potentially time dependent background component.

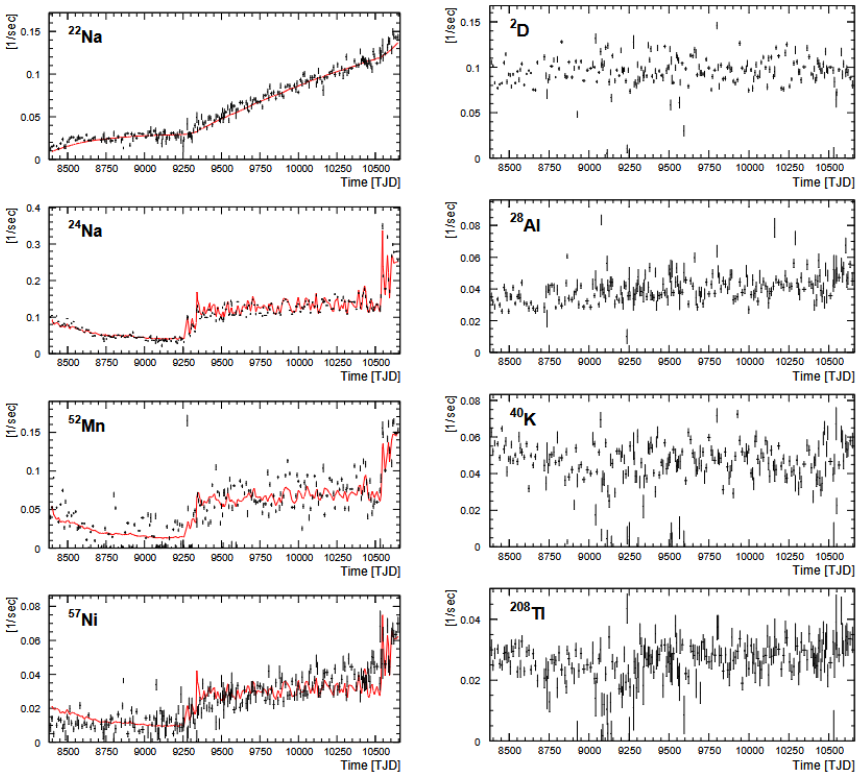


Figure 3.23: Count rates of important activation lines measured in COMPTEL over the duration of the CGRO mission. From [Wei+01].

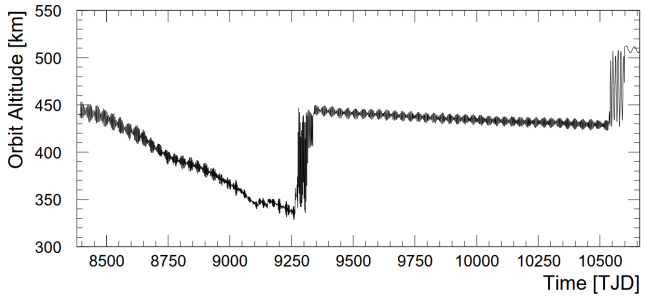


Figure 3.24: CGRO orbit altitude vs. time. Note the two orbit adjustments at t = 9280 s and t = 10560 s. From [Wei+01].

3.4 Past, Present and Future Instrumentation

3.4.1 The Compton Gamma-ray Observatory (CGRO)

The Compton Gamma Ray Observatory [Kni+88] was the second of NASA's Great Observatories. It was launched on 5 April 1991 and orbited Earth in a LEO with altitudes between 350 km to 500 km with an inclination of 28.5 deg. With a total mass of 17 tons it carried four instruments (OSSE, BATSE, COMPTEL and EGRET)[3] covering the electromagnetic spectrum in the range from 20 keV to 30 GeV. Fig. 3.25 shows a sketch of the satellite and the deployment with from the Space Shuttle Atlantis. The experiments BATSE, COMPTEL and EGRET will be described in the following. Details of the Orientated Scintilaton Spectrometer Experiment (OSSE) can be found in [Joh+89]. CGRO was de-orbited on 4 June 2000.

Figure 3.25: Left: Sketch of the instrument positions on the satellite. Right: Deployment of CGRO from the payload bay of Space Shuttle Atlantis. Mission Specialist Jay Apt for scale. Courtesy of NASA.

BATSE

The Burst and Transient Source Experiment (BATSE) [Hor91] was an all-sky gamma-ray transient monitor. It consisted of eight detector groups located at the corners of the observatory. Each module contained a Large Area Detector (LAD) and a Spectroscopy Detector (SD). The LAD consisted of a NaI crystal behind a charged particle detector. It was read out with three PMTs. In contrast to the more disk-like shape of the LAD crystal, the SD crystal had a cylindrical shape and a greater depth (7.62 cm vs 1.27 cm) to improve the energy measurement. It was read-out with a single PMT and placed underneath the LAD. The arrangement around the observatory allowed to measure each GRB with at least 4 detectors at all times.

[3] The total science payload was about 6 tons.

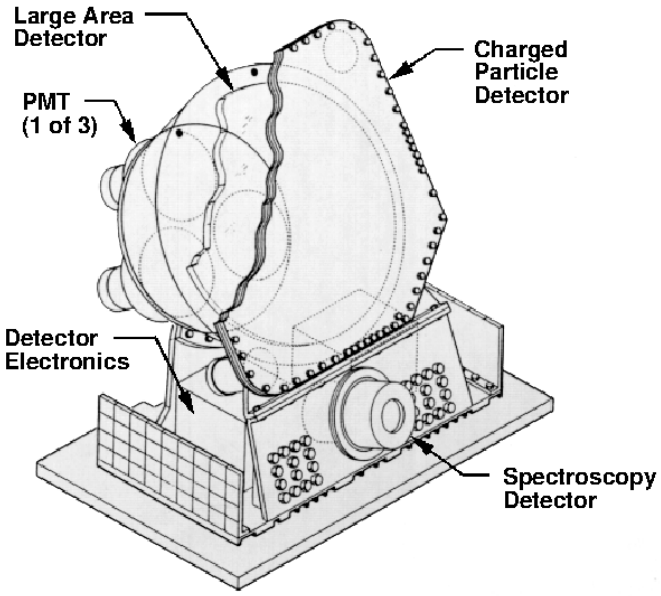


Figure 3.26: Detector module consisting of the LAD and SD. Eight of these modules were placed on the top and bottom corners of the observatory. From [Hor91].

The nominal energy coverage was between 28 keV to 1.8 MeV. This allowed for burst spectroscopy from the very low end to energies up to several hundred MeV with data collected from the other CGRO experiments, as for example for GRB990123 [Bri+99]. Over the course of its mission it was able to collect over 2700 GRBs, a feat only surpassed by Fermi-GBM due to its long mission lifetime. Its capability to localize the source of the transient signal to order of degrees delivered strong evidence for the extragalactic nature of GRBs.

COMPTEL

The imaging Compton telescope COMPTEL [Sch+93] was the first and up to now only Compton telescope on board a satellite observatory. It consisted of two detectors. The first one, D1, contained NE 213A, a low-Z liquid scintillator that acted as a scatter detector. The second detector, D2, contained NaI(Tl), a moderately high-Z crystal scintillator with very high light yield. D2 acted as absorbing detector. The scintillation light was collected with PMTs. A gap of 1.5 m between D1 and D2 allowed for sequence confirmation (D1 to D2) via time-of-flight. To suppress signals

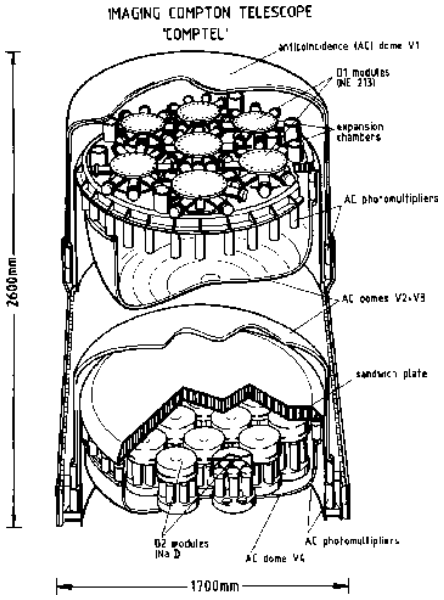


Figure 3.27: Drawing illustrating COMPTEL. From [Sch+93].

by charged particles the two detector domes were surrounded by an anti-coincidence shield.
COMPTEL had an energy range from 0.75 MeV to 30 MeV and was able to achieve an angular resolution of about 1 to 2 deg. It had an energy resolution of 8.8% at 1.27 MeV. Since it was the first and only Compton telescope in space its track record is filled with milestones. Some of them are the COMPTEL Maximum-entropy All-sky Maps [Str+97] or the first 1.809 MeV all-sky-image [Obe+96; Obe97].

EGRET

The Energetic Gamma-Ray Experiment Telescope (EGRET) [Kan+89] was the high energy detector on CGRO. It had a nominal energy range of 20 MeV to 30 GeV. In this range, pair creation is the dominant interaction mode. Figure 3.28 shows a schematic of the instrument. It used a spark chamber to create the pairs and measure the initial tracks and and a NaI(Tl) calorimeter to measure the energy contained in the electron-positron pair. Between the closely spaced spark chamber and the calorimeter was a two-layer coincidence detector which determined the direction of traversing charged particles via time-of-flight. The detector was encapsulated with a plastic scintilator anti-coincidence veto. A downward pointing charged particle in the coincidence detector together with a sufficiently high energy deposit in the calorimeter with the absence of a veto triggered the high voltage in the spark chamber and recorded the event. The energy resolution was ranging between 20% and 25% depending on incident energy and angle of the converted photon. The angular resolution ranged between 5 and 30 arcmin, depending on source intensity and

location.

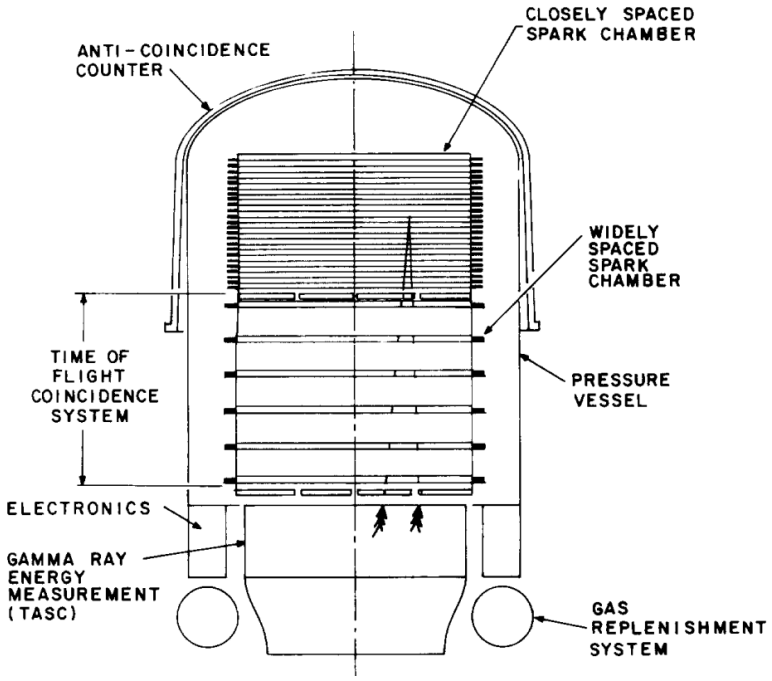


Figure 3.28: Schematic of EGRET. The detector consisted of a spark chamber on top of a NaI(Tl) calorimeter inside of a surrounding anti-coincidence veto. From [Kan+89].

EGRET performed the first all-sky survey above 100 MeV. Its revised catalogue contains 188 point sources above 100 MeV [CG08].

3.4.2 INTEGRAL

INTEGRAL[Win+03a] is the International Gamma-Ray Astrophysics Laboratory and surveys an energy range from 15 keV to 10 MeV. In addition it monitors the X-ray and optical band . The main instruments are the imager IBIS [Ube+03] and the spectrometer SPI [Ved+03]. JEM-X [Lun+03] and OMC [Mas+03] provided the monitoring of the other bands. INTEGRAL was launched on 17 October 2002. IBIS, SPI and JEM-X use a coded mask as imaging technique. It was shutdown on February 2025 after its final observation cycle. The current orbit puts it on a trajectory for re-entry into Earth's atmosphere in March 2029.

IBIS

IBIS is the Imager on Board the Integral Satellite. It is able to detect γ-sources with a good angular resolution of around 12 arc-minutes. It consists of two detector planes, ISGRI and PICsIT. ISGRI is made of $128x128$ CdTe detectors optimized for low ener-

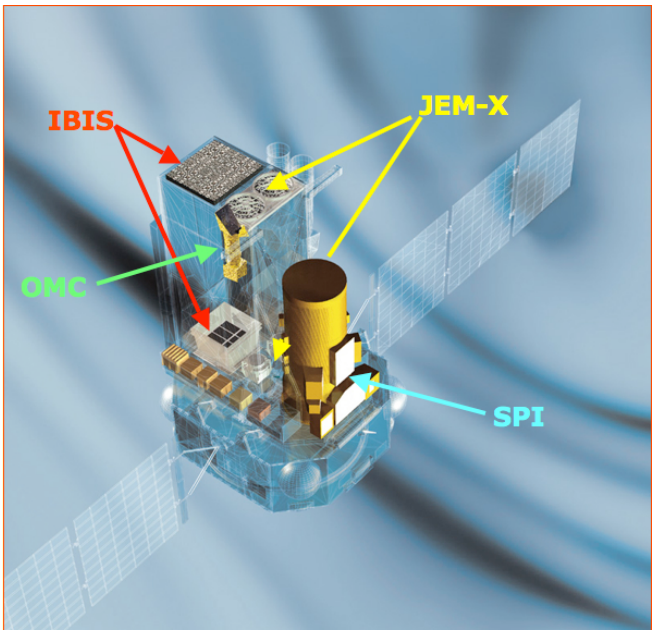


Figure 3.29: Overview of INTEGRAL with all instruments. Courtesy of ESA.

gies ranging from 15 keV up to 500 keV. PICsIT is an array of 64×64 CsI(Tl) crystals and covers the high energy portion up to 10 MeV. Both modules are separated by about 94 mm allowing the application of Compton kinematics to events that trigger in both detectors to aid in noise reduction. The modules are placed behind a coded mask. The veto system covers the whole of the high energy layer and extends just up to the low energy layer. IBIS provides a spectral resolution of 8% at 100 keV and 10% at 1 MeV.
IBIS was used to observe active galactic nuclei in the hard X-ray to low γ-range and collected over 360 AGNs in his all sky catalogue [Mal+16].

SPI

SPI is a coded mask spectrometer using 19 hexagonal high purity Germanium detectors as camera. The detector system needs a cooling system to keep the Ge at around 85-90 K. It achieves a fully coded field of view of 16° with an angular resolution of 1.5° with an energy resolution of 2.2 keV at 1.33 MeV. SPI operates in an energy range of 18 keV up to 8 MeV. To keep background levels low,the detector and most of its aperture are surrounded by an active shielding made of bismuth germanate (BGO) while the mask is also shielded with an plastic veto towards the Ge detectors. The energy resolution degrades with radiation damage. To counteract this, the crystals have to be annealed (baked at 100°Celsius) in regular intervals.
One of the key results obtained with SPI are the 511 keV all-sky maps[Kno+05], which are continuously updated (see e.g. [Sie+16a].

3.4.3 Fermi Gamma-ray Space Telescope

The Fermi Gamma-ray Space Telescope was launched on 11 June 2008 as Gamma-ray Large Area Space Telescope (GLAST) and renamed after nominal operations were achieved. It has a total mass of 4.3 tons and carries two instruments, the Large Area Telescope and the Gamma-ray Burst Monitor. The orbit has an altitude of around 565 km with an inclination of 25.6°. The energy coverage of the two instruments allows for GRB observations from 40 keV to several 100 MeV in one

Large Area Telescope (LAT)

Fermi-LAT [Atw+09] (Fermi-Large Area Telescope is the high energy γ-ray imaging pair telescope. Fermi-LAT consists of two detector units. The top is a converter/-tracker with 16 layers of converter foil made of tungsten interleaved by 18 layers of silicon strip detectors. Each of the 18 tracking layers is segmented into 4 by 4 modules with two layers (x- and y-strips) of single sided silicon strips. The strip pitch is 228 μm with a strip thickness of 400 μm.
The calorimeter consists of 16 modules with 96 2.7 cm by 2.0 cm by 32.7 cm CsI(Tl) crystals. The crystals are arranged in 8 layers with 12 crystals. The layers are aligned orthogonally to each other such that the crystal modules form a hodoscopic array. The calorimeter is therefore a 3D imaging device. The scintillation light is read out by two PIN diodes glued to each end of the crystal (a small one for high energies, a big one for low energies). The large one covers a range of 2 MeV upt o 1.6 GeV, while the small one covers the range from 100 MeV up to 70 GeV. Although the calorimeter has only a radiation depth of 8.6 radiation lengths the longitudinal segmentation allows energy measurements up to a TeV due to measurement of the longitudinal shower profile. The energy resolution is then limited by fluctuations in the shower leakage.
The modules segment tracker and calorimeter into smaller modules aligned on top of each other. Each of these stacks is wired together into a 'tower' with its own read out electronics and own tracker and calorimeter.
Tracker and calorimeter are enclosed by an anti-coincidence detector that acts as a veto for charged particles entering the detector. The veto was required to have an efficiency of at least 0.9997. In addition, it needed to be less prone to false vetoes created by back-splash (isotropically distributed secondary particles from within the detector that trigger events in the ACD and 'kill' legitimate events) that hampered the sensitivity of previous instruments above 10 GeV. To counter this effect, the ACD is segmented and only the nearest veto segments relative to the photon are taken into account.
Fermi-LAT's operational range is the region of 20 MeV up to more than 800 GeV. Major achievements of Fermi-LAT include the measurement of the spectrum of the isotropic diffuse Gamma-ray Emission between 100 MeV and 820 GeV [Ack+15] and cataloguing more than 7194 sources in the 50 MeV to 1000 GeV range [Bal+23].

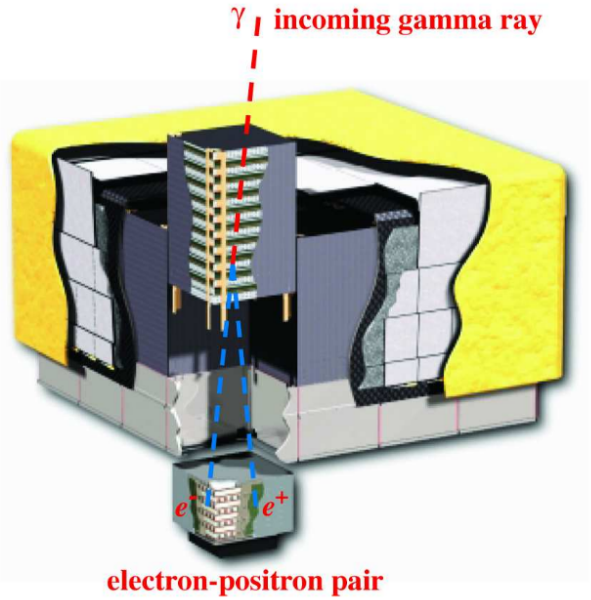


Figure 3.30: Schematic of the Fermi-LAT. The incoming photon is converted in the tracker. The created electron positron pair gets absorbed in the hodoscopic calorimeter.From [Atw+09].

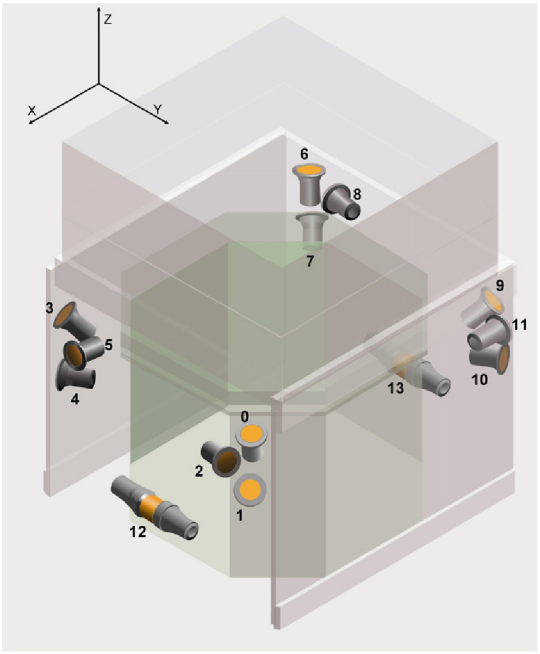


Figure 3.31: Placement of the NaI and BGO detectors of GBM on the Fermi Gamma-Ray Space Telescope. Detectors 0 to 11 are NaI, 12 and 13 BGO. From [Mee+09].

Gamma-ray Burst Monitor (GBM)

The Gamma-ray Burst Monitor (GBM) [Mee+09] on board the Fermi Gamma-ray Space Telescope consists of 12 small NaI(Tl) and two BGO detectors distributed around the satellite bus read out by PMTs. The placement is illustrated in Figure 3.31. The choice of detector materials allows for coverage across an energy range of 8 keV to 40 MeV, while the placement ensures that bursts can be detected from any direction of the spacecraft. The low end up to 1 MeV is covered by NaI detectors, the BGO detectors complete the energy spectrum at higher energies. The varying orientations of the NaI detectors (0 to 11) generate intensity patterns which allow to achieve a first GRB localization of the GRB within a few degrees. This enables Fermi to slew the LAT into a pointing, that facilitates follow up observations of the event in the LAT field-of-view with full coverage.
The Fermi-GBM GRB catalogue is constantly growing and reached at the time of writing 3885 GRBs [von+20], including GRB221009A, also known as the "brightest of all time"(BOAT) [Bur+23].

3.4.4 COSI-SMEX

The Compton Spectrometer and Imager (COSI) is a small explorer (SMEX) mission funded by NASA [Tom+23]. It is scheduled for launch in 2027. The design of the telescope is depicted in Fig. 3.32. It will operate in a nominal energy range between 0.2 to 5 MeV. It consists of four towers housing 4 layers of $8{\cdot}8{\cdot}1.5\,\mathrm{cm}^3$ Germanium strip detectors entrenched in a BGO-well which acts as active veto. The Ge-detectors are housed inside of a vacuum cryostat. The payload mass is around 400 kg. It is projected to reach an angular resolution of around 4.1°at 0.511 MeV and 2.1°at 1.809 MeV with a relative FWHM in energy of 0.80% and 0.78% respectively. Being a compact Compton telescope, it is able to reach a field-of-view of around 25% of the sky at all times, mapping the whole sky in roughly a day. Its four main science objectives are galactic surveys of nucleosynthesis lines, uncover the origin of galactic positrons, gain insight into extreme environments via polarization and probe the physics of multi-messenger events.
Its predecessor is the COSI-balloon project. On the last flight in 2016 it was able to perform measurements of galactic ^{26}Al [Bee+22], detecting and imaging 511 keV annihilation radiation from the Milky Way [Kie+20; Sie+20], performed observations of the Galactic continuum emission [Kar+23] as well as steady point sources, like the Crab Nebula and Cygnus X-1 [Rob+24].

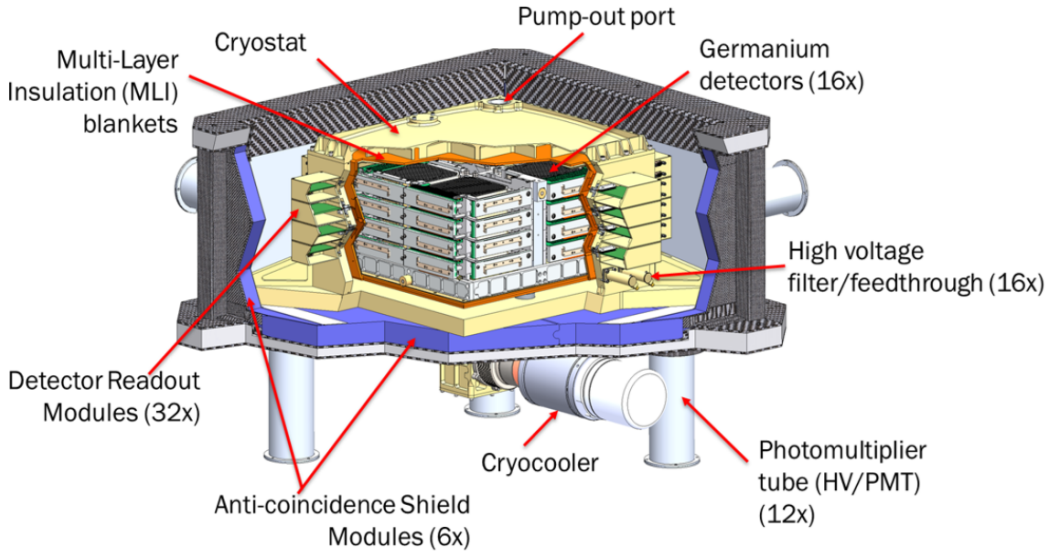


Figure 3.32: Schematic of the COSI telescope. The telescope is enclosed in a cryostat surrounded by a BGO-well that acts as active veto. Each germanium layer has $8{\times}8{\times}1.5\,\mathrm{cm}^3$. From [Tom+23].

3.4.5 e-ASTROGAM and Beyond

e-ASTROGAM (see e.g. [De +17b]) was a proposed detector design for the ESA M5-call following the Compton/Pair Telescope approach similar to the All-sky Medium Energy Gamma-ray Observatory (AMEGO [McE+19]). The design used a stack of 56 layers of Double-sided Silicon Strip Detectors (DSSDs) as tracking detector (D1). Each layer is subdivided into 4 modules with 5 by 5 wire bonded DSSDs. Each DSSD has a size of $9.5 \cdot 9.5 \cdot 0.05\,\mathrm{cm}^3$, resulting in layer size of roughly $1\,\mathrm{m}^2$. The strip pitch of 239 μm allowed for a positional resolution of less than 40 μm.
The design used a pixelated CsI calorimeter as absorber D2 consisting of 33856 crystals arranged in 529 (23 in x and y) modules with 64 crystals (8 by 8) each. The individual crystals had a size of $0.5 \cdot 0.5 \cdot 8\,\mathrm{cm}^3$ and was intended to be read out with silicon drift diodes. The geometry results in an estimated position resolution of 0.14 mm in x-y, while the z-resolution was assumed to 2 mm. The detector design used an outer detector covering five sides made of plastic scintillator as active veto. in Figure 3.34 shows the simulated angular (left) and energy resolution (right). The angular resolution for Compton as well for pair events was projected to improve on past and present instrumentation significantly. The energy resolution in Compton mode also promised a significant improvement over COMPTEL. The large payload mass of roughly 1000 kg and the $1\,\mathrm{m}^2$ area would have enabled an increment in sensitivity of more than a factor 10 in the Compton regime while enhancing the sensitiv-

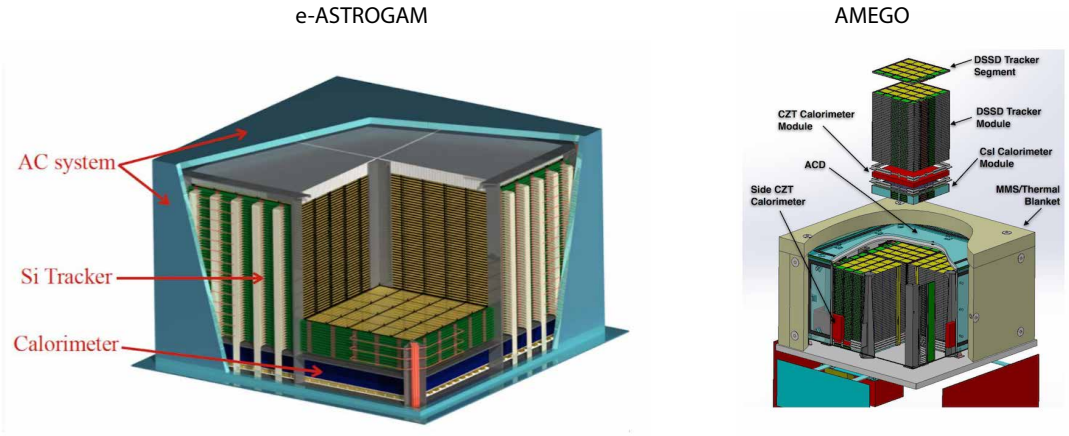


Figure 3.33: Illustrative sketches showing the design of e-ASTROGAM and AMEGO. From [De +17b; McE+19].

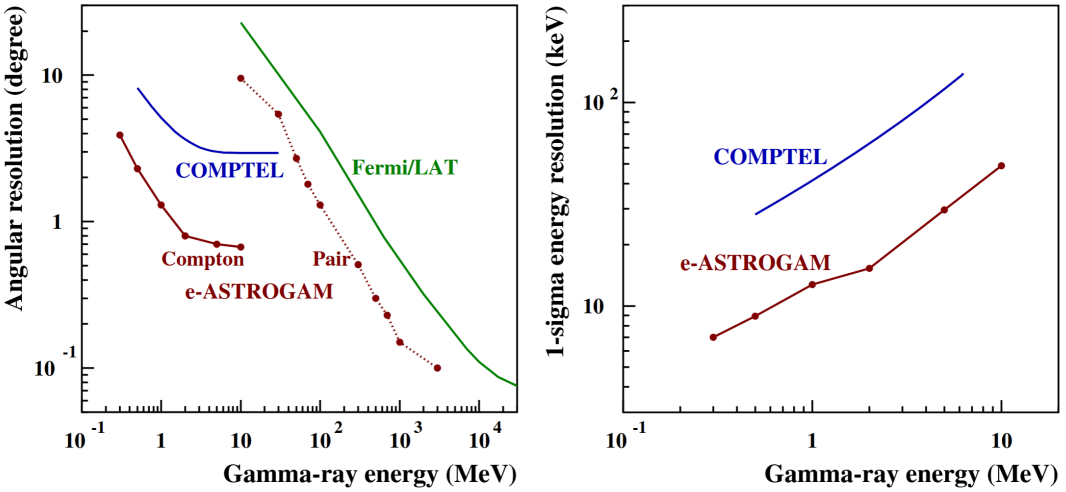


Figure 3.34: Angular (left panel) and energy resolution (right panel) as function of measured energy of e-ASTROGAM. The angular resolution is given for Compton and pair events. Values of COMPTEL and Fermi-LAT are given for comparison From [De +17b].

ity. Figure 3.35 illustrates this intended closing of the MeV Gap. The design proposal was unsuccessful in the M5 call.

A modified and scaled down version of e-ASTROGAM has been proposed for the ESA M8-call. This design incorporated a light coded mask on top of the tracker to improve the localization capabilities at low energies significantly and also make use of all interaction channels of the incoming photon [Ber+25]. This design was unsuccessful in the M8 call.

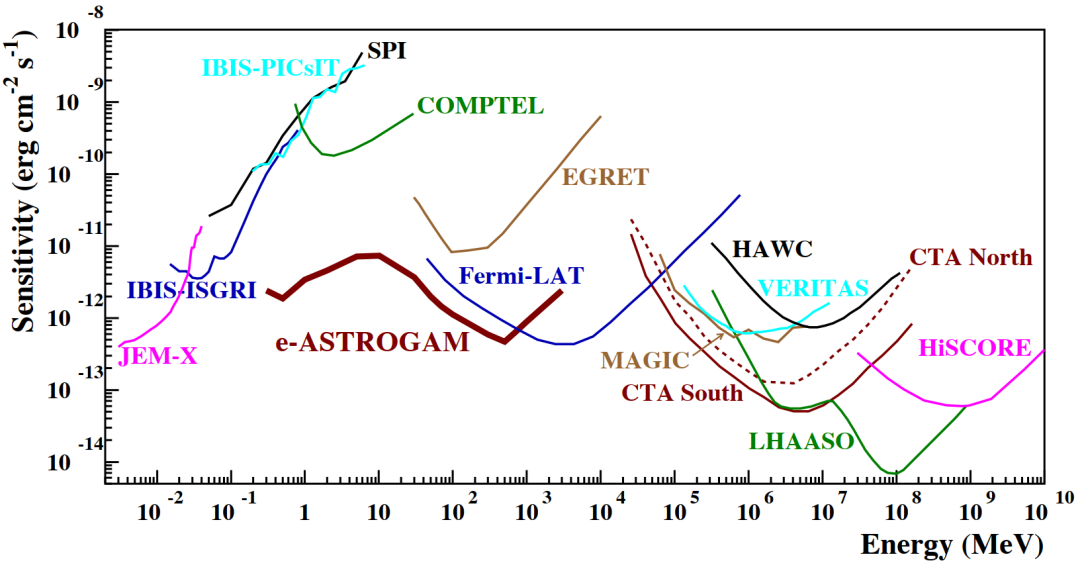


Figure 3.35: Sensitivities of past, present and planned gamma ray instruments. The landscape shows a clear void in the range between 1 MeV and 30 MeV, where only COMPTEL data is available. The curves for INTEGRAL instrumentation (JEM-X, IBIS (ISGRI and PICsIT) and SPI) are given for an observation time of 10^6 s. COMPTEL and EGRET values are given for 9 year mission duration. Fermi-LAT is given for a high Galactic latitude source with an observation time of 10 years. MAGIC, VERITAS and CTA are given for an observation duration of 50 hours. HAWC, LHAASO and HiSCORE sensitivities are given for T_{obs} = 5 years, 1 year and 1000 h respectively. The e-ASTROGAM 3σ-sensitivity is calculated for a source at high Galactic latitude and an effective exposure of 1 year. From [De +17b].

Chapter 4

Introduction to Deep Learning

4.1 Overview

The field of Artificial Intelligence (AI) is a branch of Computer Science that tries to model problem-solving and decision-making capabilities of the human mind on computers or machines in general. Following [RN21] it can be defined as a field of study that combines computer science with robust data sets and enables programs to solve problems on their own, without further human input. The rules that generate these decisions are in the case of *classical AI* tailor-made by a programmer for a specific problem.
Examples would include basically any AI in contemporary computer games or decision making in most conventional event data processing software in Physics (e.g. the default light-signal classification in the Xenon experiment [Ang+25] or the default MEGAlib event classification, see section 5.2.1). In general, these models take the shape of a tree and are very sensitive to variations of the initial conditions that

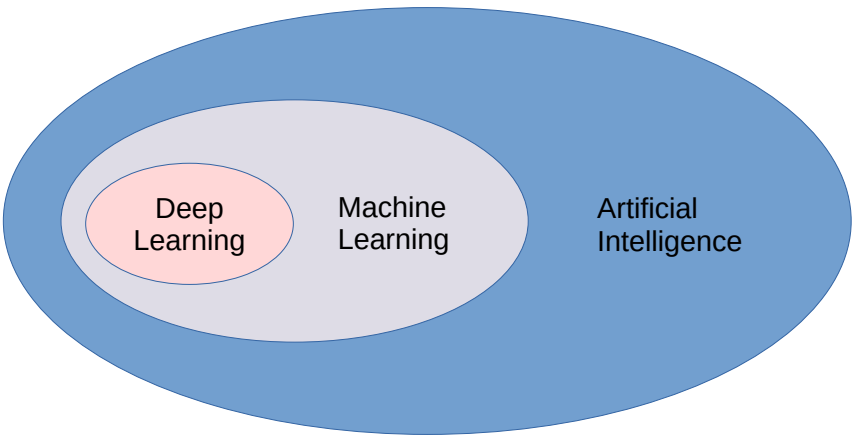


Figure 4.1: Sketch illustrating the interrelation of Deep Learning, Machine Leaning and Artificial Intelligence

the programmer used to derive his rule set. Even slight deviations can lead to cases where the implemented rule set fails. One reason for this is the tendency of the human brain to 'overfit' present patterns [Nuz15].
A way to minimize human bias in the rule sets is to let the machine itself define the rule set under which it derives a certain decision from the given data. This sub-field is called*Machine Learning* (ML). Figure 4.1 illustrates the interrelation of Machine Learning, Deep Learning and AI.
Classical Machine Learning models are widely established in physics. Well established algorithms include Random Forests and Boosted Decision Trees (e.g. [Cor+22]), Support Vector Machines (e.g. [WZ07]) and Bayesian networks (e.g. [Kro+21]). The strong point of these approaches is the transparency of the derived decision, as the "reasons" for a certain decision of the model can traced back through the model and associated with human comprehensible features. Data points in the input space can be characterized by properties. For example, an event can be described by its trigger topology, the total deposited energy, the deposited energy in each detector and number of triggered layers. The final model derived is a succession of splits on these data properties, optimized for maximum discrimination power on each split. It is always human readable and reconstructible how the model arrived to its decision by following the splits the model performs. One can also deduce the importance of a certain data-set feature for the decision making process by its discriminating power. The type of an event is determined after the last split on a data set partition. This leaf is called the terminal leaf. The class distribution after the final split, its purity, can be used as a proxy for the confidence of the model for a sample to belong to a certain class [Mit97]. Figure 4.2 shows a decision tree is for a three-class multi-classification task (Compton, Pair, non-reconstructible). At each split the algorithm tests for each attribute which condition gives the cleanest subset of data to proceed. When applied to unseen data, the track along the splits can be used to interpret the decision of the model.
The performance of the tree is dependent of the data representation used. The model found the deposited energy in the tracker to be the most important feature of an event. Removing it results in an insufficient data representation and leads to an underperforming model. The tree can only operate on the data features present in the data set. It therefore has no capability to combine present features to generate new input features. The transparency of the decision making is therefore paid for by heavily feature engineered data sets in order to find the optimal representation of the data *by hand*.

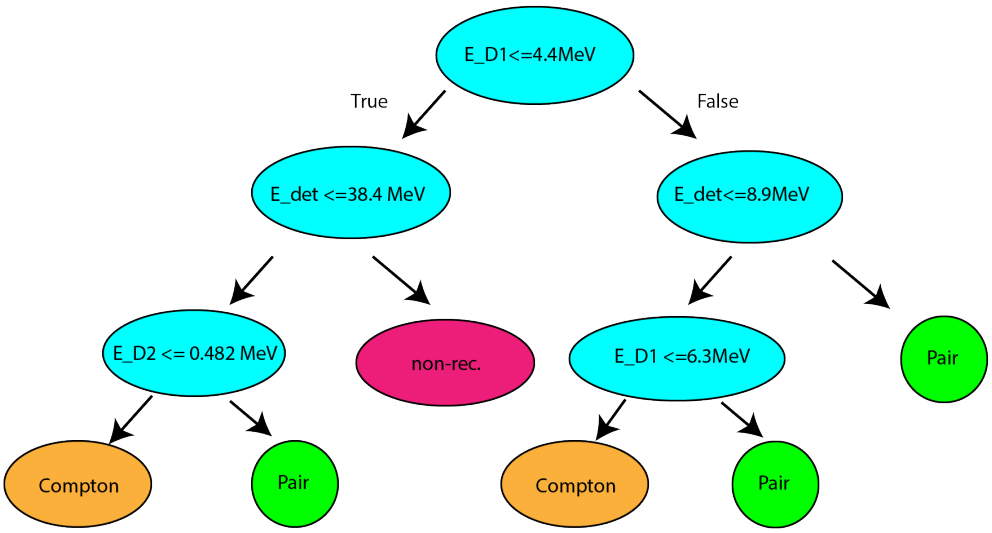


Figure 4.2: Decision tree generated with the data set described in Sec.5.6 with depth set to 3. Labels are Compton, Pair, and non-reconstructible. The classification is done by taking either the label of the samples of the remaining class in the terminal leaf (if the population is pure), or the majority of the training samples. The tree was able to classify 62% of the presented test set samples correct. Adopted from a tree grown with [Ped+11].

Deep Learning (DL) is a subset of Machine Learning that uses neural nets to model algorithms [GBC16]. Neural nets emulate complex algorithms by using simple and small building blocks called neurons to process information similar to biological nerve cells. The basic working principle is depicted in Figure 4.3. The data sample, here shown as a vector $\vec{x}$ is processed by the jth node by first applying a weighted sum

$$X = \sum_{i=1}^{n} a_{ij} x_i + b \tag{4.1}$$

Here a_i is the i-th adaptable weight of the node neuron j in question while b is called the bias term applied by the neuron. The summed value X is then processed by an activation function f, which is in general non-linear. By stacking consecutive layers of neurons and creating *deep neural nets* (DNNs), the non-linearity is able to generate new data representations, the so-called hidden representation which is helping or in some cases enabling the model to perform a given task.

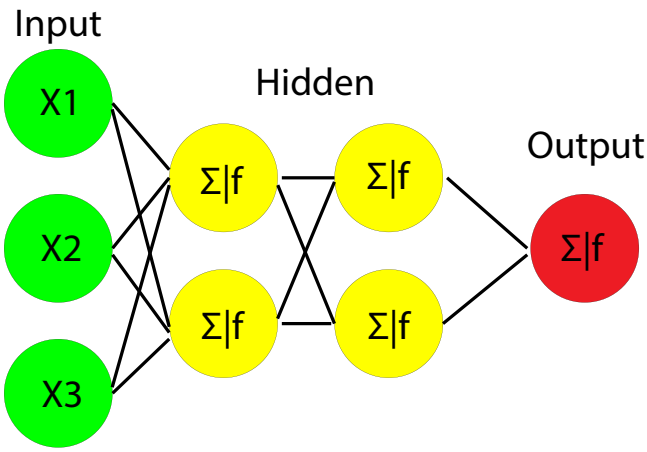


Figure 4.3: Example for a Multilayer Perceptron. The input is processed by multiple nodes in two successive layers. In a fully connected neural net, each preceding node is connected to all nodes in the next layer. In each node, the weighted sum is calculated (Σ) and put into the activation function f.

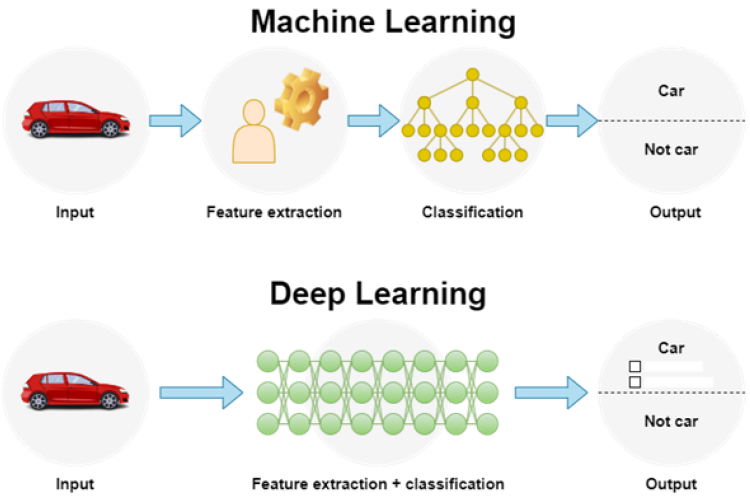


Figure 4.4: Sketch illustrating the workflow of classical Machine Learning and Deep Learning. From: [Jen+23].

Figure 4.4 depicts the difference workflow for classical Machine Learning and Deep Learning. The given task is to classify car and not-car. When employing classical Machine Learning, the analyst has to perform feature extraction and with that selects a classification approach. In the worst case, this can lead to a long loop of feature extraction and model training in order find the optimal representation. The situation simplifies for Deep Learning approaches. Here, the model is performs the feature extraction and classification by itself, reducing the task of the analyst to mainly finding the optimal model for the input data.

The following sections give an introduction into the basic ideas of Machine Learning, the concept of training algorithms on data and the relevant types of deep learning models. The content relies on discussions found in [GBC16], [Mit97] and [BB23] with other sources referenced when needed.

4.2 Machine Learning

Machine Learning algorithms are often defined as algorithms that learn from presented data. This usually leads to the question what 'learning' entails. A commonly used definition is given by Mitchell in [Mit97]:

> A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

In the following, a definition of task, performance measure and experience is tried.

4.2.1 The Task T

The task is the problem, which the algorithm is supposed to solve. The most common tasks in Machine Learning are classification and regression. In classification tasks the model has to learn a decision boundary in the problem's data space as shown in Fig. 4.5. Classification tasks are in general setup as *supervised tasks*. Every training sample is characterized by its properties encoded as a feature-vector $\mathbf{x}$ and a ground-truth label y forming a tuple $\mathbf{X}$

$$\mathbf{X} = (\mathbf{x}, y_{true}) \tag{4.2}$$

The predicted label y_{pred}, which is the output of the classifier, is then compared to the ground-truth label to assess the performance and steer the learning process. The data space of the problem is spun up by the features in the feature vector. In the shown example, the distributions align in such a way that the classes can be separated by a straight line. This is not always the case. It is then up to either the analyst (classical Machine Learning) or the model (Deep Learning) to find a suitable data space that allows to find a decision boundary to solve the problem.
Examples for classification tasks in physics are event type classification, source identification (e.g. classifying the type of astrophysical observation) or signal classification (e.g. primary or secondary scintillation in dual-phase noble gas detectors, source or background event, etc.). Note that other tasks like sequence reconstructions can be expressed as classification tasks when the model has to pick the correct sequence out of all possible sequences.
Another large group are regression tasks. Here the model is tasked to give the best estimate for an output value or set of output values with respect to a given input vector $\mathbf{x}$:

$$\mathbf{y} = f(\mathbf{x}) \tag{4.3}$$

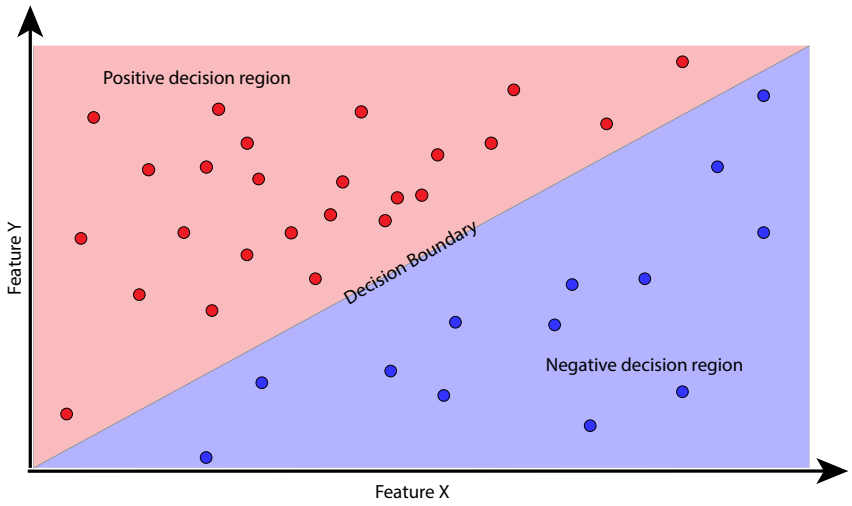


Figure 4.5: Sketch illustrating a linearly separable binary Classification problem. The data space is spanned by two abstract features X and Y. The two distributions are separable by the shown decision boundary.

This is task can be solved either in a supervised or unsupervised way. A well known supervised example is linear regression. Here the algorithm tries to find the best fitting linear approximation of some value y with respect to inputs $\mathbf{x}$:

$$y = f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \tag{4.4}$$

where $\mathbf{w}$ is the weight vector, and b is the bias term. The predicted value y_{pred} is then compared to the known data point of the sample when optimizing the weights during training.
Other tasks include clustering, anomaly detection and denoising. These tasks are in general formulated as unsupervised tasks, i.e. the samples are either their own label (anomaly detection or denoising) or the model is tasked to sort the data set into a given number of 'buckets' (clustering). In the latter case it is then up to the analyst to give the clusters a meaningful label. An example for the application of a clustering algorithm can be found in [SMT23]. In anomaly detection one tries to detect data points that deviate significantly from the average distribution and flag them appropriately. If the number of anomalous samples is statistically marginal compared to the regular samples in the given data set, one can use all samples to learn an average representation of the underlying distribution. Anomalous samples will then still be mapped significantly outside of the average distribution and caught by the model. An example can be found in [Fin+21]. Denoising attempts to reconstruct corrupted input back into the original form. The model learns here the transformation from the corrupted sample

$$\tilde{\mathbf{x}} = h(\mathbf{x}) \tag{4.5}$$

into the reconstructed sample

$$\mathbf{x_{rec}} = f(\tilde{\mathbf{x}}) \approx \mathbf{x} \tag{4.6}$$

Here h denotes the corruption function and f the denoising model. If the corruption function is known or approximated sufficiently one can use clean samples to generate corrupted samples and use the clean one as learning target. A possible application is the denoising of radio images as shown in [Dro+24].

4.2.2 The Performance Measure P

The performance on a task is desired to be as high as possible. The performance measure gives a functional procedure to assess the quality of how well the task is performed. Generally one is only interested in the performance of a given model on unseen data, usually referred to as *test set*, in order to estimate the use-case performance.
The performance on a task has to be determined with a metric fitting to the given task. A widely used metric for classifications is accuracy, which is defined as

$$accuracy = \frac{\text{Number of correct predictions}}{\text{Number of total predictions}} \tag{4.7}$$

Alternatives are, among others, Precision and Recall or F1-score and will be discussed later. Their interpretation is in general straight-forward: the better the classification is performing, the higher the chosen metric scores. In optimization problems, it is more common to *minimize* a given quantity. From accuracy one can derive an error rate as $1-Acc$ and use this to drive the weight optimization[1]. If the output of the classifier is not just a label, but instead a probability for a given sample to belong to some class, i.e. the conditional probability

$$y_{pred} = p(y_{true}|\mathbf{x}) \tag{4.8}$$

one resorts to *binary* (two classes) or *categorical* (three or more classes) cross-entropy. The binary cross-entropy, also known as log-loss is defined as

$$L_{ce}(y_{pred}, y_{true}) = -\log p(y_{true}|\mathbf{x}) = -(y_{true} \cdot \log(y_{pred}) + (1 - y_{true}) \cdot \log(1 - y_{pred})) \tag{4.9}$$

The quantity $p(y_{true}|\mathbf{x})$ can also be interpreted as the likelihood of the label y_{true} given the inputs $\mathbf{x}$. The log-loss is therefore the negative log-likelihood and its minimization equivalent to maximizing the log-likelihood and with that maximizing the likelihood. All machine learning algorithms, which use binary or categorical cross-entropy for optimizing the performance are therefore *Maximum Likelihood Estimators*.
In the case of a regression task, the performance is mostly estimated using the Mean Squared Error on a test set

$$MSE = \frac{1}{N}\sum_{i}^{N}(y_{i,true} - y_{i,pred})^2 \tag{4.10}$$

[1] Note that the error rate is in general a suboptimal minimization criterion as it is defined on a closed interval between 0 and 1. This can lead to exploding and vanishing gradients during especially the early phases of training, when the model guesses an extremal set of weights.

where N is the number of samples in the test set, $y_{i,true}$ is the ground truth-value of the ith sample and $y_{i,pred}$ is the predicted value for the ith sample. As can be seen from the definition it gives the averaged squared euclidean distance between the ground-truth and predicted values in the problem's output data space.
In the unsupervised regression task, the model usually is asked to regress to the original input values (e.g. anomaly detection) or the uncorrupted original input (denoising). It can be shown, that minimizing the MSE is equivalent to maximizing the likelihood function under a Gaussian noise distribution (e.g. Chapter 4 Section [BB23] or Chapter 5.5 in [GBC16]). A very useful property is the parabolic shape. This guarantees the presence of only a global minimum, which makes MSE and its derivatives one of the most used metrics to drive gradient descent optimizations in regression tasks.
The final performance assessment has to be done on *unseen* data, i.e. data not used to optimize the model. In order to reflect use case of the model, the composition of the *test set* has to be broad enough to form a sufficient representation of the whole possible input space. At the same time, the number of samples has to be large enough to allow for statistically relevant error estimates. When using synthetic data, a difference in size of multiple orders of magnitude between training and test set is not uncommon.

4.2.3 The Experience E

The training experience is the processing of a prepared data set in order to optimize the autonomous decision. The main choices are the representation of the data (binned data, feature vectors, sequences of variable length, etc.), the form of feedback given to the learning system and the data selection used to generate the training data set.
Physical detectors are able to give a binned representation of a given event by the discrete nature of sensor placement and the usually binned output values. This offers the possibility to present data either as an event 'image' or as feature vector, that tries to encode all the information in higher level attributes like number of hits, deposited energy, average deposited energy, average distance of hits, etc.
The data selection has to achieve a compromise between the amount of data that has to be presented (the smaller the data set, the faster the model completes the training) and coverage of the input data space (the more data space is covered, the better the overall performance on unseen data). Since a complete coverage of the input-space is in general not achievable, one has to restrict the training set to relevant data populations.

It is common to split the data available into several partitions:

- the *training set*: the data on which the actual training is performed. It has to be devoid of selection biases.
- the *validation set*: a portion of the training set, that is not used for training which is evaluated against the present state of the model to track training progress. This data set is usually only a small fraction of the training set and should at least be stratified on the very basic properties of the data set.
- the *test set*: the data used for assessing the final model performance. It consists of unseen samples that should sufficiently cover the data space . If the training and evaluation is done on synthetic data, test sets can outgrow training set significantly in order to reduce the influence of statistical fluctuations to a minimum.

The form of feedback to the learning system can be either direct (assessment is correct or incorrect) or indirect. When considering sequence reconstruction as task, direct feedback tells the model after every step of reconstructing the sequence, whether the chosen continuation is correct. In the indirect case, the learner has to deduce from the feedback on the whole reconstructed sequence, how to achieve the positive outcome. In general, learning with direct feedback is easier to achieve than in the indirect case.

4.3 Training models on Data

Training a Machine Learning algorithm on data is very akin to the fitting of data to a model. The data points are used to derive the optimal set of weights (free model parameters) of a given model function by minimizing a loss-function. Assuming a polynomial model of the form

$$y(x, \mathbf{w}) = \sum_{i=0}^{M} w_i x^i \tag{4.11}$$

the number of parameters M corresponds to the *model complexity*. Since M is of the choice of the analyst, it is a so called *hyperparameter*.
Let the data be generated by a function

$$f(x) = sin(x) + \epsilon \tag{4.12}$$

where ϵ corresponds to imposed Gaussian noise. The task of the model is to predict the correct value of the underlying sinus function for any given value of x. The error can be determined by

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} (y_{pred}(x_n, \mathbf{w}) - t_n)^2 \tag{4.13}$$

Here $\mathbf{w}$ is the set of weights, x_n is the nth training sample and t_n is the nth target value. Since E is quadratic in $\mathbf{w}$, the derivative of E will be linear in $\mathbf{w}$. For this model therefore exists a unique set of weights $\mathbf{w}^*$ that minimizes E. The result for a polynomial of degree 0, 1, 3 and 9 is depicted in Figure 4.6.
For $M = 0$ and $M = 1$ the model is not able to catch the features of the underlying sinusoidal distribution. This behavior is called *underfitting*. For $M = 3$ the model fits the data reasonably, while at $M = 9$ the model is trying to follow the data excessively in a behavior called *overfitting*. Both over- and underfitting results in a model that leads to suboptimal performance on unseen data. Underfitting can be improved by adding model complexity, in this case increasing M. Increasing M from 0 or 1 to 3 improves the resulting model considerably. Overfitting can be mitigated by either lowering the model complexity, increasing size of the data set used for training or applying regularization. As seen before, reducing M from 9 to 3 yields a significantly better model. The effect of a larger data set is shown in the left plot of Figure 4.7. Here the data set increased from 10 to 100 points yielding a better constrained model at $M = 9$. Regularization is a technique to control overfitting without explicitly reducing he model complexity.

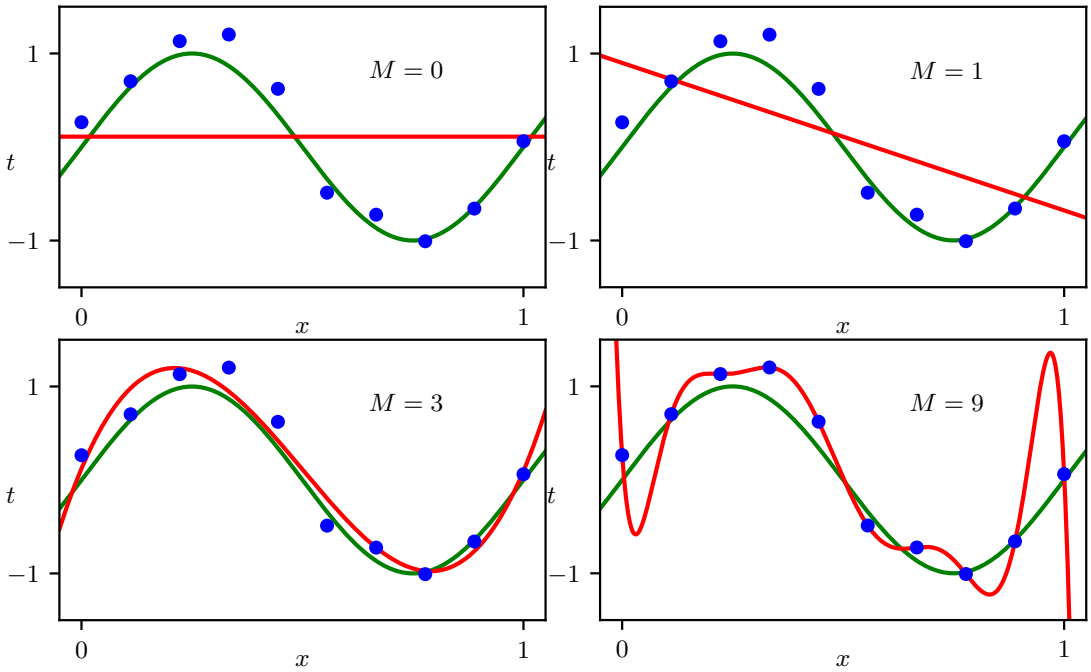


Figure 4.6: Training results for a data set with 10 points for a polynomial of degree 0, 1, 3 and 9. From:[BB23].

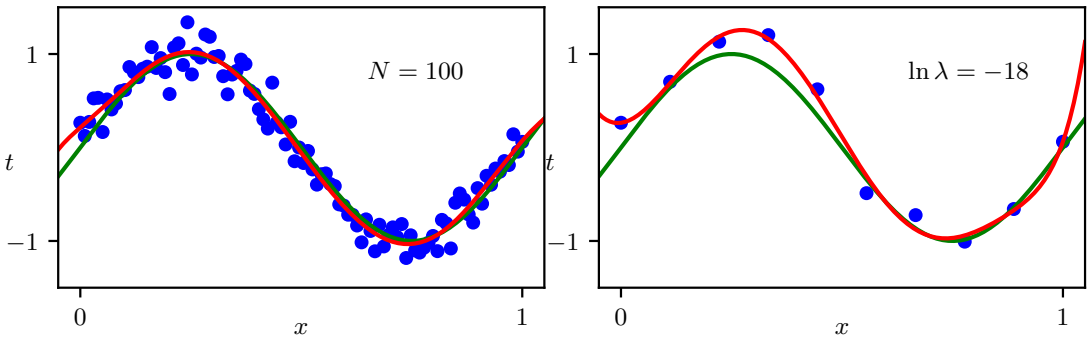


Figure 4.7: Effect of a larger training set and regularization on overfitting. From [BB23].

A very common approach is L2 regularization, which is uses a penalty term in the form of

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}(y(x_n, \mathbf{w}) - t_n)^2 + \frac{\lambda}{2}\|\mathbf{w}\| \tag{4.14}$$

where t_n is the nth target value, $y(x_n, \mathbf{w})$ is the nth predicted value and λ is the regularization strength. Since the choice of λ depends on the task to be solved, it is an additional hyperparameter. The right plot of Figure 4.7 shows the result of the model fit for $\log \lambda = 18$. Using the original data set, the overfitting is significantly reduced. Other relevant regularization techniques are discussed where applied.
The chosen example is of course oversimplifying the task in real world applications considerably. There the model to be fitted to the data, the target function, is in general unknown, such that learning a sufficient approximation of the target function is the crucial part of solving the task.

4.4 The Perceptron

A sub-discipline of Machine Learning tries to model the working principle of neural cells in code, thereby creating what is called a *artificial neuron*. The class of models made of these as basic building blocks are called *artificial neural nets*. Its earliest implementation is the Perceptron invented by McCulloch and Pitts [MP43] 1943 and first realized as hardware network by Rosenblatt [Ros58] in 1957 . A sketch of the implementation is shown in Figure 4.8. It consists of an input layer and a set of output nodes that are connected via weights reaching from the ith input node to the jth output node. Each input node represents an entry of an input vector of the given sample. The output nodes represent the result of the task that the Perceptron is given to do. Since the output nodes are directly connected to the input nodes, networks employing this topology are also called shallow neural network. Processing in each node is done in two steps. First the inputs into each node are aggregated in a

weighted sum

$$y_j = \sum_i a_{ij} \cdot x_i + b \tag{4.15}$$

. This sum is then used to calculate the value of the *activation function*, forming the output of this node. In the original publication the network was tasked with image classification. The output of the jth node was determined by taking the weighted sum of all input components and evaluating this against a Heaviside step-function giving

$$output_j = H(y_j) = H\left(\sum_{i=1}^{N} a_{ij} \cdot x_i + b\right) \tag{4.16}$$

for the j-th output node. This yields 0 if $y_j < 0$ and 1 for $y_j \geq 0$ Here a_{ij} are the weights connecting the i-th input component to the j-th output node, b is a bias-term that the model adopts to the data presented during training[2] Weight adaptation is performed with the Perceptron-Learning-Rule in two steps. First, the output for a sample is determined. If the output is correct (e.g. 1 for ground truth 1 or 0 for ground truth 0) no adaptation is applied. If the output differs, the respective nodes are adjusted after the k-th step according to

$$a_{ij}(k+1) = a_{ij}(k) + \Delta y \cdot \alpha \cdot x_i. \tag{4.17}$$

Δy is either 0 (ground truth equals output), -1 (ground truth 0, output 1) or 1 (ground truth 1, output 0), α is the learn rate set to a value greater zero. The classical Perceptron is capable of learning tasks like simple logic operations (AND, OR, NOT). Possible realizations are shown in Figure 4.9. As can be shown (see e.g. [Agg18]) single layer Perceptron networks can be used to generate models which behave equivalent to classical Machine Learning algorithms by choosing appropriate activation functions and adapting the loss function accordingly.
Classical Perceptron Networks reach their limits when trying to emulate more complex operations like the Exclusive OR (XOR). This can be seen by going through the possible input pairs. While it is possible to find weight and bias combinations to give the correct output for (0,1), (1,0) and (0,0) it is impossible to map the input (1,1) correctly to False while retaining correct results on the other three. The reason for this is the incapability of the classical Perceptron to separate two distributions with a decision boundary more complex than a straight line [Sej18]. This can be seen inspecting the data space for simple binary logical operations as shown in Figure 4.10.
Modifications to ramify this are for example adding intermediate layers (going to a *Multi-Layered* Perceptron), modification of the input variables (feature engineering) or more complex activation functions.

[2] In almost all software implementations b is realized as an additional input x_0 which is set to 1. The magnitude of b is then set by the weight a_{0j}.

Figure 4.8: Sketch of the the working principle of the Perceptron (left): The inputs are agglomerated in a weighted sum and then handed of to an activation function f. On the right: the realization of the Perceptron as hardware network. The adjustable weights were realized by potentiometers stored in the center cabinet. Courtesy of the National Museum of American History.

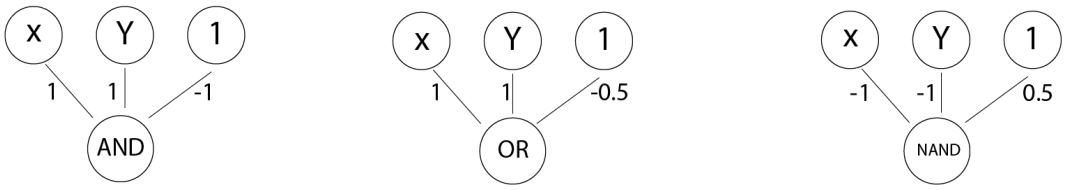


Figure 4.9: Possible Perceptron realizations of AND, OR and NOT AND. Data flows from top to bottom. The given weights realize the respective operation for the step function as activation function.

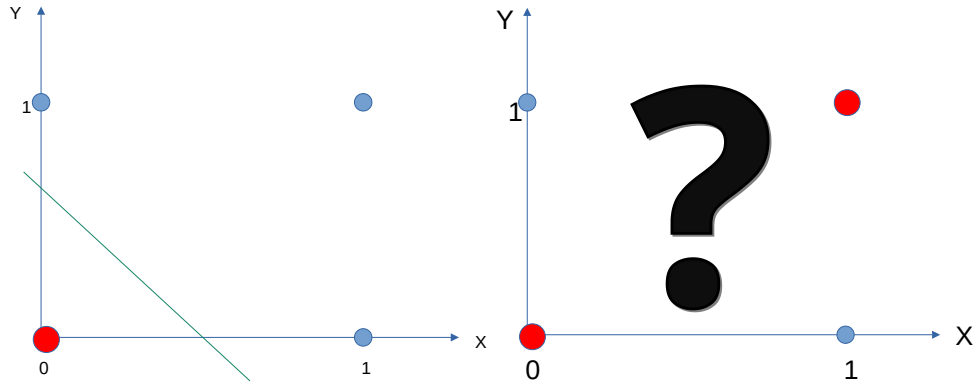


Figure 4.10: Input data space for the OR and XOR binary logic operations. The possible input tuples in this case are (0,1), (1,0), (0,0) and (1,1). For OR (0,1), (1,0) and (1,1) represent True (in blue), while (0,0) evaluates to False (red). In this case, the input tuples can be separated by the green line. In the case of XOR, also (1,1) evaluates to False. It is therefore not possible to separate the two distributions with a simple line, making it impossible for the Perceptron to perform the task.

4.5 The Multilayer Perceptron

Multilayer perceptrons are a special form of artificial neural networks, that allow the transformation of input data into an intermediate representation that benefits solving the task. This is achieved by stacking multiple layers on top of each other and connecting each node of the predecessing layer with every node of the successive layer. Layers that are neither an input layer or an output layer are called *hidden layers*. Models with more than two hidden layers are generally considered *deep*. The hidden layers store an intermediate representation that is used to perform the given task. The ability of *learning* an appropriate representation of the input data is what allows deep learning models to outperform conventional AI and Machine Learning algorithms like Decision Trees, where the analyst has to find the best suited representation for the given problem. A possible realization of a Multilayer erceptron for the XOR-problem is given in Figure 4.11.
When using the hidden representation of this example, we can see that each axis now represents a tuple of inputs (x_1, x_2). The input tuples (1,0) and (0,1) are now mapped to the same point in the data space, resulting in a distribution, that is again linearly separable .
This demonstrates, that MLPs are able to solve non-linear tasks. Later, [HSW89] showed, that an MLP with one hidden layer and non-linear and non-polynomial activation functions is capable of approximating all continuous functions in a relevant domain given a a sufficient number of nodes. They are therefore universal function approximators. Despite being one of the first deep network architectures, MLPs are still widely used for vectorized data.
The choice of activation function is of course not limited to the step-function. Other

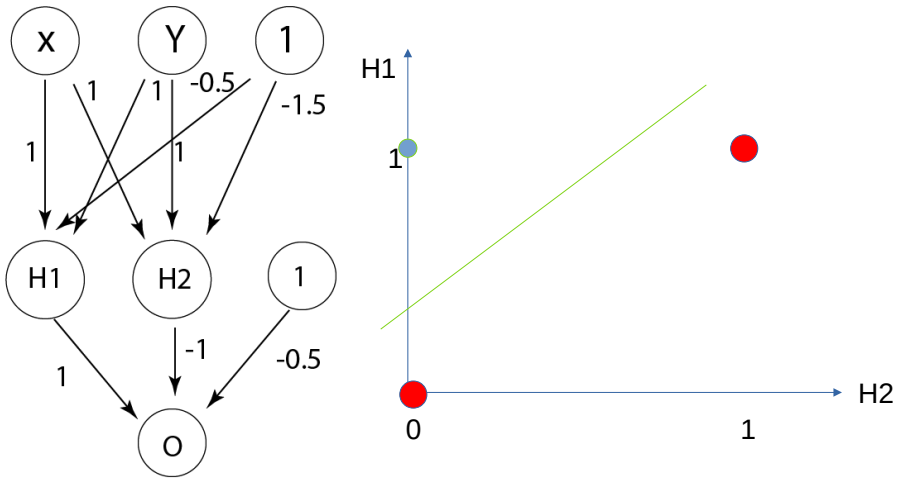


Figure 4.11: Left: Sketch of a possible realization of XOR with a Multilayer perceptron. Right: Hidden representation using H1 and H2 as input variables. The green line corresponds to the decision boundary, the blue dot represents output 1, the red dots represent output 0. Note that in the intermediate representation, the original inputs (1,0) and (0,1) are mapped to the same point in the data space, making the distributions linearly separable again.

common choices are the Rectified Linear Unit (ReLU), the hyperbolic tangent function (tanh), the sigmoid function and the softmax function (as output for multi-class classification).
A systematic catalogue of activation functions can be fund in [KK24].

4.6 Gradient Descent

Even though it is possible to optimize the weights directly in some cases, it turns out, that an indirect approach, using the gradient of the error function, is more efficient. The values of the error function can be viewed as surface in a data space spun by the weights, sometimes called *weight space*. A small change in the weight vector $\mathbf{x}$ leads, in first order approximation, to a small change in the value of the error function:

$$\delta E(\mathbf{w}) \approx \delta \mathbf{w}^T \nabla E(\mathbf{w}) \tag{4.18}$$

∇E is pointing into the direction of the steepest decline of E. The points in weight-space where

$$\nabla E = 0 \tag{4.19}$$

are called stationary points and correspond either to minima, maxima or saddle points in the error surface. Minima and maxima can either be local or global extrema of $E(\mathbf{w})$. The optimization process is performed iteratively with randomized

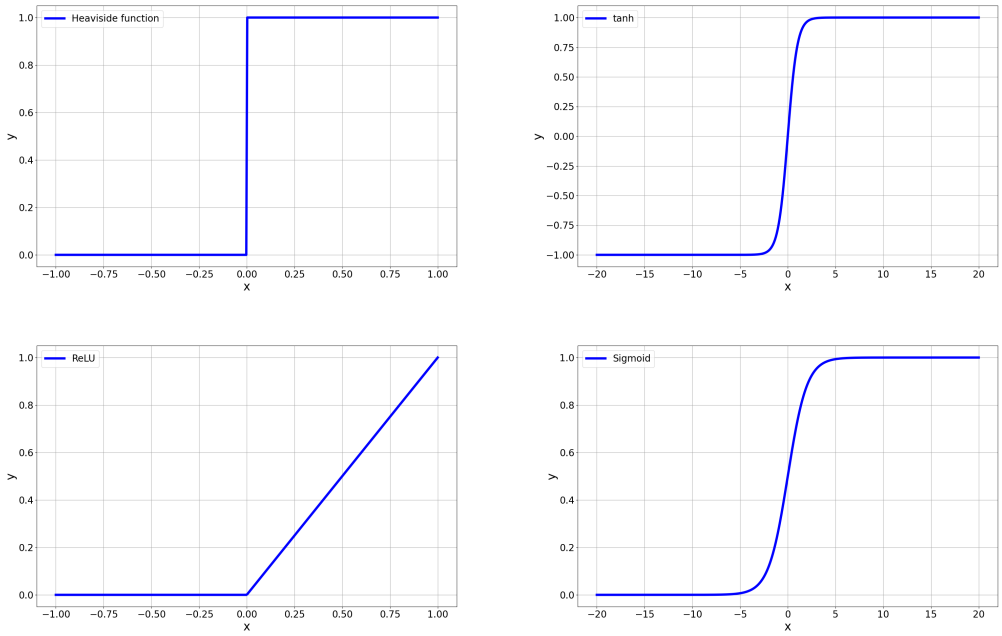


Figure 4.12: Shape of various activation functions commonly used in deep learning applications.

initial weights $\mathbf{w}^{(0)}$:

$$\mathbf{w}^{(k)} = \mathbf{w}^{(k-1)} + \Delta\mathbf{w}^{(k-1)} \tag{4.20}$$

$\Delta\mathbf{w}^{(k)}$ is called the weight upgrade applied at the k-th update step. Its actual form depends on the used optimization algorithm driving the gradient descent. The most common approaches use a form of *stochastic gradient descent*. The data set is either partitioned into small equally sized chunks called mini-batches or used sample-by-sample. The calculation of the error function then breaks down into

$$E(\mathbf{w}) = \sum_{i=1}^{N} E_n(\mathbf{w}) \tag{4.21}$$

with N being the number of mini-batches or samples respectively. The weight update-rule then becomes

$$\mathbf{w}^{(k)} = \mathbf{w}^{(k-1)} - \eta \nabla E_n(\mathbf{w}) \tag{4.22}$$

A complete pass through the training data set is called *epoch*. To prevent accidental correlations in the presented data, the order of the samples is usually shuffled after each epoch. $\eta > 0$ is called the *learning rate*, whose size has to be determined empirically, a *hyperparameter*. The choice of η is crucial to ensure convergence. A large value ensures quick initial convergence, but can lead to an oscillating behaviour as shown

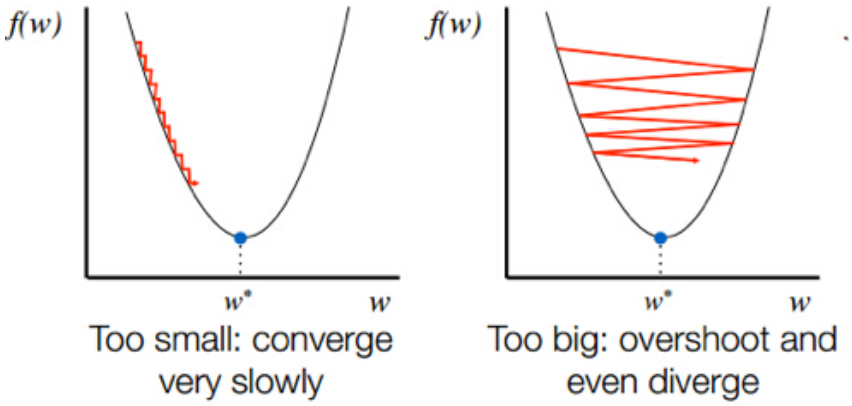


Figure 4.13: Effect of too large and too small learning rate. From[VD18]

in Figure 4.13. A small learning rate can result in either a very slow convergence or trapping of the optimizer in a local minimum.
Ideally one desires a larger learning rate at the beginning of the training process and a lower value when approaching the minimum. One method to ensure this behaviour is to *schedule* the learning rate. A direct scheduling would be training for n_0 epochs at η_0 followed by n_1 epochs using learning rate $\eta_1 < \eta_0$ and so forth. Another approach is to decrease the value of η gradually using decay. The learning rate becomes then a function of the update steps:

$$\eta(k) = \eta_0 \cdot f(k) \tag{4.23}$$

Typical choices for the decay function are linear, power law or exponential decay. The strength of the decay is again a hyperparameter.
In order to obtain better control of the learning process, it is common to control the learning rate for each weight separately. This is achieved by applying an adaptive gradient by using either the accumulated sum of the squares of all derivatives calculated for the ith weight w_i of the weight vector (this is called *AdaGrad* for adaptive gradient [Zei12]) or the exponentially weighted mean of all derivatives of the given weight (RMSProp for Root Mean Square Propagation [Tie12][3]). The weight update has then the form of

$$w_i^{(k)} = w_i^{(k-1)} - \frac{\eta}{\sqrt{r_i^{(k)}} + \delta} \left(\frac{\partial E(\mathbf{w})}{\partial w_i} \right) \tag{4.24}$$

δ takes here the role of a stabilization factor to prevent accidental division by zero. In the case of AdaGrad $r_i^{(k)}$ takes the recursive form of

[3]Note that even though RMSProp was and still is widely used, it was only published as part of lecture notes.

$$r_i^{(k)} = r_i^{(k-1)} + \left(\frac{\partial E}{\partial w_i}\right)^2 \tag{4.25}$$

while for RMSProp it is defined as

$$r_i^{(k)} = \beta \cdot r_i^{(k-1)} + (1-\beta) \cdot \left(\frac{\partial E}{\partial w_i}\right)^2 \tag{4.26}$$

Another approach is to add information of the second derivative. This is achieved by defining the weight update in 4.20 as

$$\Delta \mathbf{w}^{(k-1)} = -\eta \nabla E(\mathbf{w}^{(k-1)} + \mu \Delta \mathbf{w}^{k-2} \tag{4.27}$$

which now also incorporates the rate of change of the steepest decline. It is therefore called *momentum* and μ is called the momentum parameter.
The most commonly used optimizer is called *Adam* (derived from adaptive moment) [KB17] and combines RMSProp and momentum. It stores the momenta information for each weight by using exponentially weighted means for the respective sums and quadratic sums of the calculated gradients. The update rule then takes the form of

$$w_i^{(k)} = w_i^{(k-1)} - \eta \cdot \frac{s_i^{(k)}}{\sqrt{r_i^{(k)}} + \delta} \tag{4.28}$$

s and r are defined as

$$s_i^{(k)} = \left(\beta_1 \cdot s_i^{(k-1)} + (1-\beta_1) \cdot \frac{\partial E(\mathbf{w})}{\partial w_i}\right) \cdot \frac{1}{(1-\beta_1^k)} \tag{4.29}$$

and

$$r_i^{(k)} = \left(\beta_2 \cdot r_i^{(k-1)} + (1-\beta_2) \cdot \left(\frac{\partial E(\mathbf{w})}{\partial w_i}\right)^2\right) \cdot \frac{1}{(1-\beta_2^k)} \tag{4.30}$$

Please note that in the denominators of s and r the hyperparameters β are to the power of k. The default values for β_1 and β_2 are 0.9 and 0.99 respectively. Since Adam, in general, leads to stable convergence and reasonable results it is one of the most widely used optimizers for deep learning applications.

4.7 Backpropagation

One of the key problems when training a deep neural net is the efficient calculation of the gradient used to drive the weight optimization. Here, back-propagation [RHW86] is applied to calculate the gradient which is then used to determine the weight-adaptions performed by the gradient descent optimizer.

This is done in two phases. In the first phase (forward phase) the batch of samples is processed in forward direction (from the input layer towards the output nodes) and the error is estimated using the chosen error-function with the current set of weights. For each node in the network the pre-activated value of the node j is calculated by the weighted sum

$$a_j = \sum_i w_{ji} z_i \tag{4.31}$$

Each Z_i is the post-activation value of a preceding node or the original input value:

$$z_i = h(a_i) \tag{4.32}$$

where h is the activation function of the node.
In order to calculate the weight adaptation, the derivative of the loss with respect to all weights has to be computed. This is done in the second phase (backward phase) by applying the chain rule. Since the value of the error function of the nth sample depends only on the summed input a_j to the unit j, one can write

$$\frac{\partial E_n}{\partial w_{ji}} = \frac{\partial E_n}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}} \tag{4.33}$$

Setting

$$\delta_j = \frac{\partial E_n}{\partial a_j} \tag{4.34}$$

and observing that

$$\frac{\partial a_j}{\partial w_{ji}} = \frac{\partial}{\partial w_{ji}} \left(w_{ji} z_i \right) = z_i \tag{4.35}$$

one obtains

$$\frac{\partial E_n}{\partial w_{ji}} = \delta_j \cdot z_i. \tag{4.36}$$

The δ_j are also called errors of the jth node. Using the chain-rule on δ_j gives

$$\delta_j := \frac{\partial E_n}{\partial a_j} = \sum_k \frac{\partial E_n}{\partial a_k} \frac{\partial a_k}{\partial a_j} = h'(a_j) \sum_k w_{kj} \delta_k \tag{4.37}$$

where the sum runs over all nodes that j is feeding input to. In order to calculate δ for every hidden node in the model, one has to propagate all the previous δ backwards from the preceding layers. The gradient for a specific weight i at a given node j then becomes

$$\frac{\partial E_n}{\partial w_{ji}} = h'(a_j) \sum_k w_{kj} \delta_k z_i \tag{4.38}$$

which reduces to evaluating the errors δ, as z_i can be read of the output of the activation function of the ith node.
Implementations in Machine Learning frameworks are using AutoDiff (e.g. [Bay+17]) in order to achieve a well performing implementation.

4.8 Convolutional Neural Networks

MLPs assume no intrinsic structure in the presented data. This is apparent in the vectorized representation. Every component is assumed to be independent from the other. Interdependencies between components are found by correlating every input vector component into a respective layer with every processing node of that layer. A sizeable portion of physical data and very important cases in real-world applications, however are described by structured data with strong interdependencies between neighbouring data points. Here, the order of the input values has a profound impact on the context of the data. Examples are words in a sentences or values in a waveforms (1D data), images or sensor values derived from a X-Y sensor array (2D data) and videos or 3D detector data from e.g. a Time Projection Chamber ([Apr+08] as an example). Even though it is possible to use an MLP-approach on this type of data, building correlations between all grid points would lead,in general, to many false correlation and to a model that is overly sensitive to changes of the initial conditions. CNNs have successfully circumvented this issue by looking explicitly for local dependencies between data points. Convolutional Networks are comprised mainly out of *convolutional layers* and *pooling layers*. Convolutional layers try to extract the relevant features from the presented input. Exploiting the grid-like nature of the data and the features to extract, they slide a filter, the *kernel*, over the input data with a given stride in the dimensions of the input data, as depicted in Figure 4.14. In thee 2D case, the convolution is achieved by applying

$$C(j,k) = \sum_l \sum_m I(j+l, k+m) K(l,m) = I * K \tag{4.39}$$

where I is the input data with size (j,k) and K is the kernel with dimensions (l,m). Note that this operation is strictly a *cross-correlation* as it is not commutative.
The working principle is illustrated in Fig. 4.14 left. The filter is slid over the input patch and propagates the filter weighted sum of the observed patch to the feature map. In the given example, applying the filter results in the feature map being smaller than the original input. Assuming an input size of $L \cdot M$ pixels, a stride of $(1,1)$ (meaning the filter moves at each step one pixel further in x-direction and 1 pixel down in y when reaching $x = M - (J-1)/2$ and revert back to $x = J - 1)/2$), the size of the feature map resulting from a convolution with a filter of size $J \cdot J$ is

$$MapSize = (L - J + 1) \times (M - J + 1) \tag{4.40}$$

This is called *valid padding*, because the layer only operates values present in the original input. It is in some cases desirable to have the feature map and the input

map at the same size. One can then introduce additional rows and columns around the original input that form a *padding*. The value of the padded pixels is usually zero. The amount of padding depends on the size of the kernel. The input size then becomes:

$$InputMapSize = (L - J + 1 + 2P) \times (M - J + 1 + 2P) \tag{4.41}$$

with P being the padding calculated as

$$P = (J - 1)/2 \tag{4.42}$$

Since this padding scheme is preserving the original input size, this approach is called *same padding*. On each pixel of the generated feature map, the activation function is applied. A very common activation function in this context is ReLU. After evaluating each pixel value with respect to the activation function, one obtains the activation map, which is then further processed by the next layer. Every convolutional layer introduces the kernel size and the number of filters as tunable hyperparameters.

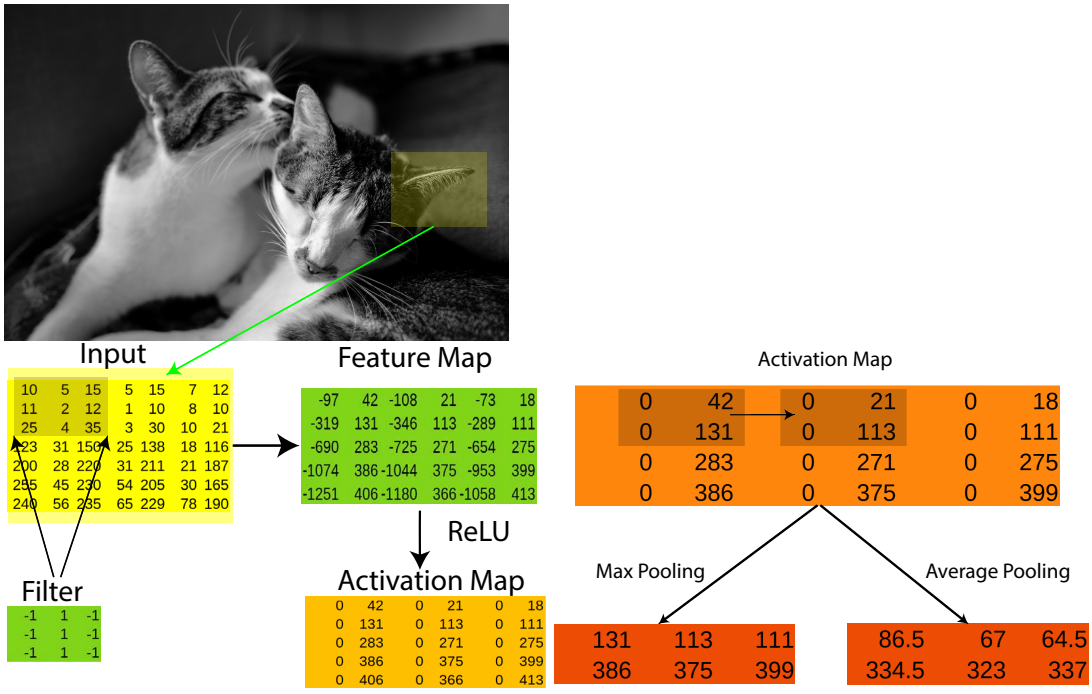


Figure 4.14: Sketches illustrating the Convolution and Pooling operation.

A convolutional layer is often followed by a *pooling layer*. The goal of a pooling layer is to ensure translation invariance in the feature detection and hierarchical learning of more and more abstract features by sampling down the activation map in between convolutional layers. The down-sampling is performed by again sliding a window

over the input and then either propgagating e.g. the maximum value in that window (*max-pooling*, [Yam+90]) or the average pixel value of all pixels under the window (*average pooling*). The amount of downsampling is controlled by the window size and the stride applied.
The right side of Figure 4.14 shows an example. Here a 2×2 window is slid over the input. The stride of $(2, 2)$ results in a processed map of half the size. Each pooling layer introduces the window size and the stride as tunable hyperparameter.
Common tasks for CNNs are image classification, object detection (finding and tracking relevant features over time, see [SSC24]) or image segmentation (pixel by pixel classification of image data, see [Bra+25]).

Chapter 5

Event Classification

5.1 Introduction

Classification of events in space-borne detectors can be done in two ways, either directly (online) in the data pipeline from the detector to down-link, or after data transmission in a ground based data center (offline). This chapter will explore the possibilities and the effect of online classification. Offline classification has infrastructural advantages over the online case. The main advantage is the amount of available computing power, which in this case is rather bound by financial limits. This allows for more powerful models that quickly reach demands impossible to satisfy with on-board computers. The task that a model has to fulfill is the same as in the offline case:

Tag each event by type so that it can be processed appropriately.

This chapter starts with presenting the simulation tools and how to progress from simulation data to event images. It then continues with discussing the architectural choices made for two classification approaches and concludes with the presentation of the results on test sets.

5.2 MEGAlib and Event Data

MEGAlib [Win+03b] is a simulation suite for low to medium energy gamma-astronomy. Its simulation framework uses Geant4 and allows the quick assessment of detector concepts and background assessments. It also offers a reconstruction engine, basic analysis tools for performance estimates and Compton imaging as well as a data pipeline for real-time analysis of real detector data. MEGAlib is used to generate the raw event data that is used to train and test the AI-models generated in this study. The most used parts for simulations and data analysis are shown in Figure 5.1. Geomega transforms the given geometry into a Geant4-model while Cosima manages source placement and event handling. The simulation output is generated as text or byte file and then handed to the reconstruction engine Revan. The engine is capable

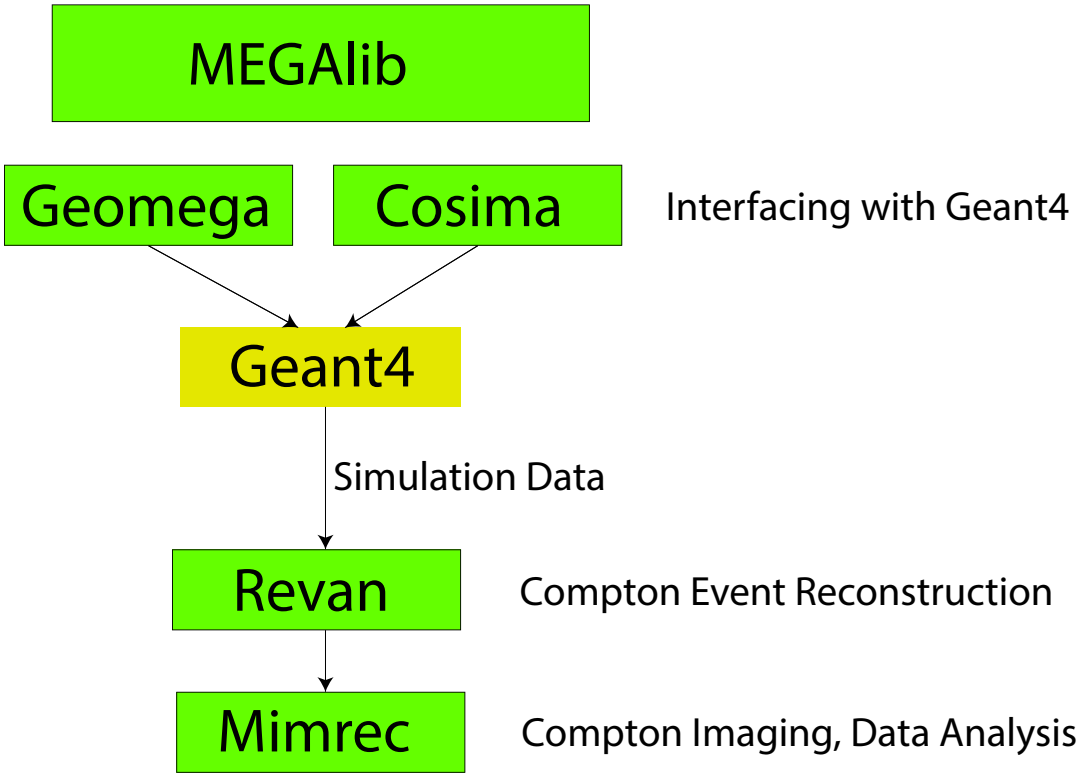


Figure 5.1: Top down view on MEGAlib. The library acts as wrapper around Geant4 and tries to ease the process of generating and evaluating mass models. Cosima and Geomega interface with Geant4 and deal with generating the geometry and managing the actual simulation. The data is then reconstructed with Revan and can be analyzed in Mimrec.

of reconstructing recoil electron tracks and Compton sequences. It is also capable of reconstructing pair events rudimentary. The output of Revan is a again text or byte-file with the reconstructed events. This file can be analyzed with Mimrec. It provides basic plotting tools for spectra, data quality and Compton imaging.

5.2.1 MEGAlib Reconstruction Chain

The MEGAlib reconstruction forms the baseline the newly developed classification algorithms. The reconstruction tries to match the measured deposits in an event with reconstruction algorithms matching different event types sorted by computational complexity as selected in the settings. More details can be found in [Zog06].

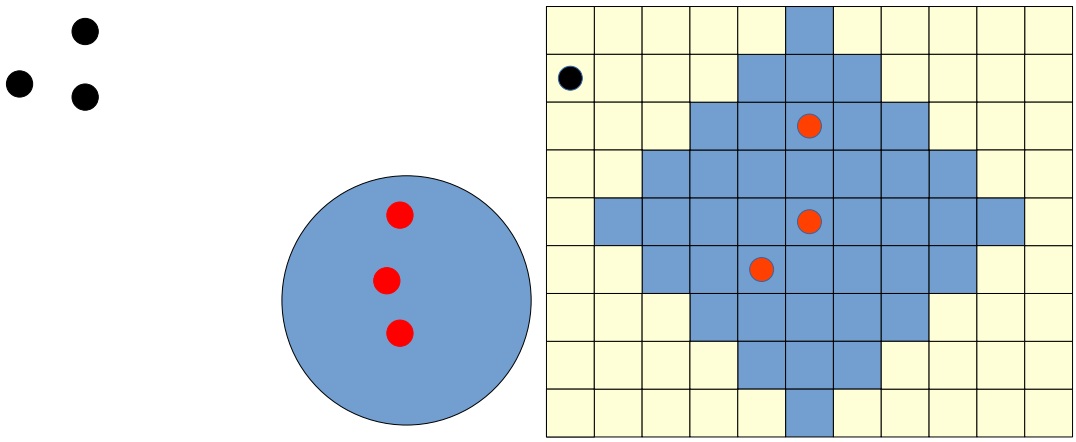


Figure 5.2: On the left: Clustering by distance. All hits in a given radius depicted in blue around a hit are merged together. The black dots are outside of the cluster. On the right: Clustering by adjacent voxels. Here all hits in n=4 neighboring voxels are to be combined into one hit. The unclustered hit in the top left corner is outside of the clustering area and therefore left out. Mathematically the first method employs the Euclidean-metric, while the second uses the Manhattan- or Taxi-driver-metric to determine the clustering area.

Stage 0: Hit Clustering

Energy deposits detected from photons or electrons in DSSDs or crystal calorimeters are usually observed as interactions of secondary particles created from e.g. the recoil electron. This can result in more than one detected 'hit' in the sensitive volume. The goal of the clustering stage is to remove this hit duplicity and reconstruct the observed hit pattern back to point interactions. MEGAlib provides for two ways of performing this task by using either a given euclidean distance or a number of voxels. The working principle is shown in Fig. 5.2.
The reconstructed point of interaction is calculated using an energy -weighted mean for the x-y-z coordinates:

$$x_i = \sum_j^N x_j \cdot \frac{E_j}{E_{tot}} \tag{5.1}$$

The result of this stage is a list of point interactions, that form the input of the reconstruction stage.

Stage 1: Muon- and Electron-Primary Track Search

In this stage the hit pattern is compared to a straight line by fitting a linear model to the hit pattern.

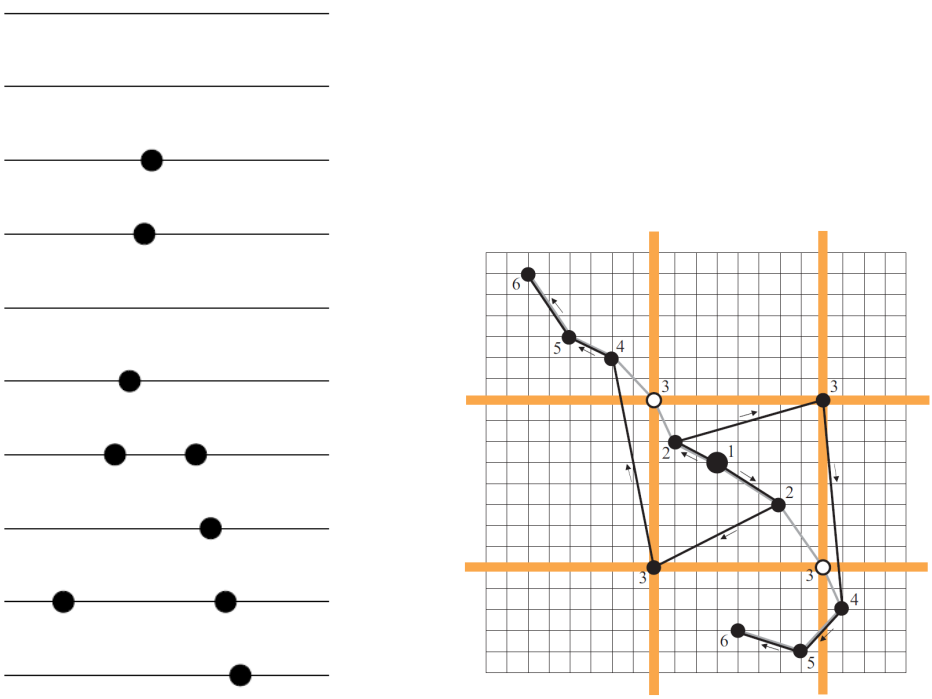


Figure 5.3: On the left: Vertex which fulfills the criteria of the pair reconstruction pattern reconstruction. On the right: Top view of a track with wrong interaction mapping due to a faulty clusterization. The pair vertex is located at 1. The white circles represent the correct interaction points in layer 3. The grey line corresponds to the true path of the pair, the black path is the faulty reconstructed one. From [Zog06].

Stage 2: Pair Reconstruction

Pair reconstruction in MEGAlib is implemented as a pattern recognition algorithm. The algorithm looks for a pair vertex from which the characteristic inverted V-shape emerges. A vertex is checked to have the following properties:

- Exactly one hit in one layer
- In the next n layers below the starting point two or more layers contain two hits
- In the m layers above the starting point no layers contain any hits

The parameters n and m can be adjusted in the reconstruction settings.
The left sketch in Figure 5.3 shows an example that fulfills the minimum criteria for n= 4 and m= 2. The less layers that have to be searched to find two layers with exactly to hits the better the quality factor of the vertex.
If a vertex was found the subsequent hits are appended under the assumption that the initial direction was towards the first layer with two hits. The remaining hits in

the chain are then arranged in such a way that they form a line as straight as possible by maximizing the Pearson coefficient(see, e.g., [Mor08]):

$$c_{Pearson}(X,Y) = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y}, \tag{5.2}$$

where X and Y are random variables and σ_X and σ_Y denote their standard deviation. It ranges from -1 to 1, where 1 corresponds to a linear correlation and -1 to a linear anti-correlation. In order to deal with defective layers or strips, the algorithm is allowed to skip a set amount of layers between hits. MEGAlib uses the Pearson correlation for the deposited energy E_{dep} and the angular deviation $\Delta\alpha$ along the track in order to calculate the track quality factor for electron and positron defined as

$$Q_{Pearson} = 1 - \frac{2 + c_{E_{dep}} + c_{\Delta\alpha}}{4} \tag{5.3}$$

For the correct sequence of hits, both correlation factors should be positive and the resulting quality factor close to zero.

Stage 3: Recoil Electron Track Reconstruction

A recoil electron track resulting from a Compton interaction is assumed to have the following characteristics:

- The angular deviation from a straight line $\delta\alpha$ increases from layer to layer due to Moilere-scattering. This leads to a bended shape inside the tracker.
- The energy deposit in the first layer is on average smaller than for the remaining hits, due to the electron being the result of a Compton interaction inside of the given layer. Therefore it does not travel through the whole layer and thereby deposits less energy.
- The deposited energy increases at the end of the track but varies greatly in between, because of the underlying Landau distribution.
- The largest deposit occurs in general in the last interaction site, because the electron deposits all of its remaining energy.

The reconstruction starts with correlating all unambiguous track segments, i.e. the algorithm appends all single hits in adjacent layers (see Figure 5.4 a). To compensate for damaged strips, the algorithm is allowed to skip every other layer.
Once an ambiguity is encountered all possible continuations have to be investigated and evaluated. This can be done in numerous ways, using for example Pearson correlation or a Spearman-Rank approach. Since the Pearson correlation assumes a linear correlation between the investigated variables, the more general Spearman-Rank is to be preferred [Zog].
The Spearman-Rank assumes a monotonous correlation of the investigated quantities. To determine the rank of a element x_i among all X all elements are sorted into

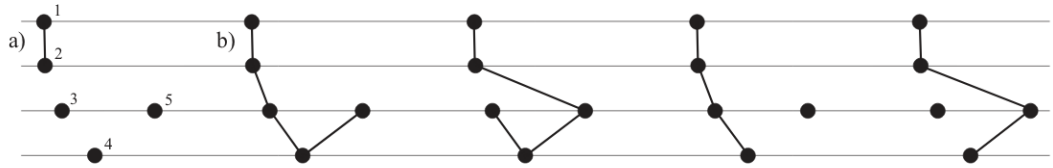


Figure 5.4: A hit pattern how it can emerge from a recoil electron at low energies. In a) the algorithm looks for a track of unambiguous hits and therefore appending hit 1 and 2. From there on all possible continuations have to be investigated, including the possibility of 'hot pixels' (as shown in the last two examples). From [Zog06].

a list from small to large, where rank 1 would correspond to the smallest entry. For given N pairs of measured quantities (x_i, y_i) the underlying distribution of the x_i and y_i is not necessarily known. The idea of a nonparametric or rank correlation is to replace the the value of x_i with the rank of x_i among all x. Here the x_i are the energy deposits and y_i are the scatter angle change $\delta\alpha$. The rank correlation is

$$c_{rank} = 1 - \frac{6D}{N^3 - N} \tag{5.4}$$

with D being

$$D(E_{dep}, i) = \sum_{n=1}^{N} (Rank(E_{dep,n}) - Rank(i_n))^2 \tag{5.5}$$

and

$$D(\Delta\alpha) = \sum_{n=1}^{N} (Rank(\Delta\alpha_n) - Rank(i_n))^2 \tag{5.6}$$

and i being the hit number along the track. With the rank correlation a quality factor can be calculated:

$$Q_{E,Rank} = 1 - \frac{2 + c_{rank}(E_{dep}, i) + c_{rank}(\Delta\alpha, i)}{4}. \tag{5.7}$$

If the assumption of the monotony association of E_{dep} and $\Delta\alpha$ is true, the correlation coefficients are about 1 and the quality factor should yield a value close to 0. Ambiguities (tracks with very similar quality factors) are resolved in such a way that the straighter line is to be favored similar to pair reconstruction.

Stage 4: Compton Sequence Reconstruction

The Compton sequence reconstruction algorithm has to look into all possible interaction sequences. This gives $n!$ possible sequences for every event. The algorithm distinguishes between three types of events:

- Two site events without a track (like in COMPTEL)
- Two site event with a distinguishable track
- Compton event with 3+ interactions

Track in this context means that the hit information is sufficient to decide whether the photon moved up or downwards. For most COMPTEL-like events this differentiation is not possible without additional information. In COMPTEL's case this additional information was the interaction delay between D1 and D2. Since the tracking detector D1 and the absorber D2 are too slow, this information is not available.
The key to distinguish the photon's direction in a two site event without the possibility of time of flight is the measurement of the recoil electron. This allows for a kinematic and a geometric reconstruction of the total Compton scatter(nomenclature analog to Figure 5.5):

$$\cos\theta_{kin} = \frac{E_1(E_2 + E_0)}{E_2\sqrt{E_1^2 + 2E_1 \cdot E_0}} \tag{5.8}$$

$$\cos\theta_{geo} = \frac{\vec{g_2} \cdot \vec{e_2}}{|\vec{g_2}||\vec{e_2}|} \tag{5.9}$$

If the difference between these two angles

$$d\theta = \theta_{kin} - \theta_{geo} \tag{5.10}$$

is sufficiently large then the evaluated ordering is rejected. As a side effect this criterion can also reject incompletely absorbed events that move from D1 to D2 because of the unambiguity of the hits involved. The kinematic reconstruction of these events results in a wrong scatter angle because E_0 is not the sum of all deposits and is hence assumed to have the wrong value. Therefore $d\theta$ is becoming large and the event gets rejected because no ordering of events.
For an event with three interactions the algorithm has to look into six different paths. The angle φ_2 can be reconstructed by kinematics and interaction geometry:

$$\cos\bar{\varphi}_2 = 1 - \frac{E_0}{E_{2+}} + \frac{E_0}{E_{2+} + E2} \tag{5.11}$$

$$\cos\varphi_{2,geo} = \frac{\vec{g_2} \cdot \vec{g_1}}{|\vec{g_2}||\vec{g_1}|} \tag{5.12}$$

Here E_2 is the energy deposit in the 2nd interaction while E_{2+} is the sum of all energy deposits after 2 (here just E_3). E_0 gives the total incident energy of the photon. A sequence is again considered good, if the difference between these two angles is sufficiently small. This approach is limited by the positional and angular resolution of the involved detectors, the energy resolution and Doppler broadening. The Double Sided Silicon Strips (DSSDs) in D1 have a better positional and energy resolution than the crystal array in D2. This can lead to cases where the upward and downward sequence are looking equally plausible. Here again the recoil electron can provide the crucial information in an analogue way as discussed for two site events.

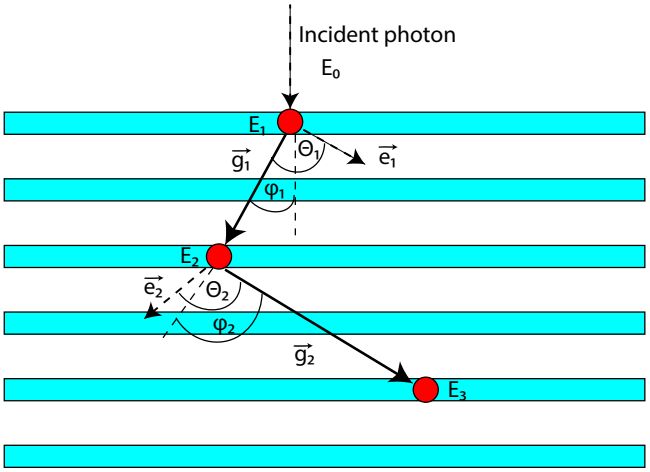


Figure 5.5: An example for a Compton event with three interactions. The scatterings occur at 1 and 2, the final photo effect happens at 3. The second angle ϕ_2 can be reconstructed by the interaction geometry or from kinematics. The recoil electrons provide information about the total scatter angles θ_1 and θ_2 by either Compton kinematics or interaction geometry.

Event Classification with MEGAlib

Event classification in MEGAlib is performed during reconstruction of the data. The possible labels are therefore depending on the chosen reconstruction settings used on the data. With the default settings MEGAlib looks for

- Muon/charged events (resulting in rejection of the event)
- Pair events
- Compton events with either tracked recoil electron (electron leaves the layer) or untracked Compton event (recoil electron stays inside the layer)

The hierarchy is shown in Fig. 5.6. Once a specific algorithm succeeded, the event is tagged accordingly and the next event is processed. If all algorithms fail the event is rejected.
For the model comparision, tracked and untracked Compton events are merged into Compton events, charged particle events as well as pair events created in the Calorimeter (D2-first Pairs) are merged into the non-reconstructible class, because neither of which yields a meaningful reconstructed event. Only D1-first pairs are considered valid pair events.

5.3 From Event Files to Image Data

The simulation output forms the basis of the training data. The output in the use-case here has to be a text file. MEGALib stores the whole Monte-Carlo (MC) informa-

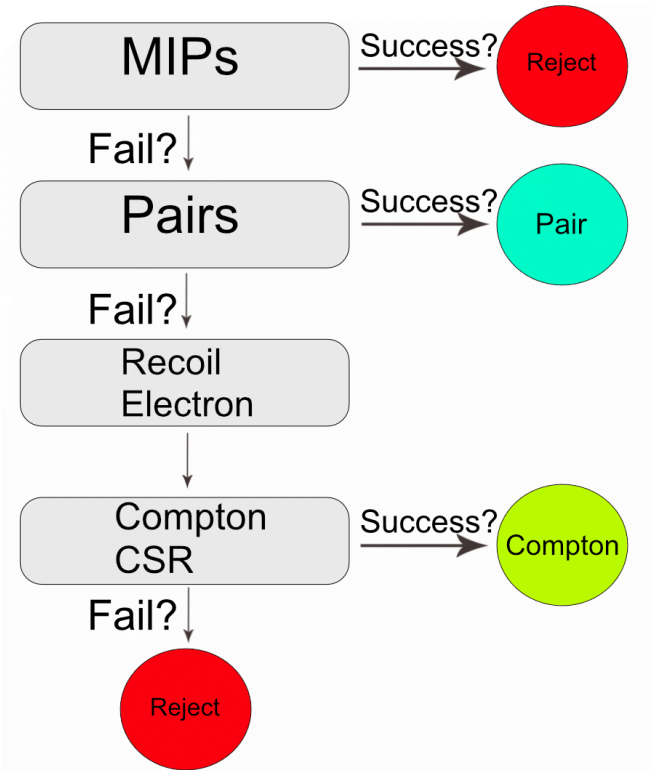


Figure 5.6: Reconstruction hierachy in MEGAlib.

tion only in this format. This information can then be turned into image data. The process is described in the following.

5.3.1 The .sim-file-format and detector simulation

The simulation file is structured into two parts for each event. A short example is shown in Figure 5.7.
The information of each event is stored in three blocks. The header block contains basic information like photon id, event id, time stamp, and deposited energy in sensitive and non-sensitive parts of the detector. The second part, called interaction section contains the MC-information. The first line gives the primary particle with initial energy and momentum. Each interaction in the geometry is stored in a separate line, containing the interaction type which can be used to define a type for the given event. The third section is the detector simulation. The main detector types were 2D-Strip, Calorimeter and Anger Camera. MEGAlib uses simplified models to turn interactions into hits. The main idea for each detector type is illustrated in Fig. 5.8. For the,e.g, 2D-Strip simulation, each hit is cast to the closest x-y-strip junction. The middle of the layer is used as general z-coordinate. This simplification is drastic but necessary since a detailed signal-simulation from channel-waveforms

```
SE
ID 24 3917
TI 0.014886922
ED 430.178
EC 0
NS 137.263
PM CircuitBoard    7.58615
PM Millipore  129.67710
IA INIT  1; 0;0;0.000000000000e+00;   3.29902;   1.94550;   6.31539;0; 0.00000; 0.00000; 0.00000; 0.00000; 0.00000; 0.00000;     0.000;1; 0.07712;-0.44442;-0.89249; 0.00000; 0.00000; 0.00000;   567.441
IA COMP  2; 1;7;1.720224020912e-10;   3.69671;  -0.34642;   1.71272;1;-0.72429; 0.10069;-0.68210; 0.35975; 0.89913;-0.24928;   369.333;3; 0.62834;-0.58414;-0.51378; 0.00000; 0.00000; 0.00000;   198.097
IA COMP  3; 1;0;2.388452907680e-10;   2.24574;  -0.14470;   0.34627;1; 0.98705; 0.15778;-0.02900; 0.15969;-0.94903; 0.27172;   167.738;3;-0.86840; 0.02150;-0.49540; 0.00000; 0.00000; 0.00000;   201.586
IA PHOT  4; 1;7;2.984720919900e-10;   4.01015;   0.13735;   0.29443;1; 0.00000; 0.00000; 0.00000; 0.00000; 0.00000; 0.00000;     0.000;3; 0.94244;-0.32017;-0.09644; 0.00000; 0.00000; 0.00000;   154.303
IA PHOT  5; 1;7;2.984720919900e-10;   4.01015;   0.13735;   0.29443;1; 0.00000; 0.00000; 0.00000; 0.00000; 0.00000; 0.00000;     0.000;1;-0.27303; 0.22998; 0.93411; 0.00000; 0.00000; 0.00000;    11.881
IA PHOT  6; 3;1;3.320736475985e-10;   0.45342;  -0.13104;  -0.34711;1; 0.00000; 0.00000; 0.00000; 0.00000; 0.00000; 0.00000;     0.000;3;-0.33862; 0.50745;-0.79236; 0.00000; 0.00000; 0.00000;     1.569
IA PHOT  7; 5;7;2.985741169073e-10;   4.00931;   0.13805;   0.29729;1; 0.00000; 0.00000; 0.00000; 0.00000; 0.00000; 0.00000;     0.000;3;-0.50485;-0.86250;-0.03486; 0.00000; 0.00000; 0.00000;     5.716
HTsim 1;   0.40000;  -0.20000;  -0.42000;  90.70656;3.31702e-10;3;6
HTsim 1;  -1.00000;  -1.60000;   0.58000; 103.30231;4.68330e-10;3
HTsim 7;   4.16500;  -0.00361;   0.70562; 236.16895;1.72022e-10;2;4;5;7
```

Figure 5.7: Example output for a 567 keV photon. The header section gives the id, interaction time, the deposited energy, escaped energy and deposits in passive material. The interaction section (marked by IA) gives the interaction type (Compton scattering, Pair creation, photo absorption, Rayleigh scattering, Bremsstrahlung), interaction position, as well as momenta before and after scattering. The last section gives the output of the detector simulation. The first keyword gives the detector type (HTsim 1 for 2D-Strip, HTsim2 for Calorimeter, HTsim7 for Anger Camera) followed by the hit position, the deposited energy, the time stamp of the hit and the contributing interactions.

to strip-matching to individual hits depends on the concrete, real existing hardware and is only valid for the specific double-sided strip detector and its read-out chain. The Calorimeter simulation uses the center of a crystal bar as fixed x-y-coordinate. If the calorimeter is depth sensitive, the z-coordinate of the interaction is written out. Anger-cameras give the weighted mean in x-y-z of all interaction in the respective volume as single hit. The deposited energy is then the sum of all hits. The Anger-Camera simulation allows for noising of the x-y-z coordinate, while in the case of the Calorimeter class only the z component is subject to noise. Note, that the x-y strips for 2D strip detectors are assumed to be discrete and are therefore not subjected to coordinate noising.
The hit section of the simulation file is used to generate the images. The information in the interaction section is basis for determining the type of an event.
Event images of are 2D or 3D representations of the interactions inside the detector. They can be generated by identifying x, y and z values with points inside a 3D grid. In the case of e-ASTROGAM, these grid-points are discrete for the tracker. A full 3D representation of a detector can quickly overwhelm the available amount of Video RAM (VRAM). A different approach is the use of projections (see e.g. [**Aurisano2016**]). Its a trade-off between lesser memory demands (2D images are in general smaller than the full 3D representation) for potential information loss due to projecting along the given axis. The design of the e-ASTROGAM tracker makes the usage of X-Z, Y-Z and X-Y projections a possible choice. In order to include as much information as possible, the input should then include multiple projections per event. In addition to the tracker information (one row of pixel per layer) the discretized calorimeter information (e.g., one row of pixel per cm of crystal height) can be included at the bottom of the tracker projection. Each projection is constructed analogues to the 3D view by identifying the relevant coordinates with pixels in the image.

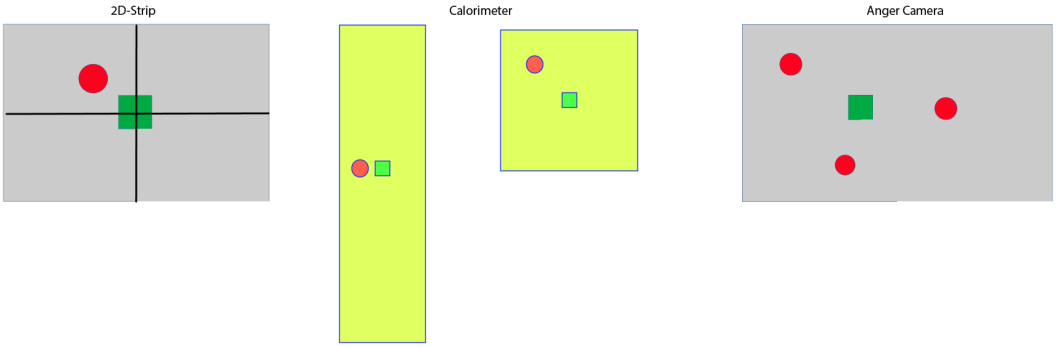


Figure 5.8: Sketches illustrating the working principle of the 2D-Strip (left), Calorimeter (center) and Anger Camera (right) detector simulation. Interactions in the sensitive volume are marked with red dots, the reported hit position is marked by the green square. In the 2D-Strip simulation the hit is cast to the next x-y-junction denoted by the black lines. The Calorimeter simulation casts each hit to the center of the x-y plane of the individual crystal. If the measurement in the crystal is depth sensitive, the z coordinate is written out. Otherwise the center of the bar is used. The Anger-Camera simulation calculates the energy weighted mean of all interactions inside the volume. The result is a single hit that is written out.

5.4 Data Set Composition

The training and test data generation is based on MEGAlib [ZAS06] simulations with the mass model shown on the right in Figure 5.9. The mass model includes a proxy model for the satellite bus (marked as ACT bus[1]) acting as a passive target for impinging particles. Since only five of the six detector sides are covered by the active veto, charged particles traversing the bus or their secondary particles can enter the detector from bellow. The presence of the bus ensures the presence of these events in the data set.
The DSSD and calorimeter crystals are simulated as fixed-grid detectors, where the smallest unit is given by the strip-pitch and the crystal size respectively. This results in a X-Y position resolution of the DSSD and Calorimeter of

$$\Delta_{XY,DSSD} = \frac{b_{DSSD}}{\sqrt{12}} = 69\mu m \tag{5.13}$$

and

$$\Delta_{XY,Calo} = \frac{b_{Crystal}}{\sqrt{12}} = 0.14mm \tag{5.14}$$

[1] ACT stands for Advanced Compton Telescope. It was a call for mission concepts to design the next generation Compton telescope after CGRO seized operations. The results are summarized in [Bog06]. The bus model is a remnant of these efforts.

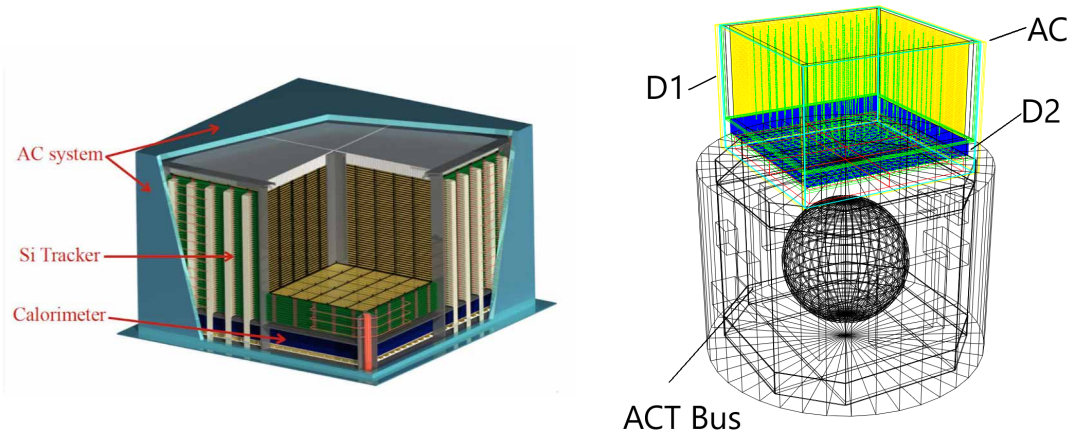


Figure 5.9: Left panel: Schematic view of e-ASTROGAM, consisting of 56 layers of DSSD-tracker (D1) on top of a pixelated CsI(Tl) calorimeter (D2). D1 and D2 are surrounded by an active veto made of plastic scintillator [De +17b]. Right panel: Monte Carlo mass model generated with MEGAlib [ZAS06].

The depth resolution of the individual crystal is assumed to be

$$\Delta_{z,Calo} = 0.2cm \tag{5.15}$$

The assumed energy resolution used for the simulations is shown in Figure 5.10 as function of deposited energy.
In general, one can assume the energy resolution to follow [Leo12]:

$$\sigma_E = \sqrt{(\sigma_{statistics})^2 + (\sigma_{Electronics})^2 + (\sigma_{constant})^2} \tag{5.16}$$

In the case of spectroscopic measurements, the Poison limit is often not the limiting factor as the ionization events are not independent from each other. This leads to a modification of the Poison term by [Fan47]:

$$\sigma^2_{Fano} = F \cdot \frac{E}{w}, \tag{5.17}$$

where F is called the Fano factor, E is the deposited energy and w is the average ionization energy of the detector material. F can also be defined as the ratio of observed variance relative to the expected variance from Poisson statistics.
The input parameters for the data set generation are identical to the proposal modelling in [De +17b]. For silicon in the DSSDs the Fano factor is assumed to $F = 0.125$ with an average ionization energy of $w = 3.64 \cdot 100^{-3}$ keV. The electronic noise is assumed at $\sigma_{electronics} = 1.346$ keV, leading to

$$\sigma_{DSSD} = \sqrt{\frac{F \cdot E}{w} \cdot w^2 + (1.346\,keV)^2} \tag{5.18}$$

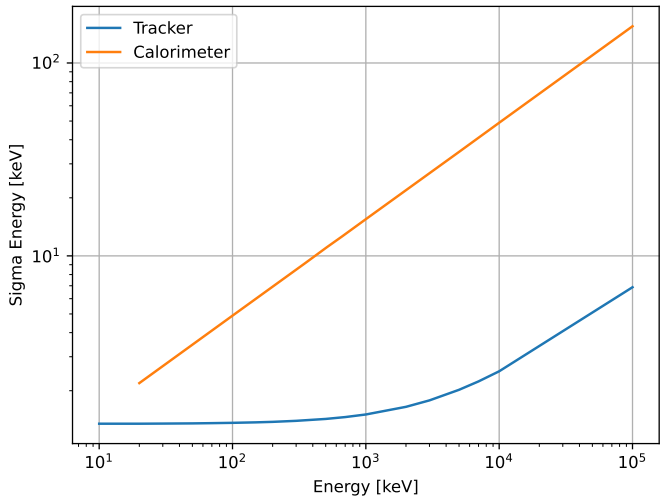


Figure 5.10: Modelled energy resolution as function of deposited energy given in σ of tracker and calorimeter used for the simulations.

The off-set term $\sigma_{constant}$ is assumed to be zero.
For the calorimeter crystals the energy resolution was derived from a test measurement with Cs_{137}. For the photo-peak, the width was determined to $\sigma = 12.6\,\text{keV}$. The total energy resolution is extrapolated by

$$\sigma_{calorimeter} = 12.6\,keV \cdot \sqrt{\frac{E}{662keV}} \tag{5.19}$$

The event topology changes with incident angle and energy. In order to avoid undesired biases in the data sets, the simulations are split into energy and directional bins. Table 5.1 shows the simulated energy and incident energy ranges per event used for the data sets. Since 5 of the 6 sides of the detector geometry are enclosed by an active veto, charged particles can only enter the detector from below without triggering the veto, therefore the angular range for electrons has been reduced to 90-180 degrees.

Table 5.1: Data binning used for dataset balancing and simulation management.

Type	Low energy	Medium energy	High energy	Θ
Compton	0.1 - 1 MeV	1 - 5 MeV	5-10 , 10-15, >15 MeV	0-180°
Pair	10-200 MeV	0.2 - 1 GeV	1 - 3.5 GeV	0-180°
Electron	10 - 100 MeV	0.1 - 1 GeV	1 - 10 GeV	90-180°

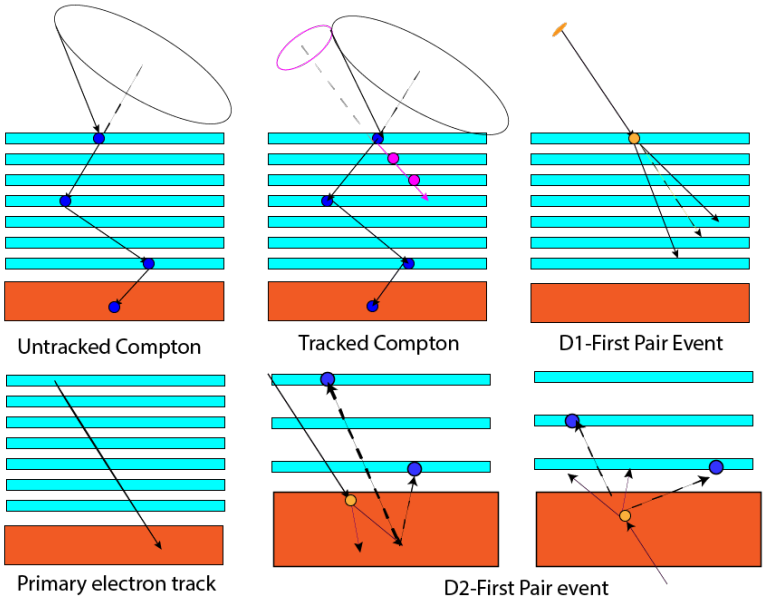


Figure 5.11: Event topologies encountered in Compton/Pair Telescopes. Compton scatters can occur without and with a trackable recoil electron (top left and center). The correct sequence restricts the photon's origin to a cone. For tracked Compton events, the recoil electron track must be isolated from the photon hit pattern and reconstructed separately. This defines a second cone of origin. Its intersection with the photon cone further constrains the origin of the photon. Pair events use the opening angle between electron and positron for reconstruction, providing the photon's origin and energy if the creation vertex occurs in the tracker (top right). Background events (bottom row) include charged particle tracks (bottom left) and pair events created in the calorimeter (bottom center and right): Tertiary particles and electron-positron tracks can mimic signal events with no correlation to the original source photon's energy and origin.

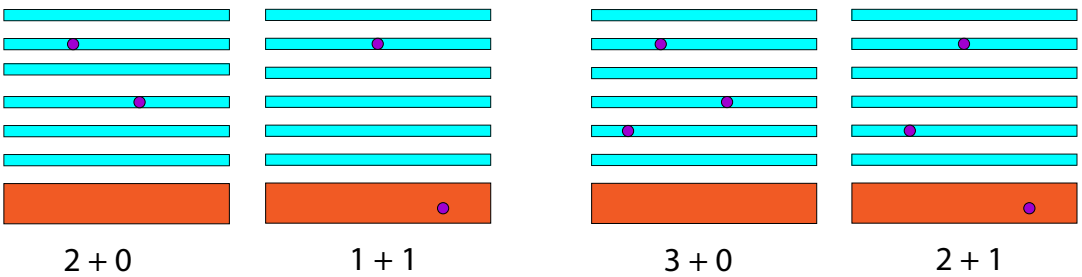


Figure 5.12: Enforced trigger pattern to ensure a minimum amount of reconstructable information in each event. Compton events have to fulfill at least 1+1 or 2+0, pair events have to fulfill at least 3+0 or 2+1.

In addition to enforcing the absence of a veto trigger, the training and test data is restricted to events that can be reconstructed in principle, as shown in Figure 5.12. In order to reconstruct a Compton sequence one needs at least two hits. Since the position and energy resolution in the calorimeter is lower in general, a coincidence of at least two individual tracker layers or one tracker layer and the calorimeter is required for Compton events included in the training set. Pair events require at least three resolvable hits.
In this study, the following definitions of event types are used:

- *Compton event:* The first interaction in the sensitive part of the detector is a Compton scatter and the veto was not triggered (Figure 5.11, top left and center).
- *Pair event:* Electron and positron interacted inside the sensitive part of D1 first and the veto was not triggered (Figure 5.11, top right)
- *Non-reconstructable:* Events that cannot be reconstructed meaningfully. Either the event is resulting from an electron primary traversing the geometry without triggering the veto, or a pair was created inside D2 and did not trigger the veto. D2-first pairs are a case of lost source photons, events triggered by electron primaries are a genuine background component (Figure 5.11, bottom row).

It was also deemed important to enforce clean samples from which the model is tasked to learn the characteristics of good Compton and Pair events. The very first interaction of a photon in each training set sample, independent of the type of event, has to occur in sensitive detector material. Valid pair events have to be created inside the tracker. Since vetoed events in space-borne detectors are usually discarded immediately, they are not considered in the datasets of this study.
The resulting stratified training set comprises of 400 000 events in total, while the test set is consisting of a total of 2.56 million events.

5.5 Performance Measures

The outcome of each binary event classification falls into one of the following four cases:

(tp) True positive: the sample belongs to the positive class and is classified correctly (positive).

(tn) True negative: the sample belongs to the negative class and is classified correctly (negative).

(fn) False negative: the sample belongs to the positive class and is classified as negative.

(fp) False positive: the sample belongs to the negative class and is classified as positive.

From this, several metrics can be derived, each bound between 0 and 1, with 1 being a perfect performance under the given metric. The most common one is *Accuracy*, defined as

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fn + fp}. \tag{5.20}$$

It gives the fraction of correct predictions relative to all predictions. Accuracy is a workable metric on class-wise balanced datasets but may give deceptive results when the data is skewed towards one class. For example, consider a case where the signal-to-background ratio is expected to be very small. A (useless) 'classifier' that would tag all events as background would reach an accuracy of almost 100% while failing the given task completely. On unbalanced datasets, performance must be measured with more robust metrics. Two of those, from the context of binary classification and information retrieval, are precision and recall. *Precision* is defined as

$$\text{Precision} = \frac{tp}{tp + fp} \tag{5.21}$$

and is a measure of how reliable a classification into a certain class is, i.e., it is a measure of the purity of the events in that class. It can be influenced by the classifier's performance on the samples of a class itself or by misclassification of samples belonging to other classes.
The other metric is the *recall* on a given class. It is defined as:

$$\text{Recall} = \frac{tp}{tp + fn}. \tag{5.22}$$

Recall, also called true positive rate, is a measure of how well a given class is recognized by a classifier, i.e., the efficiency with which these events are properly classified. In general, it is impossible to optimize both precision and recall to arbitrarily high values unless the samples in the dataset are perfectly separable. Usually, precision and recall reach an equilibrium state where improvements in one metric imply performance loss in the other.
Both metrics can be extended from binary to multi-class scenarios by applying a one-against-all scheme. The class to be evaluated is considered the positive class, while the other classes are combined into the negative class. False positives then have to be defined as the combined misclassification into the investigated class from all other classes. Table 5.2 shows the confusion matrix for the 3-class task to be performed.

5.6 Classical Machine Learning Approach

Before applying deep learning approaches to a given task, it is useful to try classical machine learning approaches or conventional algorithms on a given problem. This helps to establish a baseline and allows for direct feedback if a chosen approach results in a better performance. The chosen algorithms are the classical decision tree and boosted decision trees. This family of algorithms is able to capture non-linear dependencies and is therefore suited for the present classification problem.

Table 5.2: Layout of the confusion matrix for the 3-class task in this study. True positives align on the main diagonal. Each column corresponds to all samples predicted into a given class, each row provides the classification result for a given (true) class. C is for Compton event, P is for pair event, NR is for non-reconstructable event. Every off-diagonal matrix element can be interpreted as either a false negative or a false positive of a given class.

Truth\Classification	Compton	Pair	non-reconstructable
Compton	tp C	fn C, fp P	fn C, fp NR
Pair	fn P, fp C	tp P	fn P, fp NR
non-reconstructable	fn NR, fp C	fn NR, fp P	tp NR

5.6.1 Decision-Tree Algorithms and Random Forests

Decision trees are among the most widely used learning algorithms. They sort a data set using attributes associated with the respective sample from a root node into terminal leaf-nodes. Nodes that are connected to other nodes further down the tree are called internal nodes. Figure 5.13 shows an illustrative sketch.

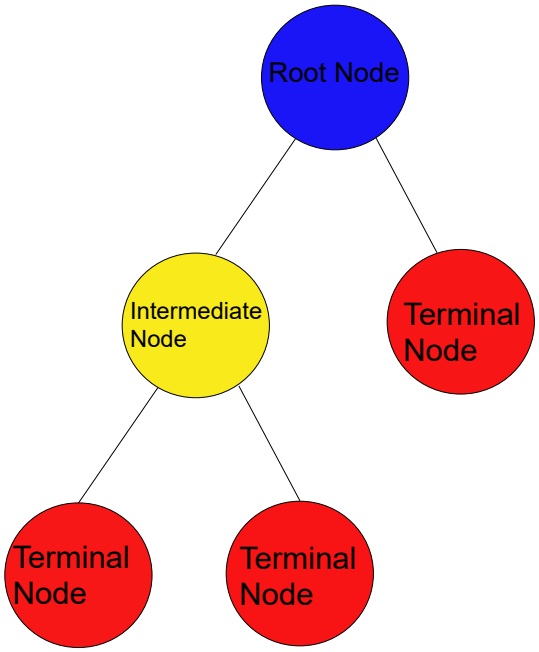


Figure 5.13: Sketch of a decision tree. The root node (blue) is the beginning of the tree. Daughter nodes can either be internal nodes (yellow) or terminal nodes (red). A sample is tagged to belong to a certain class either by belonging to the single surviving class of the sorted samples (pure node) or the majority of the samples in the respective terminal node.

At each non-terminal leaf the algorithm tests the remaining samples against different thresholds on the available data set features. Depending on the used training settings, the splits are either continued until all nodes are pure or a stopping criterion like the number of levels in the tree is reached. If the terminal leaf is not pure, the majority of the samples determines the classification of this leaf. For this trial the Scikit-Learn [Ped+11] implementation was used. It uses an adaptation of the Classification And Regression Tree (CART) algorithm to grow decisions trees which is going to be described in the following.
The algorithm was first described by Breiman et al [Bre+84]. The distinct difference from other training algorithms is the binary selection enforced at each split. An alternative approach is the C4.5 algorithm [Sal94] that allows for more than two possible separations at each node.[2]
The criterion on which attribute to split at a given node is determined by the impurity after the split or the information gain. In the classical CART algorithm this is measured by the Ginni index. The Scikit-learn implementation also supports Cross-Entropy with information gain. If one denotes the fraction of samples belonging to class i at a given node t as $p(i|t)$ and a number of classes c one can define impurity measures as follows [Tan+19].
The Gini index [Gin55] is defined as

$$\text{Gini(t)} = 1 - \sum_{i=0}^{c-1} (p(i|t))^2 \tag{5.23}$$

Smaller values are indicating a higher purity after the split. In the case of an even distribution between classes, the range is between 0 (pure) and $1 - \frac{1}{c}$ (randomly distributed). It gives the probability of classifying a randomly drawn sample out of the data partition at the node into the wrong class.
A different measure is the information-entropy, also called Shannon index [Sha48] and Information-gain. The information-entropy on a data partition in this context can be defined as

$$E = -\sum_{i=0}^{c-1} (p(i|t) \cdot log_2(p(i|t))) \tag{5.24}$$

It gives a measure of disorder of the data partition at the given node after the split, where lower numbers correspond to more order in the data partition. In order to assess the quality of the split when using information-entropy, the information gain given as

$$IG(t, X) = E(t) - E(t|X) \tag{5.25}$$

This gives the change in entropy after the split X with respect to the entropy of the data partition Y before .

[2] Note that every non-binary split can be rewritten into successive binary splits. For unlimited depth, both algorithms arrive at equivalent trees.

The best performing split is then applied in the given node. This manner of determining the best attribute and threshold to split on is also called "greedy".
This can be summarized as follows:

1. Test at each node if a stopping criterion has been reached. If so, make the node terminal, if not proceed.
2. Test the data partition at the node against all attribute splits.
3. Select the split with the highest quality factor
4. For each daughter-node, test if the remaining partition is pure (make this node terminal) or not (create a new node and proceed at 1.)

The applied splits can be represented as a cascade of if-then-statements, that turn the decision tree into a human-readable form, allowing to directly follow the reasoning of a given prediction. This is in contrast to Deep Learning models, where tracing the reasoning is often only possible by employing indirect methods like Class Activation Maps (CAM, [Zho+15]), Gradient Class Activation Maps (GradCAM [Sel+16]) or saliceny maps [SVZ14]).
A single decision tree is prone to overfitting the presented data. A way to remify this is to extend the single decision tree into a committee of multiple decision trees. This set of multiple trees is called a random forrest and was first descibed by Breimann in 2001 [Bre01]. The idea is to grow a set of trees with boot-straped training sets, i.e. one draws a new training set from the data set present with put-back. This introduces an individual bias into every tree, which averages out if the number of trees is sufficiently high. The main hyperparameters for a random forrest are the maximum depth of the individual trees and the number of trees that comprise the committee.

5.6.2 Training Procedure and Hyper-parameter selection

The dataset is generated from a set of binned simulations and is stratified in directional bins, energy and topology. It contains a set of features that can be extracted from the raw data that would be available from the detector without knowledge of the Monte-Carlo truth of the given event. The features can broadly divided into energy driven and topological aspects of an event:

- deposited energy of the whole event
- deposited energy in the tracker (D1)
- deposited energy in the calorimeter (D2)
- trigger pattern (see Fig. 5.12)
- number of triggered layers in D1
- average energy deposited in all tracker layers containing at least one hit

Model optimization was done along tree depth and number of trees as hyper-parameters using the Gini index. The model is tasked to classify between Compton events, D1

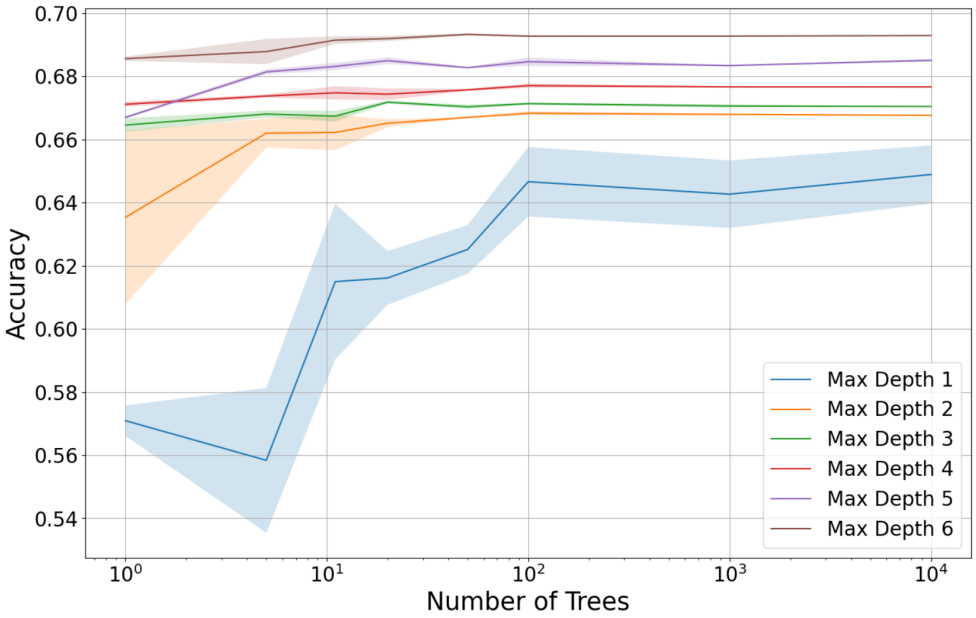


Figure 5.14: Classification accuracy as function of number of trees for random forests with maximum depth from 1 to 6. Given is the averaged accuracy obtained in three separate runs with redrawn training sets. The band represents the standard error of the mean.

first Pairs, and non-reconstructible events consisting of D2-first Pairs and Lepton caused events. The full training set contains a total of 300k events with an evaluation set of 100 k unseen events.
The best performing model is then tested against a data set with 2.56 million samples. In order to find a suitable set of parameters a grid search is performed and averaged over three reruns.. Since the training set is boot-straped for each of the runs, one obtains a spread, allowing to estimate the variance of the resulting classifier. Figure 5.14 shows the result of three training runs. As expected, using a single node leads to the most unstable and worst results. However it illustrates the working principle of random forests. A single node can only classify into two classes as it is only able to perform a single split. It can only cover two of the three classes in the data set. With increasing number of estimators, the whole committee obtains coverage of all three classes, resulting in a better performance on the whole task. The performance obtained from this configuration showed the largest spread in accuracy, since the performance of each individual node depends now heavily on the composition of the boot-straped training set presented For larger maximum depth, the resulting models become reliably stable for forests of around 30 trees. The best performing parameters were found for a maximum depth of 6 and 50 individual trees.

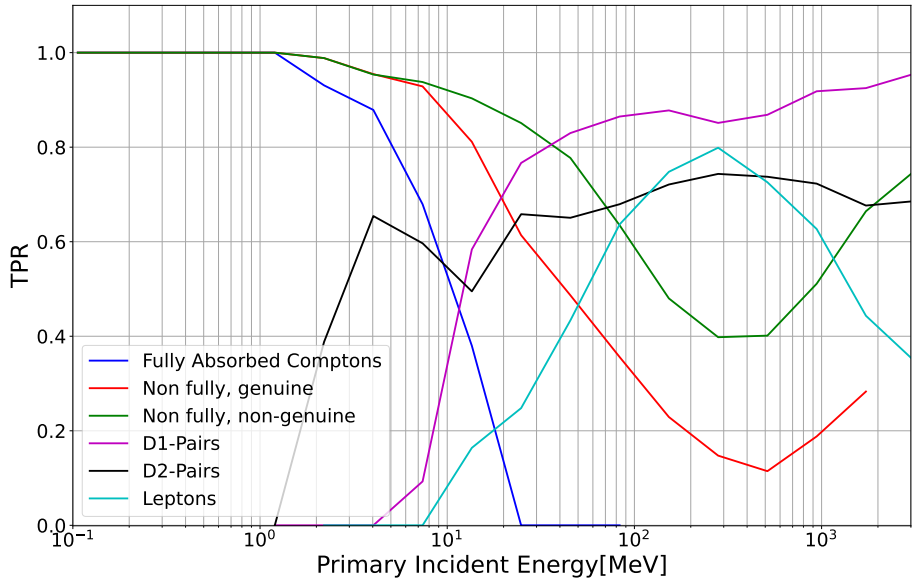


Figure 5.15: Directional averaged recall (TPR) for Compton, Pair and non-reconstructable events.

5.6.3 Results

Figure 5.15 shows the class coverage of the selected tree classifier as function of energy. Compton events are split into three sub topologies. The most important sub topology are the fully absorbed photoins. Incomplete absorption can be split into out-scattering of the incident photon after starting the event with an interaction in the sensitive part of the detector and incomplete absorption due to prior interaction outside of the sensitive volumes. The model performs a sharp cut between Compton and D2-pair events between 1 and 1.5 MeV. The performance on fully absorbed Compton events then drops shaprly to roughly 50% at 10 MeV. All events below are tagged as Compton events. It starts to become sensitive for D1-pair events only after roughly 7 MeV while reaching a performance of 40% at 10 MeV. At around 500 MeV the performance reaches 90% or higher. D2-first pair events are detected at around 60 to 70%, while leptons peak out at around 80% 300 MeV incident energy.
From this one can conclude that a random forest can yield to a workable classifier for the given problem. It also shows, that the problem is ”learnable” and that further improvements can be expected from changing the approach and data representation.

5.7 Going Deep with Convolutional Neural Nets

The data generated by the detector is intrinsically binned in space and energy. Finding a suitable vectorized data representation is non-trivial. Convolutional neural nets (CNNs) are specialized architectures for this type of data as they make use of local dependencies of adjacent bins. They are therefore a natural choice for an alternative deep learning approach.
Convolutions can be applied either on 2D images or 3D volumes, which could be stacks of images forming a video or a binned 3D representation of a detector. A full 3D representation of the tracker module of e-ASTROGAM using only one 8-bit color channel for the deposited energy results in 3D images with a size of around one GB. This would lead to a very large training dataset and intense VRAM usage during training and inference. Therefore a multi-view approach was chosen. For optimal performance, each view is processed in a separate stack of convolutional modules. The result of each stack is then merged and further processed. The design of the tracker makes the use of X-Z, Y-Z and X-Y-projections of an event a natural choice. In addition to the tracker information (one row of pixels per layer), discretized calorimeter information (one row of pixels per centimeter of crystal height) is included at the bottom of the tracker projection. In order to avoid large memory and processing overhead, the X-Y projection was down-sampled by a factor 10.
This reduces the memory footprint of the input data from around 1 GB down to 0.71 MB per event using the same precision for the deposited energy and a major factor for the chosen approach.

5.7.1 Model Architecture and Training

A preliminary investigation using the exact approach presented by [Aur+16] led to the conclusion that the use of Inception modules (see [Sze+14]) would lead to a rather slow and computationally expensive model. The SqueezeNet architecture [Ian+16], however, enables the design of small, yet powerful models. It has been demonstrated by [Gsc16] to run on small-scale hardware like FPGA-accelerated System-On-Chips (SoCs). Their design pattern was chosen for the convolutional modules. The presented model makes heavy use of SqueezeNet's Fire modules shown in the top right of Fig. 5.16. The final model is sketched in Fig. 5.16 left.
A common restriction in Feedforward Neural Nets in general lies in the fixed input size. In the case of CNNs, this restriction is self-imposed by only using a dense layer to perform the classification. If used conventionally this layer takes the stack of final feature maps, flattens them, and uses the resulting 1D tensor with a length equal to the sum of the product of the width and length of each feature map as input. The number of free parameters introduced by the dense layer for processing is

$$\text{parameters} = \text{input length} \times \text{output nodes} \times \text{maps}$$

and can quickly reach millions because of the size and number of the feature maps to flatten. For example, an incoming stack of 32 $100 \cdot 100$ pixels feature maps as in-

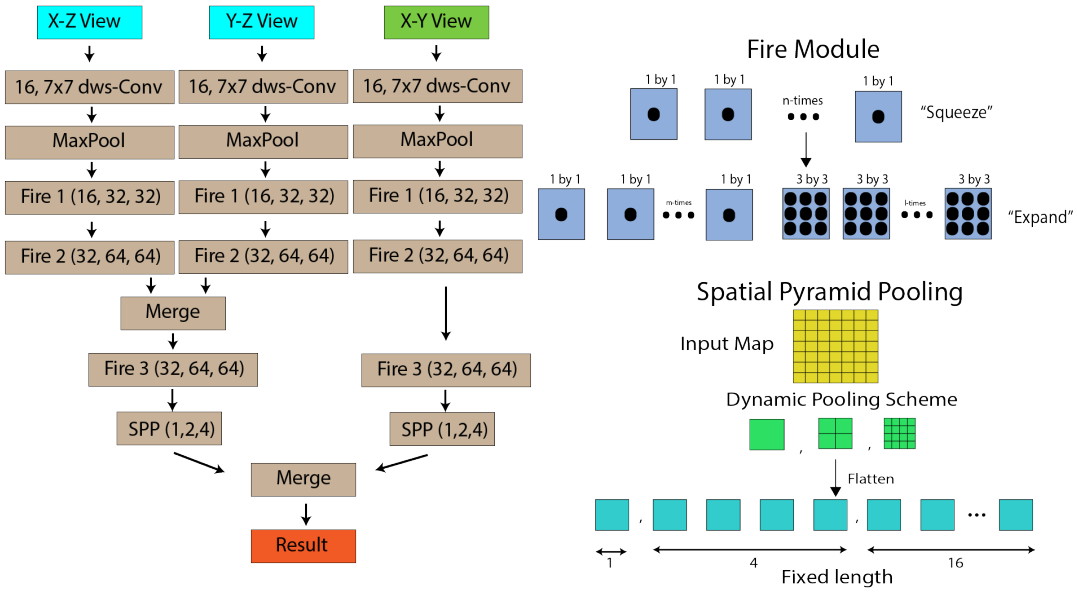


Figure 5.16: Left: Sketch of the CNNCat model architecture. Each view is processed in a separate stack of convolutional (dws-Conv: depth-wise separable convolution) and pooling layers with identical layout. The results are merged and processed again. The final output is a classification vector whose components give the probability for the shown sample to be either a Compton, Pair or non-reconstructable event. Top Right: Sketch of the micro architecture of a SqueezeNet Fire Module micro-architecture. The number of input maps is reduced to n by the 1×1 "squeeze" filters that act as information bottle necks. The resulting data is processed by m 1×1 and 3×3 "expand" filters. The stacked output is handed to the next layer. Adapted from: [Ian+16]. Bottom right: Sketch of spatial pyramid pooling. The incoming feature maps are binned via pooling into 1×1, 2×2, 4×4, ... representations, which are then flattened into the input of the dense layer. Here the length of the outgoing vector is fixed to $n \cdot (1 + 4 + 16)$, where n is the number of feature maps to process. Adapted from: [He+14]

put for a classification into seven classes results in a dense layer with 2.2 million free parameters and the need for a potentially very large training set. Due to their nature of connecting every input node with every output node, dense networks correlate each pixel in a feature map with every other pixel, which is also a source of faulty correlations either by pure chance or simulation biases. A more robust approach called Spatial Pyramid Pooling was introduced by [He+14]. The final feature maps are binned by applying Pooling with a set of dynamically sized windows. When flattened, this results in a fixed input size for the dense layer, regardless of the original image size. Each value now corresponds to regions of the processed feature map, effectively binning the image into $n \cdot n$ equally sized rectangles. This facilitates de-

coupling of the input size of the classifier (the dense layer) from the output of the feature generator (CNN). In addition, the size of the input vector of the dense layer is generally smaller than with conventional flattening. This reduces the number of free parameters. He et al. [He+14] also demonstrated that contemporary image classification networks using this approach outperform the designs with conventional flattening. In this study, max-pooling was used to generate the binned representation. Another application of this pooling scheme is to resize intermediate input with respect to each other. The feature maps of the X-Y filter tree have a different size with respect to the X-Z and Y-Z maps. This can be ramified by applying the dynamical pooling scheme in order to generate compatible input for the dense layer. One go-to method for regularizing neural networks (i.e., to prevent overfitting) is Dropout [Sri+14]. During training in layers with Dropout applied (usually the dense layer or the first few convolutional layers), randomly selected neurons are switched off for each training step. To compensate, the model has to learn to propagate vital information for the classification over various paths in the graph. During inference, no Dropout is applied, i.e., the full model is used. Using weight adaptation only during training results in a more robust model. The CNNCat model was implemented using TensorFlow [Aba+15] Version 2 with TensorFlow's internal keras implementation. Spatial Pyramid Pooling was implemented with a modified version of [Hen17].

5.7.2 Dataset Generation and Selection Criteria

The simulation output was processed into X-Y, X-Z and Y-Z images using custom-written python scripts. Each Z-projection has a size of 66 rows and 4267 columns with 3 color channels. The 66 rows correspond to 56 DSSD layers atop of 10 calorimeter slices. The width of 4267 pixels include the 3840 DSSD channels extended by 427 pixels to accommodate the 10 cm wider calorimeter. The X-Y projection has a size of 427 by 427 and also 3 channels. Channel 1 contains the binned deposited energy in the pixel saturating at 2048 keV, channel 2 contains the deposited energy relative to the total energy deposited in the tracker, while channel 3 contains the deposited energy relative to the total energy measured in the calorimeter. If there is no deposited energy in the calorimeter for the given event, the values of channel 3 is set to 0. For image augmentation, Gaussian noise with the estimated energy resolution of the proposed detector was applied to the input images to further solidify the model against overfitting.[3] For pair events and charged particle events, only incident energies above 10 MeV were considered. For D1-pairs, the track topology at low energies shows strong resemblance to Compton recoil electron tracks as the positron generates only a short track or no track at all. Pair creation at these energies is also suppressed compared to Compton scattering. The track of electrons at this energies shows the same morphology of recoil electrons due to Molíere scattering. At energies below 10 MeV, charged primaries also don't reach the detector, from the unshielded bottom side.

[3] Since the D1 detector is highly inhomogeneous due to its layered design, the common technique of rotating the image cannot be applied.

5.8 Results

The presentation of the results begins with an overview of the computational demands followed by a performance analysis on two test sets. Test set 1 comprises the full observational range from 100 keV to 3.5 GeV from all incidence angles. Test set 2 considers a simulated Crab observation.

5.8.1 Computational Demands and Architectural Optimizations

Since one of the possible applications should be an online data classifier used on the satellite itself, all the design choices were driven toward optimizing the number of computations per event. The choice of Fire-Modules as micro-architecture for the convolutional modules allows to control the size and hence the computational demands of the model, due to their built-in bottlenecks. Dropout as a regularization method introduces no computational overhead at run-time while effectively constraining the model and preventing overfitting of the training data.
To further optimize the number of computations per convolutional layer, depth-wise separable convolution (DWS-Conv) [Sif14] was applied wherever feasible, i.e., on the 7×7 and 3×3 convolutions[4].
A conventional convolutional layer applies m $n \times n$ convolutions on l incoming channels on an image with $j \times k$ pixels, where m denotes the number of filters to be applied and n is the filter size along one axis in pixels. Assuming $m = 16$, $n = 3$, $l = 3$ and $j = k = 100$, this setup would result in

$$\text{parameters} = n^2 \times l \times m = 432 \tag{5.26}$$

and

$$\text{ops} = j \times k \times n^2 \times l \times m = 7 \times 10^6 \tag{5.27}$$

operations used in this layer.
Depth-wise separable convolution, on the other hand, tries to simplify this step without sacrificing too much performance. In this scheme, only one full $n \times n$ convolution is applied to each incoming channel. The intermediate data is put back together and then stretched to the required dimension by applying m 1×1 convolutions. In the setup described above, this results in

$$\text{parameters} = n^2 \times l + l \times m = 75 \tag{5.28}$$

and

$$\text{ops} = n^2 \times l \times j \times k + l \times j \times k \times l \times m = 1.2 \times 10^6. \tag{5.29}$$

Applying this to the precursor model allowed to reduce the number of trainable parameters to 64 k and the number of FLOPS per event to 0.9 GFLOPS. ESA identified

[4]This speed-up may not be realized with non-optimized TensorFlow models since the underlying libraries (e.g. CuDNN) may not be optimized for this convolution type. Running the model through an inference engine like tensorRT offered by NVIDIA will show the expected behaviour.

Table 5.3: Best training parameters found after optimization. Delta gives the minimum improvement from the previous best loss-value, patience is the number of epochs in which the improvement has to be achieved.

Learn Rate I	Learn Rate II	Dropout	Batch Size	SPP	Delta	Patience
10^{-4}	10^{-6}	0.3	1	[1,2,4]	0.001	5

the necessity to explore systems for onboard processing. They funded the GPU4S (GPUs for Space) project that surveyed a variety of use-cases and possible embedded systems for implementation [Kos+20]. Follow-up studies also investigated potential improvements of commercial off-the-shelf embedded systems (e.g. [Ort+23]). Ortiz et al. also studied the Jetson AGX system which delivers around 1.3 TFLOPS per Watt at FP16 or 2.6 TOPS at INT8, equivalent to a rate of 1.4 kHz or 2.9 kHz per Watt, respectively. Note that either a smaller detector or a coarser binning of x-y strips in the input data would increase the amount of rate per Watt significantly as the computational complexity for CNNs scales mostly with $j \times k$ of the input data. According to NVIDIA's claims, space-graded systems are to be released. Therefore, a model similar to the one presented here is suited to run on soon to be available embedded systems.

5.8.2 Training and Validation

The training set consisted of 400 000 events evenly divided over classes, incident energy and incident angle binning. Of these events, 300k stratified events were used for training, 100k similar selected events were held back for validation for the purpose of monitoring the training progress. The Adam optimizer [KB17] was used for parameter adaptation. Early stopping was used to prevent overfitting due to excessive training. The progress during training is shown in Figure 5.17. The relevant training and hyper parameters are summarized in Table 5.3
Training for the best-performing model took 39 epochs in total with a total duration of around 5 days on an NVIDIA GeForce RTX 2080 Ti graphics card. The training was performed in two stages. The first stage at Lear Rate I was terminated by early stopping at epoch 10. It was then restarted at Learn Rate II with the weights of epoch 5. After the second stage, he model was reset after early stopping to the best-performing weights after epoch 34, which were then used for evaluation against the test sets.

5.8.3 Classification Performance

The test set contains 2,562,406 events containing photons and electrons ranging from 100 keV to 3.5 GeV simulated from a flat spectrum uniformly from all incident angles. The energy range was divided analogously to the training-set generation. This dataset is reserved to analyse the energy dependency of the classification result averaged over all incident angle bins and was not used during training or validation.

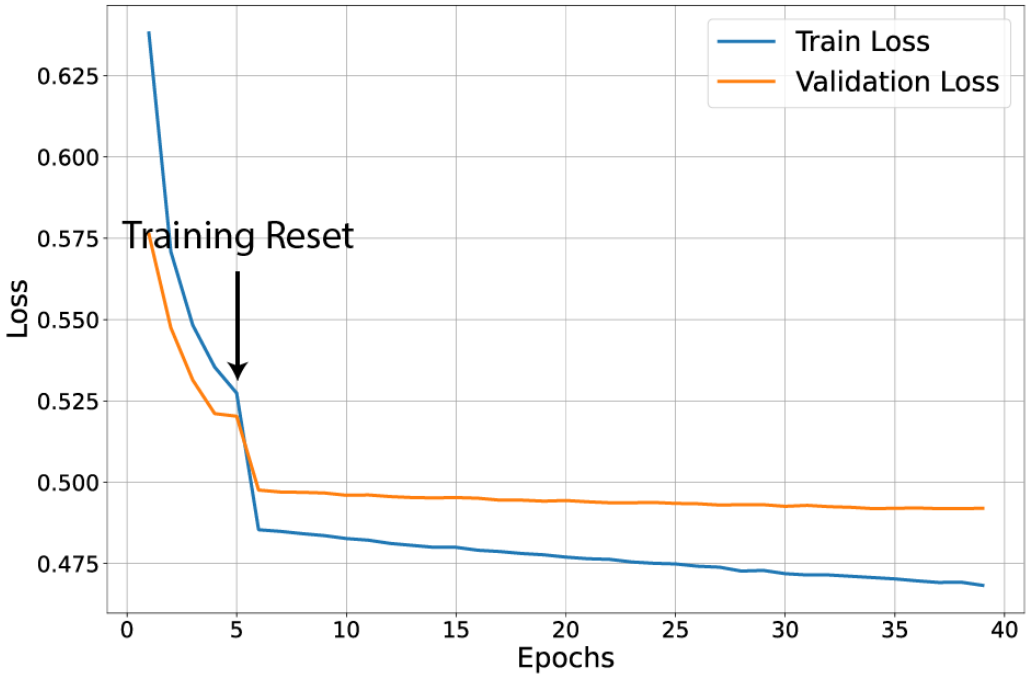


Figure 5.17: Training Loss as function of epochs. Training was started with Learn Rate I. After epoch 7 early stopping terminated training run 1. The model was reset to the previous best weights. The training was then restarted with Learn Rate II for the remaining 34 epochs.

The performance of CNNCat is evaluated over the full observation range of the instrument using the average over all directions as a function of incident energy of recall, precision, and class-wise false negative rates. The resulting plots are energy-dependent representations of rows of the confusion matrix, shown in Figures 5.18a – 5.18c. On the input side, three types of Compton events are further distinguished: fully absorbed (golden) events, non-fully absorbed events starting in sensitive material ("genuine" Compton events - potentially reconstructable), and non-fully absorbed events starting in passive material ("non-genuine" Compton events), which, effectively, are source photons turned into background, as the initial scatter angle and energy loss is unknown.
Figure 5.18a shows the model performance on Compton events. The recall of all Compton subsets is very high (above 90 %) for incident energies up to 4 MeV. The recall of non-fully absorbed genuine Compton events steadily declines over the remaining energy range, while it stabilizes at 60 % at incident energies between 7 to 20 MeV for fully absorbed Compton events . Non-fully non genuine event are less accurately tagged with increasing incident energy but stay above 70 % over the shown energy range. Fully absorbed Compton events which are not correctly tagged are mostly classified as Pair event at energies around 3 to 12 MeV. Beyond that, the ma-

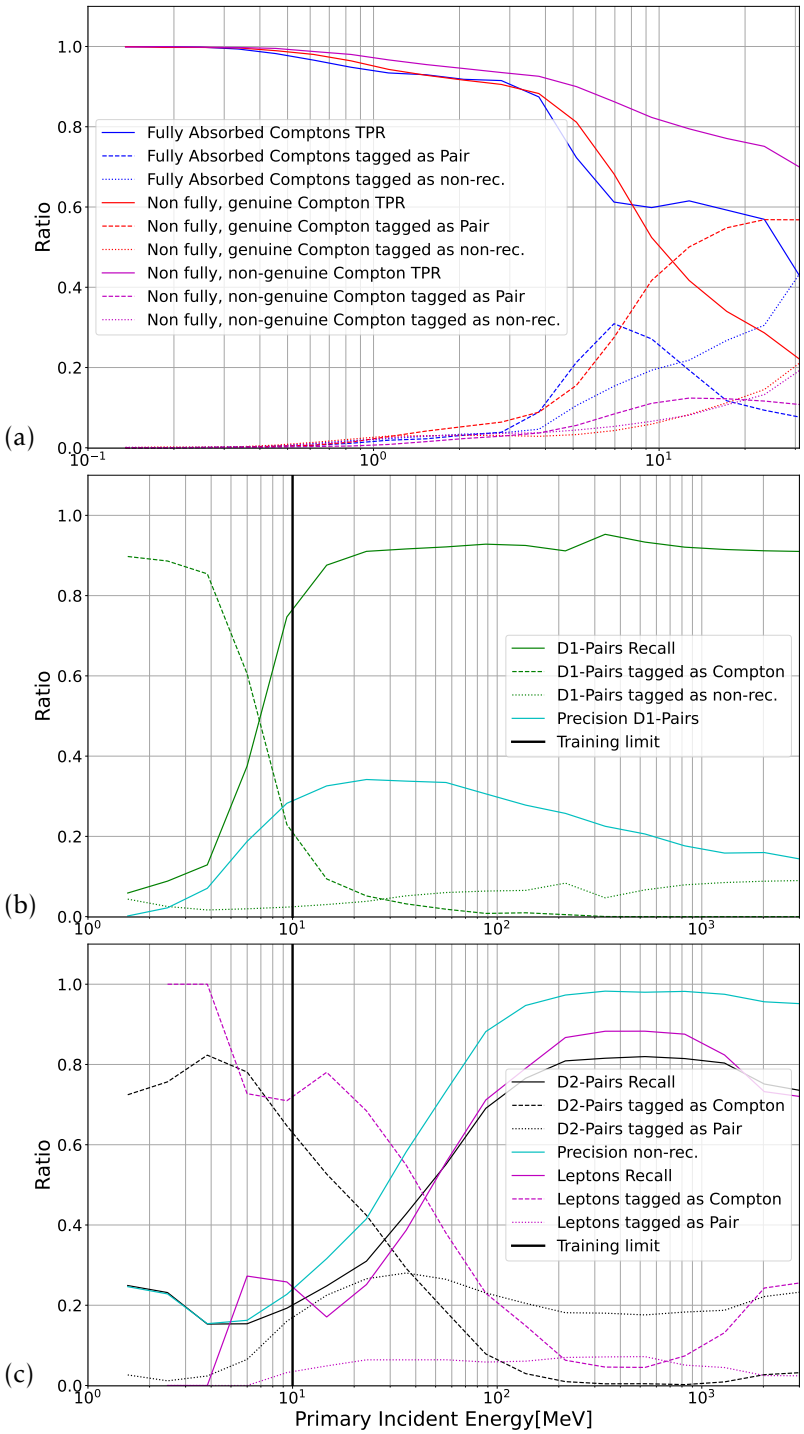


Figure 5.18: Precision, recall and class-wise false negative rates vs. incident energy.
(a) Compton events. Recall for the three sub-categories of Compton events and corresponding false negative tags vs. incident energy up to 30 MeV.
(b) D1-first pair events. The model was trained on events with E_{inc} >10 MeV, denoted by the bar.
(c) Non-reconstructable events. Misclassified events are mostly tagged as Compton for E < 40 MeV and as Pair above. The model was trained on events with E_{inc} >10 MeV, denoted by the bar.

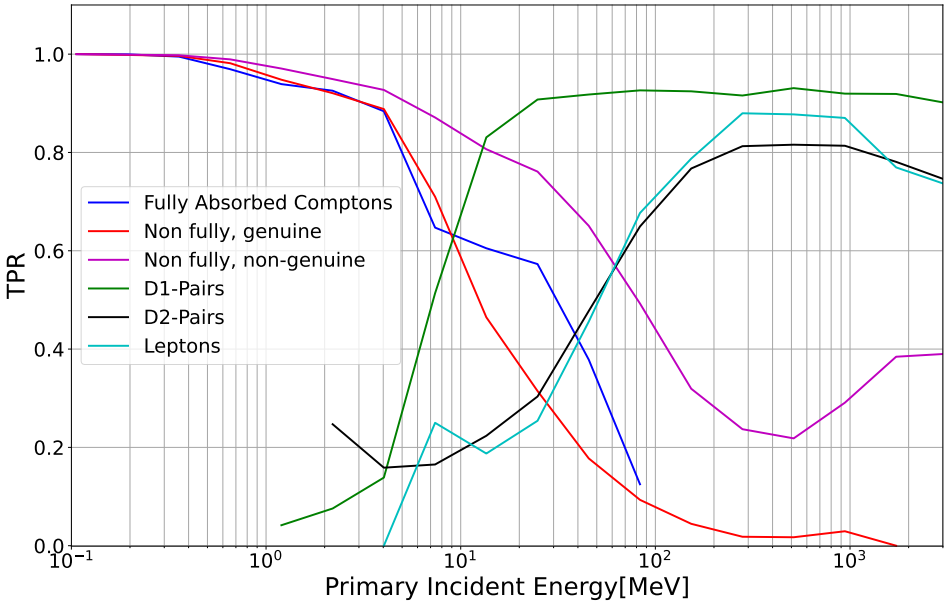


Figure 5.19: Recall vs. incident energy for the three Compton event populations, D1- and D2-first pairs, and lepton events.

jority of false negatives is tagged as non-reconstructable. The vast majority of false negative non-fully absorbed genuine Compton events are wrongly tagged as Pair. This can be explained by the very narrow scatter angle distribution at higher energies, which will give photon hits a topology similar to tracks which together with a recoil electron track can confuse the model. Another possibility is the generation of multiple recoil electron tracks before the photon leaving the detector. Non-genuine non-fully absorbed Compton events show a behaviour similar to fully absorbed Compton events, but shifted to higher energies. Here the turnover from Pair to non-reconstructible tagging occurs around 18 MeV.

Figure 5.18b shows the model performance on pair events with their first interaction measured in D1. The recall on pair events at incident energies below 7 MeV is very low. This is not surprising since the model only had pair events with incident energies larger than 10 MeV in the training data. Electrons and positrons in this energy range are very susceptible to multiple scattering in solid material and show a topology similar to recoil electrons in Compton events. Also, positrons are generating shorter tracks, leading to even greater similarities between Compton and pair events. At 10 MeV, however, the recall is already passing 78%. For energies higher than 100 MeV the recall saturates at 92%. Precision is in general low due to the large amount of false positives from Compton events at low energies and D2-first pair events over the whole energy range.

Figure 5.18c shows the model performance on non-reconstructible events. The recall

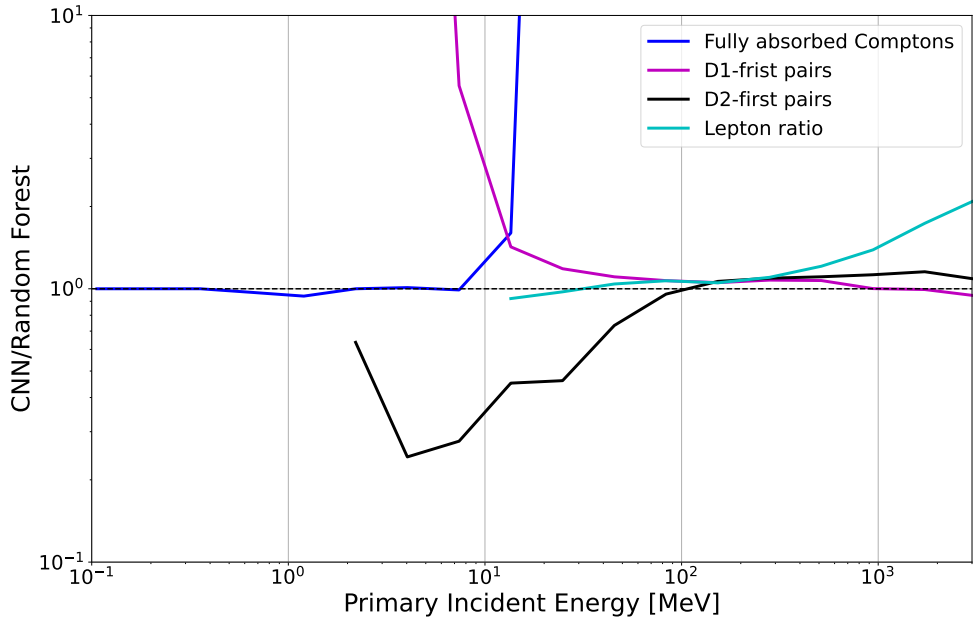


Figure 5.20: Recall ratio for the trained model with respect to the random forest as function of energy. Values below 1 indicate better class coverage of the random forest, and vice versa. Discontinuities in the performance ratio indicate blind spots of the random forest classifier.

on D2-first pair events is rising to around 80% at around 100 MeV leptonic events are recognized reliably from around 70 MeV and higher with a peak recall of 92% around 200 MeV. In the energy range up to 50 MeV, the majority of misclassified events is tagged as Compton. Over the remaining range, most of the misclassified events are tagged as pair. Overall precision for energies higher than 100 MeV peaks at 92%, leaving a relatively small number of potentially reconstructable events that were falsely rejected.
Figure 5.19 summarizes the recall vs. incident energy for all classes, distinguishing the three Compton event populations, D1- and D2-first pairs, and lepton events.
Comparing the obtained class coverage with the random forest shown in Figure 5.15 one can see (as shown in Figure 5.20) that the performance for fully absorbed Comptons is equal or a bit worse for energies up to 2 MeV. Beyond that the convolutioal neural networks performs better. The steep rise of the recall ratio indicates a “blind spot” of the random forest, i.e. the random forest does not classify any fully absorbed Compton events correctly with energy higher than 26 MeV. This can be explained by the “greedy” nature of the random forest algorithm which should help it to model hard thresholds like the pair-creation threshold, but makes it difficult to handle overlapping classes. D1-pair events are captured better by the CNN for energies up to 2 GeV. The largest improvement ratios can be found in the low energies. The random forest again displays a “blind spot” for pair events below 4 MeV. Only at the highest

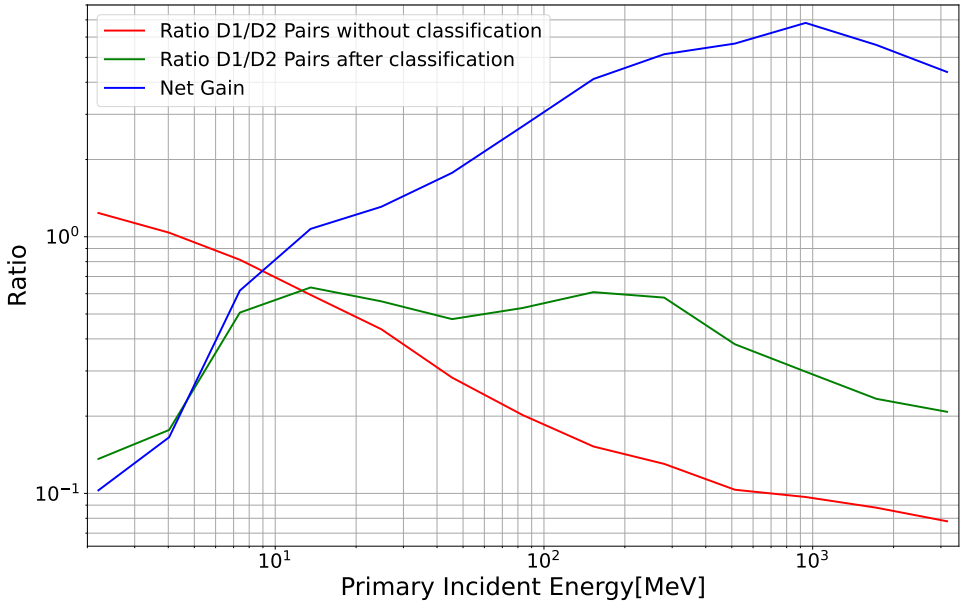


Figure 5.21: Ratio of D1-first to D2-first pair events vs. incident photon energy. At lower energies (below 15 MeV) the effect of the classification on the ratio is negative. This is because of the lower recall of the model on pair events in this energy region. For higher energies, the model is able to improve the ratio of D1-first vs. D2-first pairs by a factor ranging up to 8 at 1000 MeV.

energies the random forest performs slightly better. The coverage of leptonic events is in general better with the CNN, with the exception of energies below 30 MeV. Recall on the D2-first pairs is mixed compared to the random forest. While the overall coverage of the random forest is better at energies up to 70 MeV, the CNN has a better coverage at higher energies. Overall the CNN outperforms the random forest with at some energies very significant margins. It is an improvement over the random forest classifier.

Fig. 5.21 shows the ratio of D1-first ('good') to D2-first ('bad') pairs over the observed energy range. Since e-ASTROGAM was designed to measure both Compton and pair events, the detector design had to compromise between the demands of both detection channels. For instance, the tracker layers lack tungsten converter foil, which enhances the creation of pairs in Fermi-LAT, since this would severely hamper the measurement of (low-energy) Compton events. As a result, many high-energy source photons traverse the tracker without interaction and create pairs only in the calorimeter as indicated by the ratio before classification. This skews the ratio of D1-first to D2-pairs heavily towards non-reconstructible events that are potentially dangerous. While initially above 1, the ratio soon drops to values below 0.1, manifesting the large surplus of D2-first pairs. This ratio greatly improves after classification for

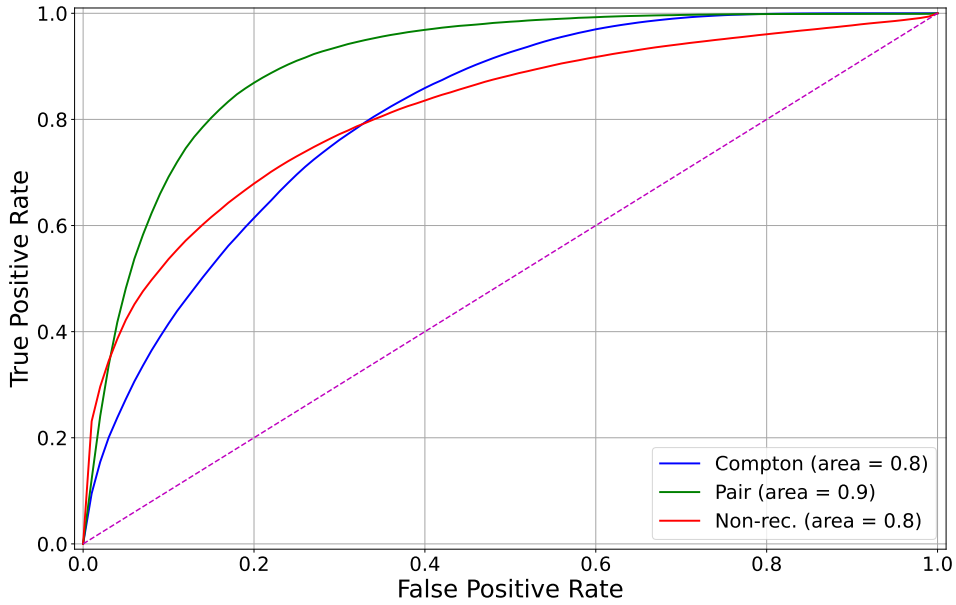


Figure 5.22: Directionally averaged ROC curves for each class for the test set. The performance of a random classifier corresponds to the dashed line. The Area Under the Curves are given in the legend. Classification of pair events performs best, with an AUC-score of 0.9. The other classes are still performing very well as indicated by scores larger than 0.8.

incident energies greater than 15 MeV. Above 20 MeV, i.e., in the intended observation range in pair-imaging mode, the model greatly improves the 'good'/'bad' pair ratio by up to a factor 8, shown by the blue line. The weak performance at energies below 10 MeV is tolerable, since, at these low energies, pair events are difficult to reconstruct and are dominated by Compton events.
Overall, one can conclude from this test set that a small-scale model is capable of distinguishing the presented events by type. As expected, the decision to classify an event as either Compton or pair is biased to some degree by the energy the model can see in an event. This can be seen from the fact that almost no events with an incident energy below 1 MeV are classified as pair event. The classification has a beneficial effect on the ratio of D1-first pairs to D2-first pairs for all but the lowest energies.
To assess the total performance of our classifier, the Receiver-Operator-Characteristic Area-under-the Curve (ROC-AUC) is calculated by determining the True Positive Rate and False Positive Rate for decreasing decision thresholds for the positive class. In multi-class scenarios this is done in a so called 'one vs. all' approach. There is only one positive class and the other classes are merged into a single negative class. This score can be used to assess the performance of a classifier. The resulting curves for this test dataset are shown in Figure 5.22. All three classes are well separated

from random guessing (indicated by the dashed line). There is no task-independent interpretation of the score, but one can employ a rule of thumb found in [HL13]. Performance could be considered 'poor' for a score below 0.7, 'acceptable' for 0.7 to 0.8, 'excellent' for 0.8 to 0.9, and 'outstanding' for scores larger than 0.9. The scores for each class are around a value of 0.8, which indicates a good performance on the presented dataset. The best performance is achieved by the pair classification, which achieves an area of 0.9. According to the cited performance standards, our model reaches the second highest performance category.

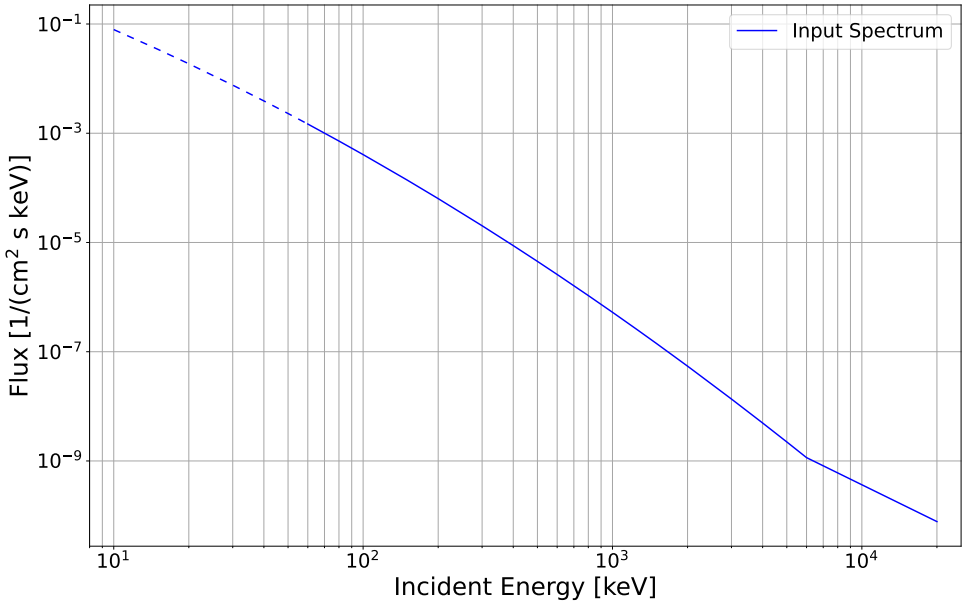


Figure 5.23: Input spectra for the Crab-Nebula dataset. The Crab spectrum uses [JR09] at $E < 6$ MeV and a power-law with index 2.24 above. The simulation range is indicated by the solid line.

Table 5.4: Upper limit of the simulation bins, medians and contributing events per bin. The minimum photon energy was set to 60 keV. The total size of the dataset is 2.85 million events.

E_{max}[MeV]	0.13	0.26	0.45	1.1	2.4	4.7	10	20
Median [MeV]	0.07	0.16	0.31	0.56	1.35	2.85	5.72	12.03
Events	106k	289k	378k	415k	422k	422k	413k	407k

5.8.4 Crab Nebula Simulation

This part investigates the impact of prior classification on the reconstruction of Compton event data. To do so, the standard MEGAlib reconstruction chain is applied on a simulated Crab Nebula measurement. The Crab spectrum was modelled according to ref. [JR09] for energies up to 6 MeV, and extended to 20 MeV by a power-law with index 2.24, as shown in Fig. 5.23. The Crab simulation was performed in eight separate energy ranges. For each bin, a simulation with 500 k pre-triggered events was performed, resulting in a total of 2.85 M events with a distribution shown in Table 5.4. The source was set up as a far-field point source with an incident angle $\theta = 30°$. MEGAlib was tasked with reconstructing Compton and pair events and eliminating events from minimum ionizing particle tracks (MIPs), using the standard reconstruction chain.

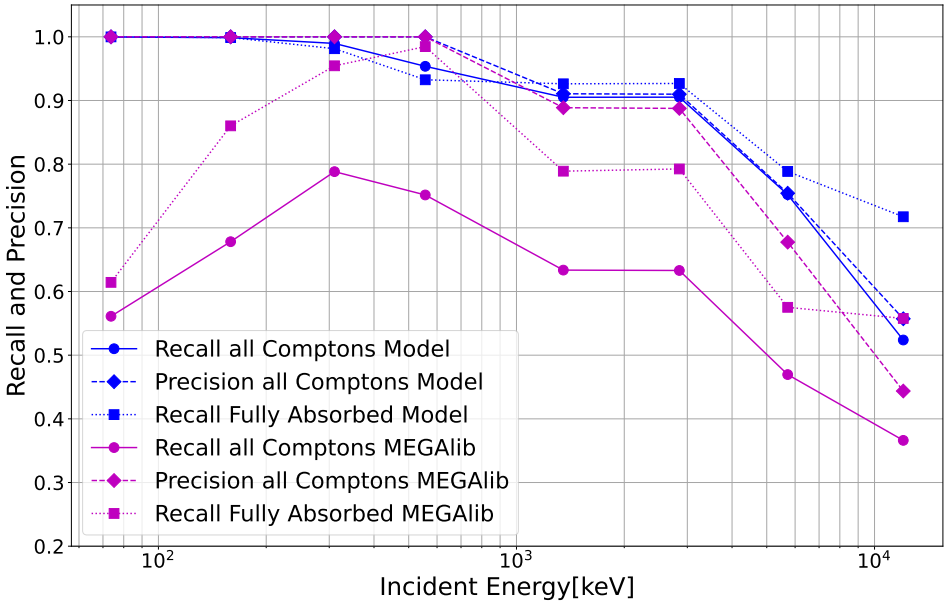


Figure 5.24: Precision and Recall for Compton events as function of energy for the Crab-source dataset.

An event was tagged as Compton when the reconstruction successfully attempted a Compton reconstruction. The event was considered tagged as pair, when the pair reconstruction accepted the event. An event was considered tagged as non-reconstructible when it was either tagged as MIP or all reconstruction algorithms rejected the event. The performance on the Compton dataset is shown in Figure 5.24. The default reconstruction tags up to 80% of all Compton events correctly at its peak around 550 keV and then drops off steadily at higher energies. The classification by our model outperforms the default classification, giving significantly higher recall values for each energy bin of the crab spectrum on the full Compton dataset. The precision on the low-energy bins is nearly perfect in both classifications with a slight edge at higher energies for the model classification. An important subset constitute fully absorbed Compton events, which have the highest likelihood to be correctly reconstructed and therefore contain the proper directional information of the event. Here the model classification outperforms the default classification in every but one bin at 550 keV (93 % vs 98 %) by up to 38 percent-points.
The performance on D1 and D2 pairs is depicted in Figure 5.25. At energies above the pair threshold and below 5 MeV, the model as well as the default classification show very little discrimination power. Nonetheless, the model's recall on D2-first pairs shows an improvement over the default case. The D1-pair recall shows large improvements over the default classification. The model is starting to recognize D1-pairs reliably at energies larger than 5 MeV and shows a very good class-coverage of

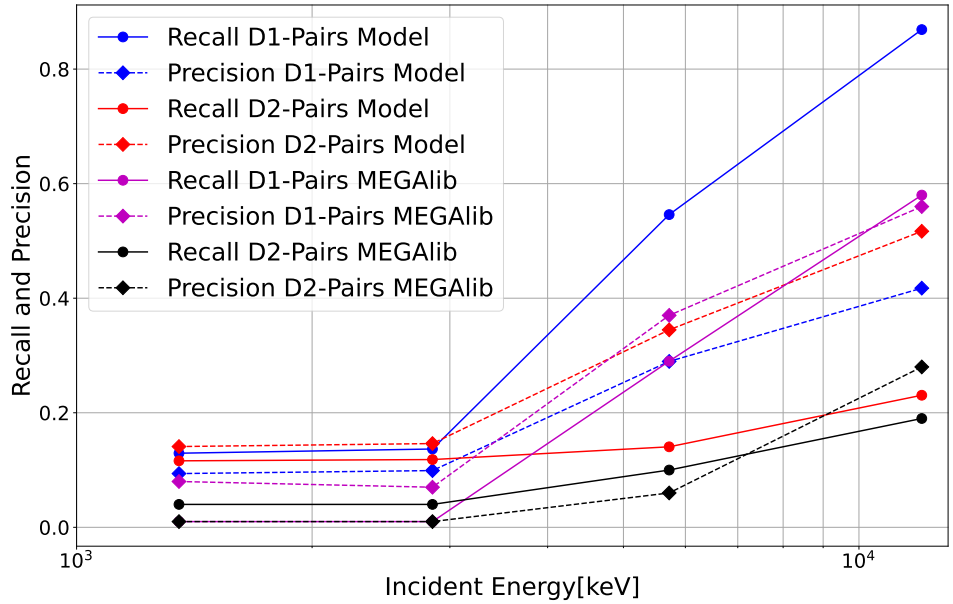


Figure 5.25: Precision and Recall for Pair and non-reconstructable events as function of energy for the Crab-source dataset.

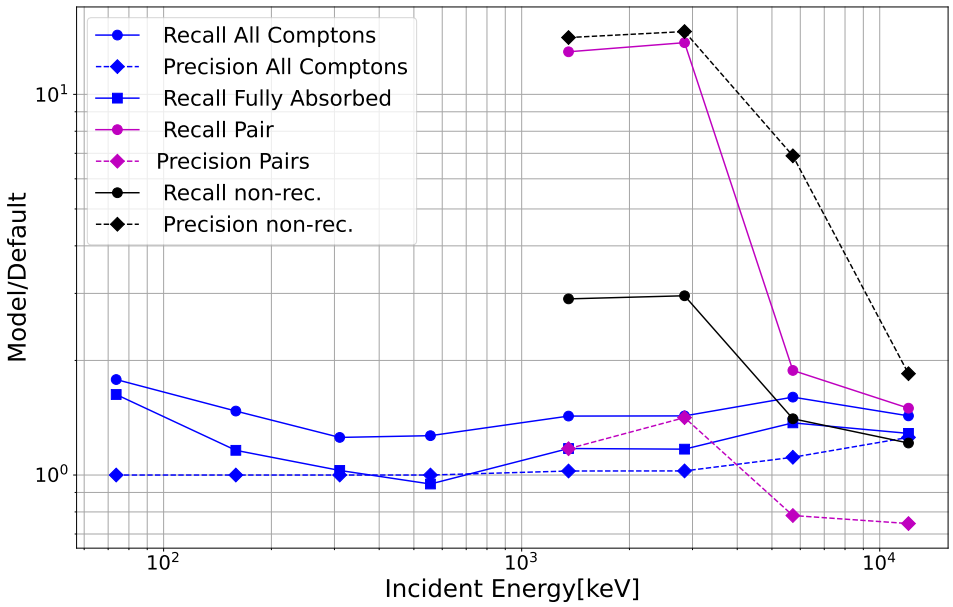


Figure 5.26: Relative improvement of the model classification vs the default reconstruction as function of energy for the Crab-source only dataset.

87 % at the highest simulated energies. The Precision of the model classifications on D1 and D2-pairs shows a slightly worse purity of the pair datasets after classification than in the default case, which can be explained by the relatively large amount of incorrectly tagged D2-pairs which have been tagged as D1-pair and vice-versa (see Figs. 5.18a, 5.18b and 5.18c).
Figure 5.26 shows the relative improvement of the model classification over the default case, given as the ratio of model performance relative to the default. Data points above 1 indicate improvement, below 1 indicate worse performance than the default. The recall on all Comptons, D1 and D2 pairs shows an overall improvement of around 25% to 60% over the default classification. In the low energy bins, the improvement ratio for D1-pairs reaches beyond a factor 10 while maintaining an improvement ratio of greater than 1.5 for high energies. Even though the recall on D2-pairs in this energy range is still sub-optimal, the model's performance is an improvement over the default classification over the whole inspected energy range, improving the class coverage by a factor 3 for lower energies to a factor of 1.3 at the higher energies. The observed behaviour of the model classification is consistent with the expectations from the omni-directional test set.

5.9 Summary and Outlook

The primary goal of this study was to explore the possibility to detect event types and event topologies with CNN-driven models. The successful model presented here is capable of distinguishing different event types and undesired event topologies, as shown on an omni-directional test set as well as on a simulated Crab measurement.
This model was trained to distinguish Compton, pair and non-reconstructible events. The latter consisted of signal photons with a non-reconstructible event topology and events caused by leptons, like cosmic and albedo electrons and positrons. Of course there are even more types of background events, like events resulting from delayed decays of activated materials, and events caused by background hadrons. These components will be the subject of a follow-up study.
The presented model has a computational overhead of about 0.9 GOPS per event. which fits well into the computational envelope of recently developed embedded systems. An architecture like the presented one is capable of processing even larger rates for smaller inputs. Studies are under way to test these embedded systems, making an online-approach viable, e.g., for CubeSat scenarios.
The class-wise recall of the trained model represents an improvement compared to the class coverage of a previously trained random forest.
One particular aspect of this architecture remains unexplored. Since Spatial Pyramid Pooling enables the use of dynamic image sizes, it is possible to decrease the number of necessary operations per event even further by cropping the input images to a smaller size. For computational estimates, the average image size then becomes a variable of the measurement environment. Such a study would make sense once a mission is funded and various orbit scenarios have to be explored.

Chapter 6

Simulations studies for the COMCUBE concept

6.1 CubeSats

The planing time line of a spacebourne science experiment is, depending on the main agency and the scale (e.g. NASA SMEX vs ESA M-Calls), in the range of 4 - 15 years. The financial envelope of these mission calls can quickly reach the order of 100 Million €. Small entities like individual universities are not able to accommodate the financial demands of a full scale science satellite mission. But not all science topics need to have a singular, large detector platform. Some would even benefit from orchestrating smaller swarms of units to achieve better coverage. The main cost driver for space missions is the wide usage of single-use components designed specifically for the developed mission. This results in each launched science satellite essentially becoming its own prototype. Stanford University and California Polytechnica tried to mitigate this problem by defining the CubeSat standard [Rev22] in 1999. It defined the basic 1U cube with basic dimensions ($11.35 \cdot 10 \cdot 10\,\mathrm{cm}^3$ with a usable volume of $10 \cdot 10 \cdot 10\,\mathrm{cm}^3$), the location of connectors and fixed the maximum total mass of 1.33 kg as shown in Figure 6.1.
Larger volumes can be achieved by increasing the number of Us to 2U, 3U, etc. The immediate impact of this standardization was a significant drop in component costs due to the fact of the presence of a standard to produce against. Their small size (a 6U-CubeSat has $30 \cdot 20 cdot 10\,\mathrm{cm}^3$ dimensions, which is barely larger than a shoe-box) enables to be launched as secondary payload together with larger satellites or multiple CubeSats, which reduces the launch costs per unit drastically. The reduced costs led to a proliferation of satellite usage to smaller entities like small countries, private groups and university run research and development projects like UWE3 [Bus+15] or UWE4 [KBS20]. UWE stands for University Wuerzburg Experimental satellite. They are used as technology test-beds and demonstrators. UWE3 is a 1U CubeSat launched in 2013 and is used to test attitude determination and control systems for swarm formations. UWE4 was developed together with the Technical University Dresden and

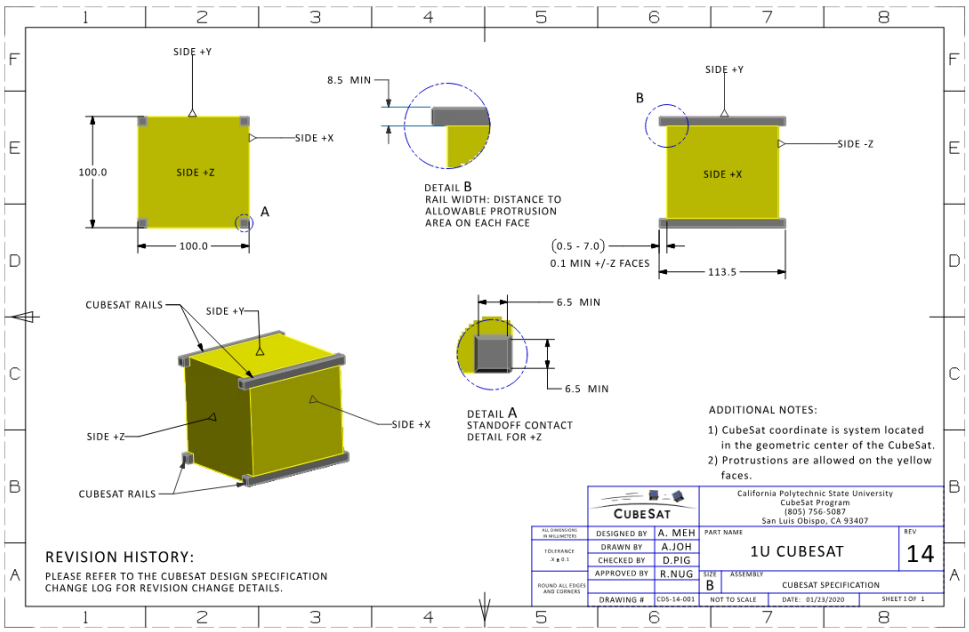


Figure 6.1: Specification of a 1U CubeSat. The 1U is the building block for cube-sat nano and mini satellites. Larger volumes can be created by either stacking 1U volumes in x-y-z thereby creating respectively larger 2U, 4U, 6U and so forth satellites. From [Rev22].

launched in 2018. It is the first 1U CubeSat with an electric propulsion system called NanoFEEP (Nano field-emission electric propulsion system [BT18]). Both satellites are still in orbit.

The low cost of these systems allows to test new designs and equipment in realistic conditions and therefore accelerate the acquisition of a sufficient technical readiness level for new systems to be deployed in large scale missions. The NASA developed BurstCube mission [Mye+24] was launched in March 2024 and ceased operations in mid September 2024. It is an example of a CubeSat as scientific detector. The payload consists of 4 cylindrical thalium doped caesium iodide crystals read out by an array of silicon Photomultipliers (SiPMs) arranged in a pyramidal geometry. The 45 degree angle between each crystal module enables the position reconstruction of the measured GRB similar to Fermi-GBM. A sketch of a Single Quarter Detector (SQD) is shown on the left of Figure 6.2. A sketch of the satellite bus is shown on the right of Figure 6.2. A picture of the whole satellite is shown in Figure 6.3 The total volume of the satellite was 6U.

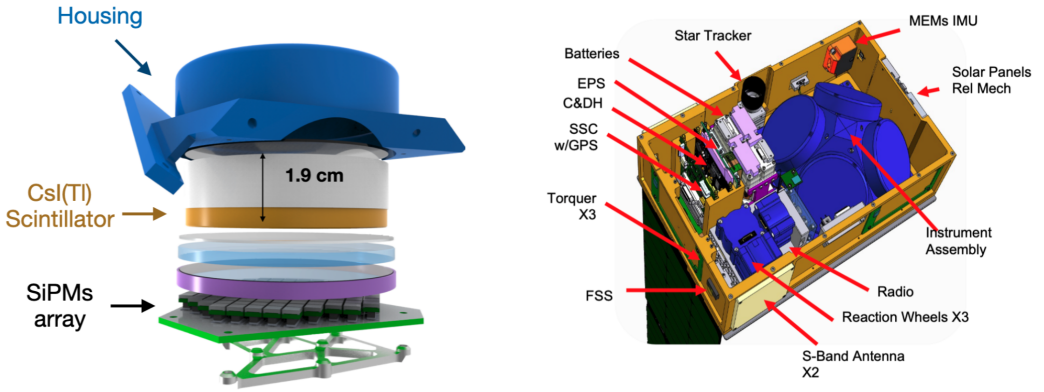


Figure 6.2: Left: Schematic of a SQD. Each module consists of a doped CsI cylindrical crystal with 90 mm diameter and 19 mm thickness. Each SQD is read out by an array consisting of 116 Hamamatsu 13360-6050 SiPMs. Right: Schematic of the payload. The 6U volume is shared between the detector (roughly 4U on the right) and the auxiliary systems on the left. The satellite systems include a star tracker, inertial measurement units, sun sensors, three reaction wheels for attitude control allowing 0.5 degree pointing accuracy and 0.1 degree pointing knowledge and a GPS module to track the orbit position. Onboard computing is provided from a Zynq-7020 SoC. More details can be found in [Mye+24].

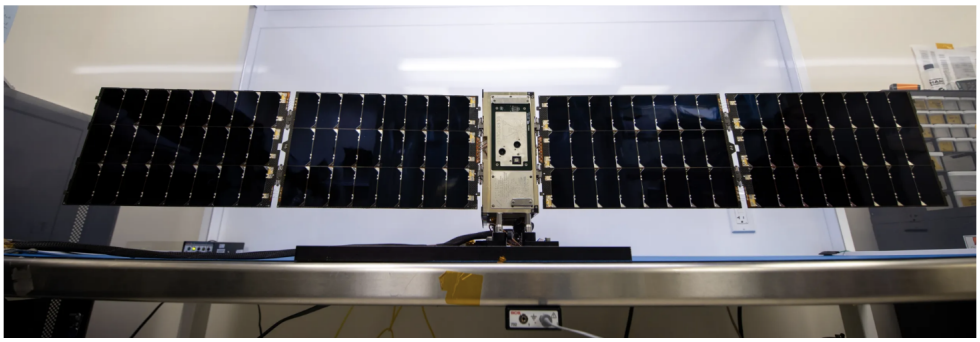


Figure 6.3: Photo of the assembled satellite with deployed solar arrays.

6.2 COMCUBE Design Proposal

COMCUBE [Lav+21] is a design study for a GRB monitor similar to BurstCube funded by the EU AHEAD2020 project. Its main science goal is to improve our understanding of astrophysical jets by giving access to integrated and time resolved GRB polarimetry. The design is using Double-sided Silicon Strips Detectors (DSSDs) as scattering detector and $CeBr_3$ crystal arrays as calorimeter arranged in 1U modules. The instrument design is able to use Compton scatter as its primary detection mode. The

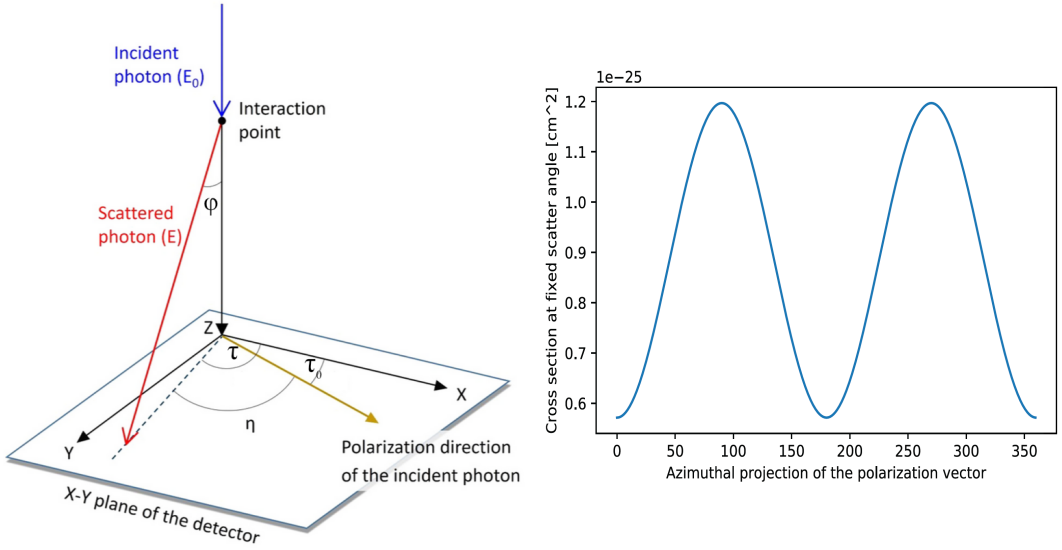


Figure 6.4: Left: Sketch of the kinematics of a Compton scatter of a linearly polarized photon. Adapted from [Kom+17]. Right: Modulation of the cross section as function of the polarization angle for a fixed scatter angle. The cross section reaches the maximum for a polarization perpendicular to the scatter angle.

Klein-Nishina cross-section for linearly polarized photons is given by (see eq. 3.8):

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2}\left(\frac{E_\gamma}{E_i}\right)^2 \cdot \left(\frac{E_\gamma}{E_i} + \frac{E_i}{E_\gamma} - 2 \cdot \sin^2\varphi \cdot cos^2\eta\right)$$

The polar angle φ is given by the position of the source in the sky relative to the detector. The cross section therefore varies with the azimuth projection of the scattered photon relative to the polarization direction of the incident photon η, i.e. the polarization of the incident photon leaves an imprint in the azimuthal scatter angle distribution. If there is no direct access to η, one can measure the projection angle with respect to a fixed axis τ and determine the polarization angle τ_0. The azimuthal scatter angle is then $\eta = \tau - \tau_0$. The right panel of Figure 6.4 shows the cross section modulation for a fixed scatter angle. The respective scatter angle reaches its maximum, whenever the direction of the scattered photon is perpendicular to the linear photon polarization direction and minimal for parallel directions. The azimuthal scatter angle distribution is therefore showing a harmonic oscillation as depicted in Figure 6.10. The polarization angle can be read of by the distance from 0 to the first minimum, the amplitude corresponds to the strength of the linear polarization. Therefore COMCUBE would be able to not only work as GRB-monitor but also as GRB-Polarimeter with intrinsic localization capabilities. This is in contrast to, e.g., POLAR [Pro+18], which was working solely as polarimeter and depended on source localization from other instruments via temporal coincidence.
The baseline design of COMCUBE is a 6U satellite similar in size to BurstCube. The

Table 6.1: Dimensions of the detector subsystems in the mass models. For details of the detector simulation see Section 5.3.1.

Detector	Size	Detector Simulation
Tracker(Si)	$2\ 68 \cdot 68 \cdot 15\ \mathrm{mm}^3$	2D Strips
Calo ($CeBr_3$)	$4\ 25 \cdot 25 \cdot 20\ \mathrm{mm}^3$	Anger Camera
Side ($CeBr_3$)	$2\ 51 \cdot 51 \cdot 10\ \mathrm{mm}^3$	Anger Camera
Plastic (p-Terphenyl)	$4\ 25.6 \cdot 25.6 \cdot 10\ \mathrm{mm}^3$	Anger Camera

detector occupies 4U of the satellite volume. Each U of the detector volume contains two DSSD layers with $68{\cdot}68\mathrm{mm}{\cdot}15\,\mathrm{mm}^3$. Each layer is read-out with 32 strips in x and y. At the bottom is a $CeBr_3$ calorimeter with $51{\cdot}51{\cdot}20\,\mathrm{mm}^3$ segmented into a 2 by 2 crystal array . In addition to this, the baseline includes two $CeBr_3$ detectors with half the thickness of the bottom calorimeters on the sides. Each U is arranged such that the active wafer of the DSSD layers are rotated to the inside. A small 6U satellite offers of course only limited sensitivity to γ-rays. The envisioned operation mode of COMCUBE is therefore in a swarm of multiple satellites and with that allowing for an ideally full coverage of the sky at all time and the combined sensitivity of a larger satellite. There were two design variation inspected over the course of the design phase. One was a control design, in which the side crystal arrays were removed (dubbed "no side" or "step down"). The purpose of this design was to investigate whether the removal of the side detectors is an option for stepping down the complexity of the instrument and inspect the effect of the additional detectors on overall performance and polarization response in particular. The other design option incorporates a layer of segmented plastic scintillator between the DSSDs and the bottom $CeBr_3$ calorimeter (dubbed "with plastic" or "step up"). Figure 6.5 depicts the investigated mass model options, while Table 6.1 gives dimensions and MEGAlib detector modelling for each detector system.

6.2.1 Performance Estimates of the single 4U Detector

To determine the final design a comparison study was performed. The data set comprises of mono-energetic far-field point sources simulated in the range of 0.1 to 1.5 MeV for incident angles between 0 and 90 degrees. The performance metrics evaluated were

- Relative energy resolution vs. incident energy
- Angular resolution vs. incident energy
- Effective Area vs. incident energy
- Polarization response to fully linear polarized photons at different energies and incident angles

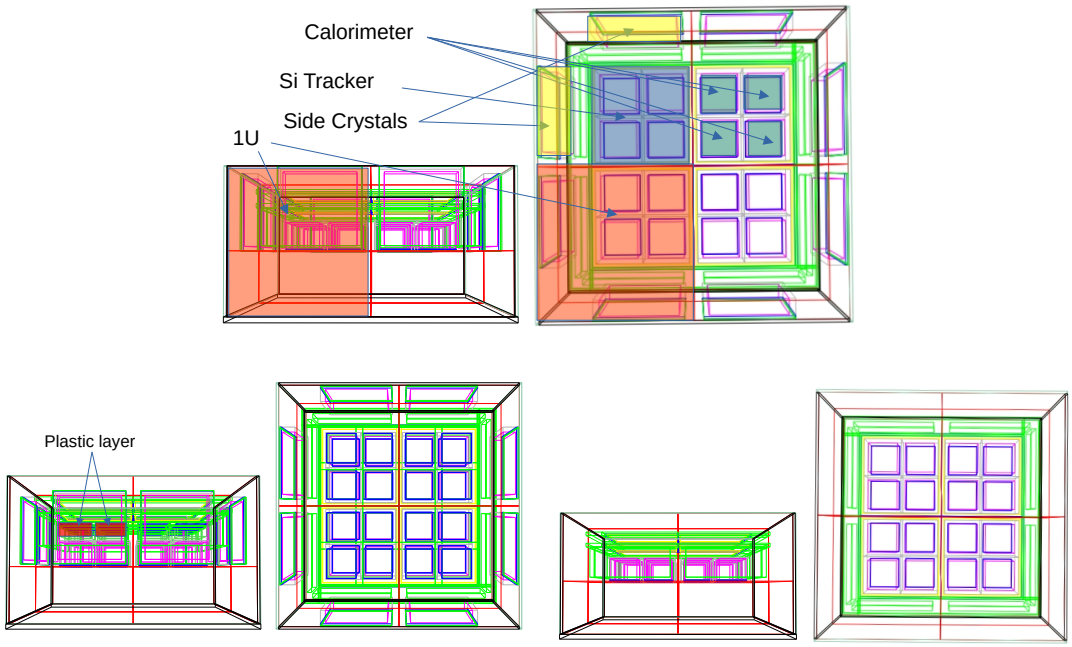


Figure 6.5: The three investigated design option. The baseline design (top) consists of DSSDs as scattering detectors and $CeBr_3$ crystal arrays on the bottom and the side as calorimeters. The design option with plastic scintillaror in between the DSSDs and the bottom crystal array (step-up) is depicted in the bottom left. The step-down option without side calorimeters is shown on the bottom right.

The basis of these performance estimates are MEGAlib simulations (see Section 5.2 for an overview) performed with mono-energetic sources from a range of incident angles. Table 6.2 summarizes the inspected energies and angles. It is normally sufficient to simulate 10^6 events with hit patterns sufficient for the event topologies laid out by the trigger criteria (pretriggered events) to derive statistically stable energy and angular resolution with mono-energetic sources. The trigger criteria used here equate to events with a coincident interaction with any other detector. The actual trigger is checked during reconstruction *after* noising of the deposited energies. To ensure sufficient statistics for evaluating the polarization response, the number of pretriggered events was set to 10^7.
The energy and position resolution of the individual detectors are either given by

Table 6.2: Simulated energies and incident angles.

Incident Energies [MeV]	$0.1, 0.16, 0.29, 0.50, 0.63, 0.79, 1.0, 1.26, 1.58$
Incident Angle [Deg]	$0, 26, 37, 46, 53, 60, 66, 73, 84, 90$
Polarization	100% linear, randomly polarized

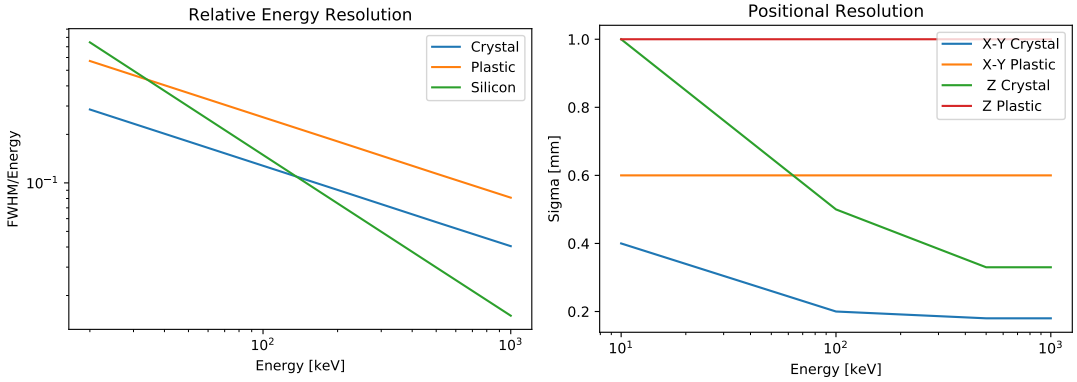


Figure 6.6: Energy and position resolution of the individual detectors as function of energy. The curves are used in the simulation for noising during reconstruction.

their geometry or as input for the simulation. Figure 6.6 shows the input resolutions as function of measured energy of the individual detectors. The position resolution of the DSSD layers is given by the strip pitch b of the layer and equates to

$$\Delta_{x,y} = \frac{b}{\sqrt{12}} = 0.58\,\text{mm} \tag{6.1}$$

The z-coordinate is fixed to the center of the layer. Both calorimeter modules have an energy dependent x-y-z resolution, while the plastic scintillator is assumed to show a flat resolution. The energy resolutions of the individual detectors are modelled with the same approach as in Equations 5.18 for the DSSD and 5.19 for the calorimeter. For the DSSD an electronic noise component of 6.37 keV is assumed, giving

$$\sigma_{DSSD} = \sqrt{\frac{F \cdot E}{w} \cdot w^2 + (6.37\,keV)^2} \tag{6.2}$$

For the $CeBr_3$ crystals a test-bench measurement of 5% at 662 keV was used to estimate the energy resolution leading to

$$\sigma_{CeBr_3} = 14.1\,keV \cdot \sqrt{\frac{E}{662\,keV}} \tag{6.3}$$

For the plastic scintillator no test-measurement existed and a factor 2 worse energy resolution compared to the $CeBr_3$ detectors was agreed upon.
The energy resolution for each simulation is evaluated by fitting the spectrum around the known photo-peak with a Gaussian. The width of the fit σ is then used to calculate the FWHM of the distribution. All events with a deposited energy around one FWHM are then selected to determine the angular resolution.
Since the source position is known, one can determine the Angular Resolution Measure (ARM) defined by

$$ARM = \bar{\varphi} - \varphi_{geo} \tag{6.4}$$

where $\bar{\varphi}$ is the scatter angle calculated from the deposited energies of the first two hits using Compton's formula and φ_{geo} is the scatter angle derived from the geometric locations of the first two hits. If the reconstructed sequence is correct for at least the first two hits, the difference between the two should be small. The resulting distribution is ideally centred at a deviation of 0 and in general not symmetric. It is then fitted with a double Lorentzian + skewed Gaussian. The angular resolution for the given point source is then the FWHM of the fit or the distribution, if the fit fails. Figure 6.7 shows an example for a 300 keV on-axis source. Shown are the spectrum and the ARM-distribution after energy selection.

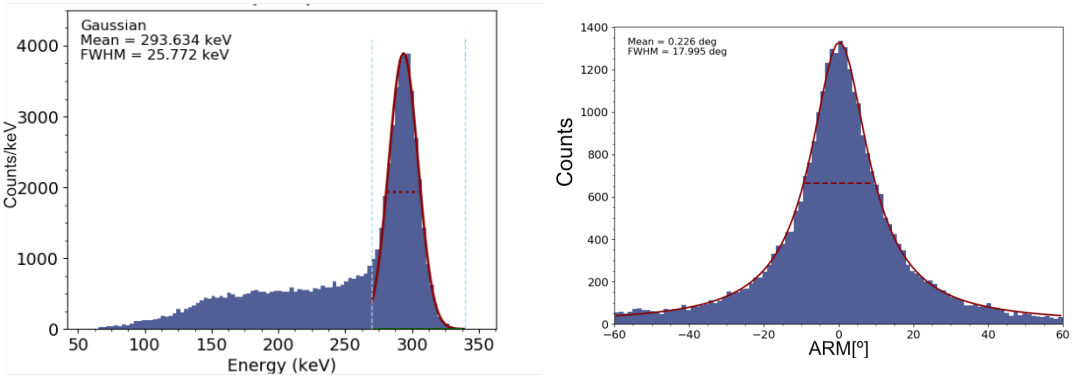


Figure 6.7: Example output for an on-axis 0.3 MeV source. Left: Energy spectrum with the fitted Gaussian. The light blue dashed vertical lines denote the fit region. Right: The ARM distribution after energy selection using the FWHM of the of the underlying spectrum. The dashed red line denotes the width of the FWHM of the fit in solid red.

The energy and angular resolution at each energy and incident angle is then used to determine the effective area. The effective area is the starting area (A_{init}) weighted ratio of emitted photons from the source ($N_{emitted}$) and the correctly reconstructed photons ($N_{rec.}$):

$$A_{eff} = A_{init} \cdot \frac{N_{emitted}}{N_{rec.}} \tag{6.5}$$

The effective area is a measure for the response of the detector to a specific source (here mono energetic farfield point sources). It allows to reconstruct the flux from the number of measured source photons by

$$Flux = \frac{\text{measured photons attributed to the source}}{\text{Effective Area for the Source} \cdot \text{Observation Time}} \tag{6.6}$$

The source setup in the simulation is shown in Figure 6.8 for a farfield point source. The detector is illuminated from a farfield point source in the simulated sky. A computational efficient representation of the sky is a sphere around the detector that encloses it completely, the so called surrounding sphere. Farfield point sources are extended sources assumed at infinity. Their photons arrive therefore in parallel from

a disc tangential to the point of origin on the surrounding sphere. The area of this disc (the starting area of the photons) is given by the radius of the surrounding sphere. Note that the calculated value of the effective area is independent of the size of the surrounding sphere as long as the photons from the disc illuminate the detector completely. A too large radius of the surrounding sphere will only result in a non-efficient simulation.

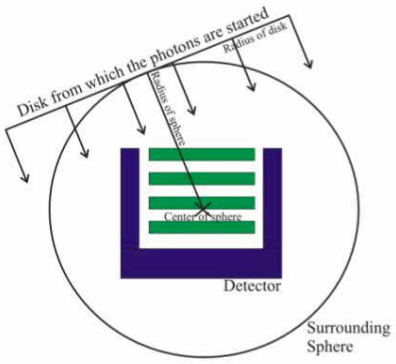


Figure 6.8: Sketch illustrating the source setup in MEGAlib. The detector is placed into the surrounding sphere representing the sky. Photons from a given point in the sky arrive in parallel starting perpendicular from the depicted disc. From [ZAS06].

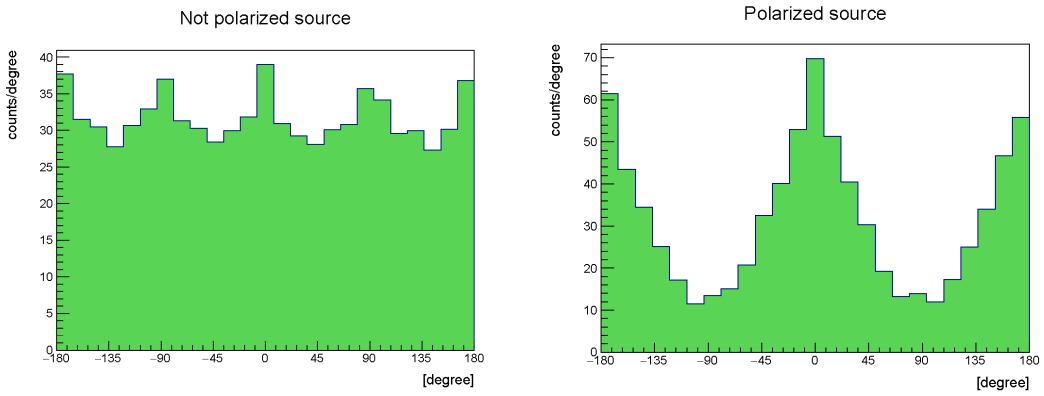


Figure 6.9: Polarigram for an unpolarized (left) and 100% linearly polarized 150 keV source. Generated with MEGAlib [ZAS06].

The response to linear polarized photons is determined using two data sets for each energy and incident angle with the same amount of statistics. One data set is simulated with randomly polarized photons. Since the polarization angles for each photon are randomly drawn, every visible modulation pattern in the scatter angle projection has to be caused by the geometry itself and represents a baseline, that has to be

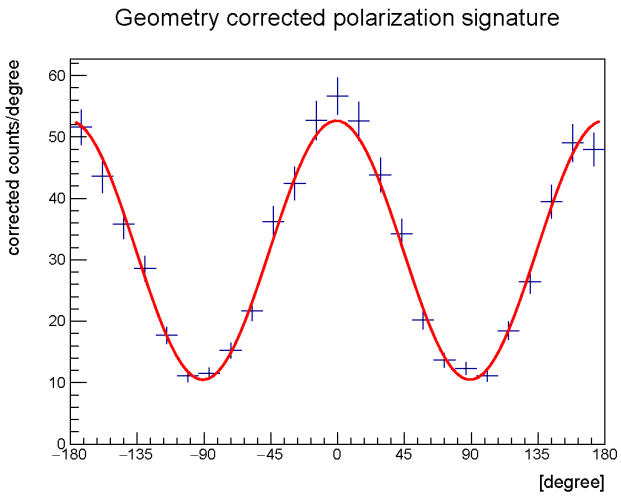


Figure 6.10: Corrected distribution of the azimuthal projection of the scatter angle for 150 keV photons and 100% linear polarization perpendicular to the incident direction of the photon. Generated with MEGAlib [ZAS06].

subtracted. The other data set contains 100% linearly polarized photons with polarization angle perpendicular to the direction of travel of the photon. Since the source position is again known, one can calculate the scatter angle using the first two interactions. Figure 6.9 shows the distributions obtained for the scatter angle projections of an unpolarized (left) and 100% linearly polarized source (right). Even though the modulation pattern is clearly visible in the latter case, also the unpolarized photons show a weak modulation pattern, illustrating the need for a subtraction of the unpolarized response. The resulting pattern is shown in Figure 6.10. It is fitted with a cosine whose amplitude gives the response to 100% linearly polarized photons μ_{100}. The polarization analysis was done with MEGAlib's built-in mimrec analysis tools.

Energy Resolution

The relative energy resolution as function of incident energy for on-axis and 37°off-axis is depicted in Figure 6.11. The energy resolution for both incident angles is roughly equal for each of the designs. The relative energy resolution reaches 5% at around 0.8 MeV and reaches roughly 3% for the highest energies. The baseline design as well as the step-down design without side detectors show a similar energy resolution. The inferior intrinsic energy resolution of the modelled plastic scintillator leads to a slightly worse performance of the step-up design.

Angular Resolution

Figure 6.12 shows the angular resolution as function of incident energy for two incident angles obtained from the ARM distribution after energy selection. The values

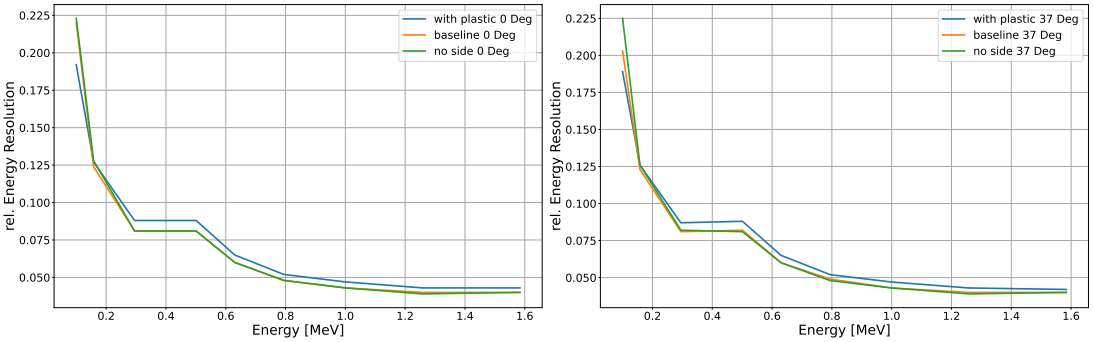


Figure 6.11: Energy resolution as function of incident energy for an on-axis and 37°off-axis source. Error bars are omitted due to size.

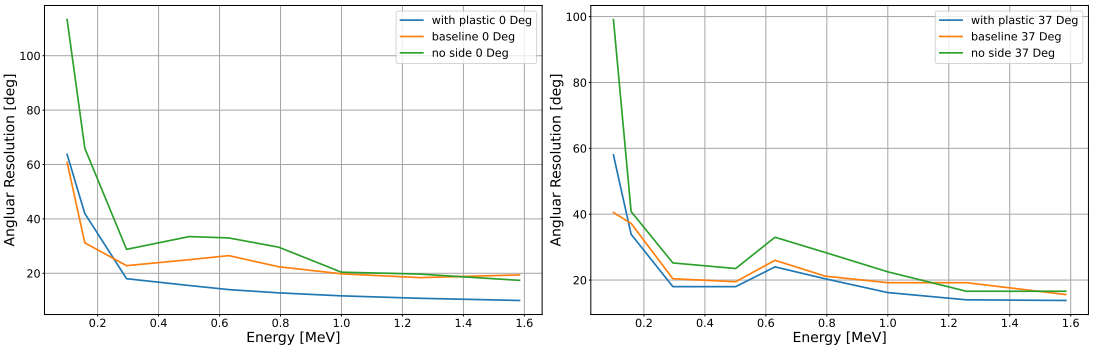


Figure 6.12: Angular resolution as function of incident energy for an on-axis and 37°off-axis source. Error bars are omitted due to size.

obtained for the baseline and step-down design show a kink around 0.6 MeV, where the performance actually worsens compared to lower and higher energies, while at least on-axis the step-up design shows a smooth curve. At 37° all designs show a kink at this energy. This can be explained by the path the respective photons have to take in order to deposit their energy. In the on-axis case the plastic scintillator can take the role of the scatter detector at a certain energy, while the other designs have to rely on the $CeBr_3$ crystals to provide first hit of the sequence. This is amplified in the step-down design, as the photons now have to scatter from the bottom calorimeter either to a neighbouring crystal or up into the DSSDs. In the off-axis case the side-to-bottom calorimeter and side-to-tracker coincidences become more dominant, resulting in a 0.6 MeV kink in all three designs. Because of the larger amount of mass in the path of the photons and the only slightly worse position resolution of the $CeBr_3$ modules compared to DSSDs, the angular resolution improves at this angles for all three designs. The step-down design shows worst performance.

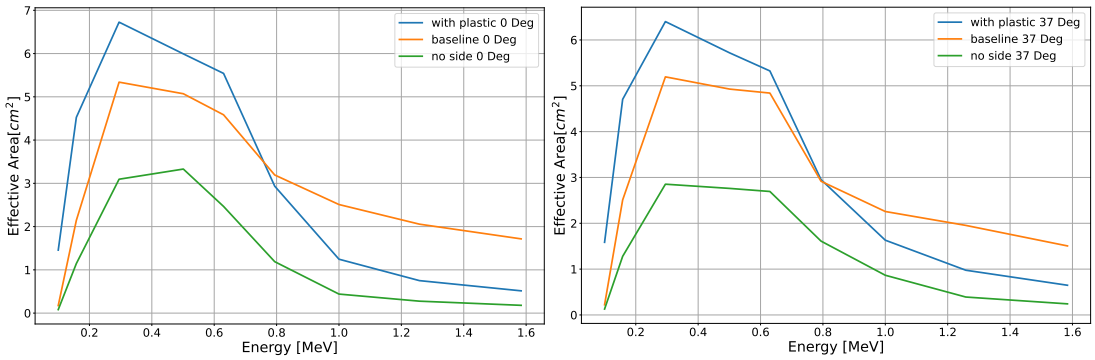


Figure 6.13: Effective area as function of incident energy for an on-axis and 37°off-axis source. Error bars are omitted due to size.

Effective Area

Figure 6.13 shows the effective area as function of incident energy determined after energy and angular selection for the two selected incident angles. The peak effective area for most designs is reached in both cases around 0.3 MeV. The step-down design reaches its peak value at 0.5 MeV in the on-axis case. Overall, the step-up design has a better effective area up until 0.8 MeV, then the baseline design performs better. As expected, the step-down design shows the worst performance over the whole energy range.

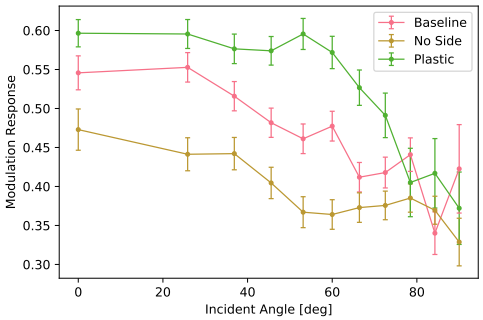
Polarization Response

The response to linear polarized photons is shown in Figure 6.14. Shown is the modulation response μ_{100} as function of the incident angle. At 0.1 MeV the step-up design shows a superior μ_{100} over almost the whole range of incident angles. The baseline design has a roughly 10-15% lower μ_{100}, while the step-down design gives a worse performance over the whole angular range. The picture changes at only a slightly higher incident energy of 0.158 MeV drastically. Here the baseline design is able to outperform the other two designs significantly. For higher energies the baseline design is able to outperform the step-up design, but with lesser decisive margins. The step-down design is able to give a similar response to the step-up design at 0.3 MeV and above.
Effective area and μ_{100} are the two main parameters of the minimum detectable polarization (MDP). It is defined as [WEO10]

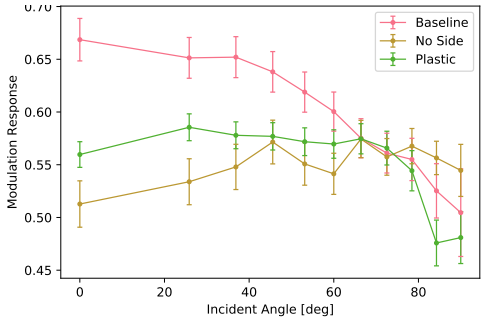
$$MDP_{99} = \frac{4.29}{\mu_{100} \cdot S} \cdot \sqrt{S + B} \tag{6.7}$$

For very bright GRBs the background becomes negligible and one obtains:

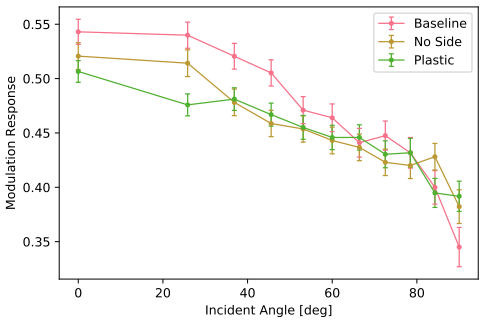
$$MDP_{99} = \approx \frac{4.29}{\mu_{100} \cdot \sqrt{S}} \tag{6.8}$$



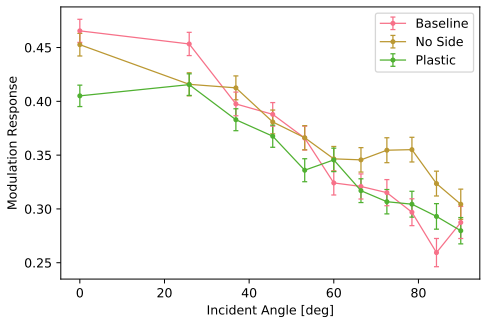
(a) 0.1 MeV



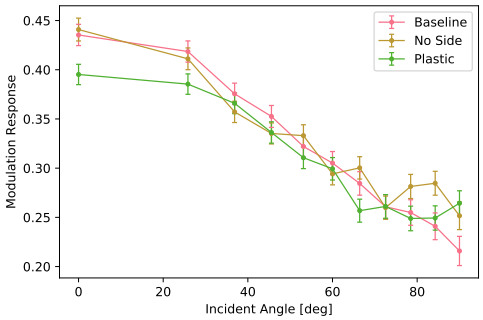
(b) 0.158 MeV



(c) 0.294 MeV



(d) 0.5 MeV



(e) 0.6 MeV

Figure 6.14: Response to 100% linear polarized photons for various energies.

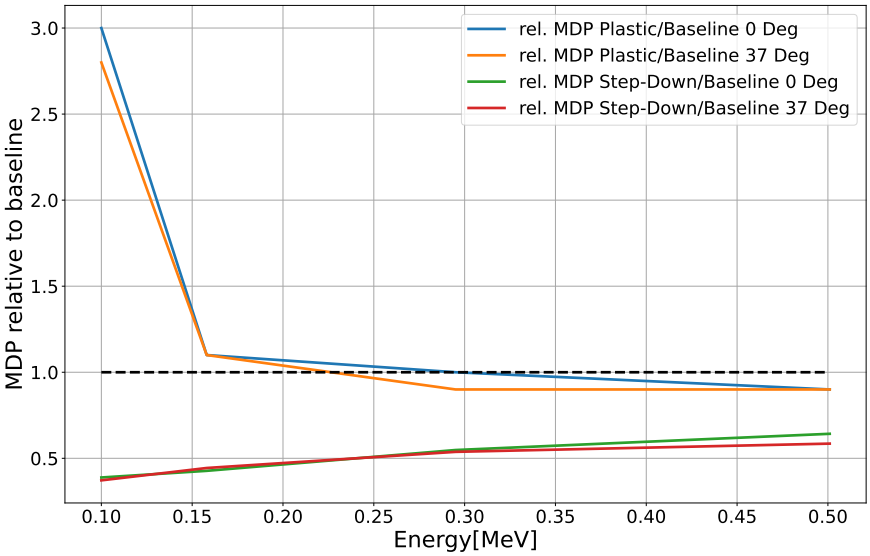


Figure 6.15: Expected improvement of the MDP relative to the baseline design.

Since the source count is linearly proportional to the effective are one can estimate the relative improvement in MDP at a given energy for two designs by

$$\frac{MDP_{99}(A)}{MDP_{99}(B)} = \frac{\mu_{100}(B) \cdot \sqrt{A_{eff}(B))}}{\mu_{100}(A) \cdot \sqrt{A_{eff}(A)}} \tag{6.9}$$

Figure 6.15 shows the relative improvement in MDP at given energies as ratio relative to the baseline design. The step-up design increases the MDP at 0.1 MeV over the baseline design significantly. This is due to the very low effective area of the baseline design at this energy, since the detector is close to its measurement threshold. The step-up design benefits the most from the additional mass and the lower trigger threshold of the plastic modules at energies below 150 keV. For higher photon energies, this effect wears off. At 0.16 MeV the expected improvement is in the order of 10%. The relative MDP is even dropping below the baseline design for higher energies. The still larger effective area can not compensate sufficiently for the loss in polarization response at these energies. The presence of the side-detectors is turning out to be crucial for the MDP as well. The step-down design is hampered by a slightly lower μ_{100} as well as a significantly lower effective area over the whole energy range, yielding the worst performance of all three inspected options. The decrease in performance is in the range of 60% to 40%.

6.2.2 GRB Sensitivity for a constellation of 4U

The performance of multiple baseline COMCUBE-satellites flying in a swarm was estimated by A. Ulyanov [Uly20]. The average GRB-response was determined using a reference GRB with $\alpha = -1.1$, $\beta = -2.3$ and $E_{break} = 300\,\text{keV}$. The fluence was set to $10^{-2}\,\text{erg}\,\text{cm}^{-2}$ as parameters of the Band-spectrum (see eq. 2.5.2). The simulated GRBs are generated in 10 incident angles with equal solid angle. For determining the polarization response μ_{100}, the energy band from 100 keV to 1 MeV was selected and an ARM-cut of 30° was applied. To estimate the background in each directional bin, the built-in MEGAlib background generator was used to model the orbit conditions at an altitude of 560 km and a cutoff rigidity of 5 GV. The MDP at a confidence level of 99% is calculated using Equation 6.7

$$MDP_{99} = \frac{4.29}{\mu_{100} \cdot S} \cdot \sqrt{S + B}$$

where S is the signal count and B is the background count for the given angular element derived from the background simulation. To now obtain the signal count for a constellation with a given number of COMCUBE satellites the Fermi-GBM [von+20] and CGRO-BATSE [Mee+98] catalogues are used. For each GRB, S and B are calculated by

$$S = 0.9 \cdot \frac{F}{F_0} \cdot S_0 \tag{6.10}$$

and

$$B = T_{90} \cdot b, \tag{6.11}$$

where b is the simulated background rate, T_{90} is the time interval in which 90% of the photons where detected, $\frac{F}{F_0}$ is the ratio of the fluency of the given GRB in the energy range relevant for the polarization analysis F relative to the reference GRB fluency (for BATSE GRBs 100 to 300 keV and $F_0 = 2.18 \cdot 10^{-3}\,\text{erg}\,\text{cm}^{-2}$, for GBM GRBs 100-700 keV and $F_0 = 7470\,\text{cm}^{-2}$) and S_0 being the source counts of the reference GRB in the given band. The first step determines the number of able observing satellites and the directional bin of the individual observation. The individual MDP are taken from the response simulations and combined using the inverse square sum:

$$\frac{1}{MDP^2} = \sum_i \frac{1}{MDP_i^2} \tag{6.12}$$

The procedure is then repeated 30 times for each GRB in the catalogues. The resulting weighted cumulative distribution is shown in Figure 6.16.
A swarm with a size of four satellites would be able to observe five GRBs with a polarization of 30% or less, a constellation of 16 satellites is projected to observe 20 GRBs at an MDP of 30% or less.

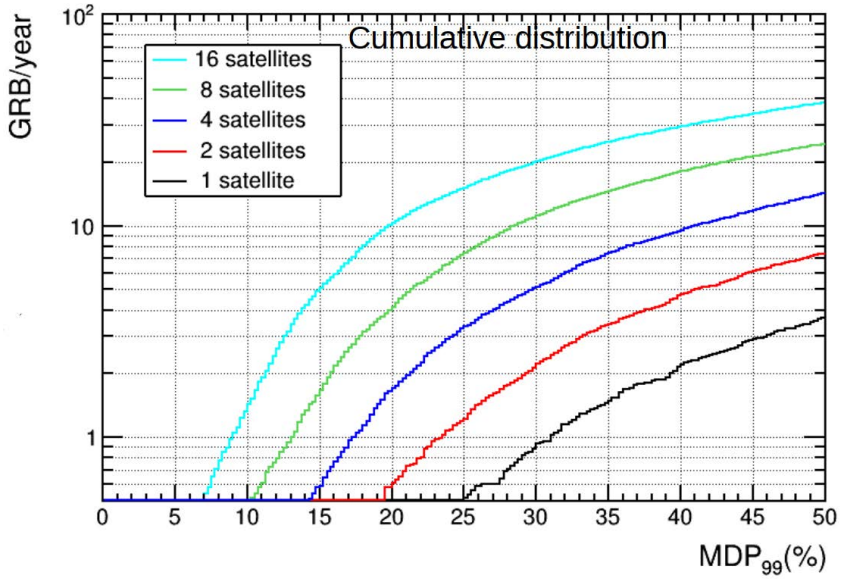


Figure 6.16: Expected GRB per year as function of MDP for various swarm sizes. From [Uly20].

6.2.3 Summary

This comparison study led to the selection of the design with side detectors and without plastic scintillator. Even though the additional mass promises a larger effective area, the impact of the increased source count is off-set by the lower μ_{100} at energies above 250 keV. The design without side detector was found to be lacking in both polarization response and efficiency. The presence of the side detectors was found to be of utmost importance. Figure 6.17 shows the peak-energy obtained from Band-function fits of GBM-catalogue. The histogram contains 3512 GRBs. The peak energy in the Band-spectrum denotes the point in the spectrum where the emission mechanism is assumed to change, e.g. from synchrotron emission with relatively low amount of linear polarization to inverse Compton scattering with potentially large amounts of linear polarization.
The improvements in MDP in the step-up design are in the range below the average break energy, while the baseline design promises a better polarization sensitivity for photons after the break energy. All designs reach their peak effective area at 0.3 MeV, which is roughly at the break energy. At this energy and above, the baseline design promises a better polarization response. The better response, the lesser strain on the mass budget and the lower degree of payload complexity led to the decision to pursue the baseline design. Figure 6.18 shows the simulated polarization response of the baseline design in comparison to the simulated response of the POLAR experiment [Pro+18]. POLAR's simulated response to linear polarization is lower than the estimates for the baseline COMCUBE due to measuring only the projection of

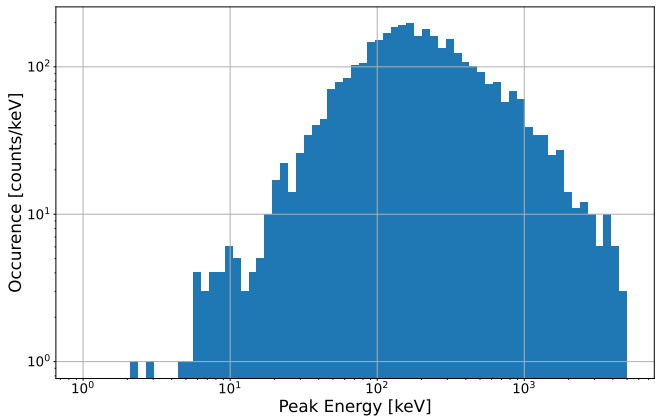


Figure 6.17: Distribution of the peak-energy of the fitted Band-function from the Fermi-GBM catalogue. The histogram contains 3512 GRBs. Generated from [von+20].

the scatter angle. POLAR's sensitivity towards polarization is mostly driven by its large effective area due to the large detector mass. On the other hand, the direct access to the Compton scatter angle enables COMCUBE to be more sensitive towards the polarization of the measured photons. A sufficiently large swarm of COMCUBE satellites is therefore able to detect GRBs with an MDP that is competitive with large scale experiments.
A modified design of COMCUBE is currently in a Phase A study at ESA for technical feasibility dubbed COMCUBE-S [Fra+25].

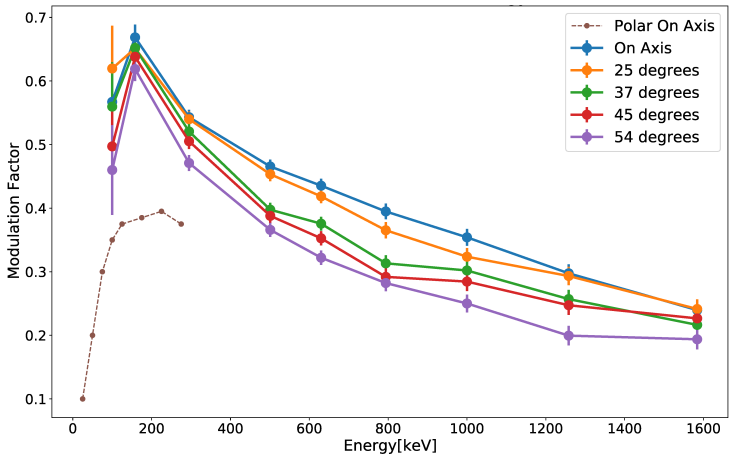


Figure 6.18: Modulation factors of the baseline design for multiple energies as function of incident angle. For comparison the POLAR modulation factor is given from:[Pro+18].

6.3 1U-Prototype Flight

The AHEAD-2020 funding included allocations for a test of a prototype on a stratospheric balloon flight. The Strato-Science campaign managed by the french aerospace agency CNES was chosen as prime candidate. The 2023 campaign was scheduled to be held in Timmins (Ontario). The site has the advantage of low population density, which allows to fly larger gondolas. Since the balloon launch was shared with other experiemtnts, the following flight constraints had to matched by the COMCUBE prototype:

- 30 kg total masss
- -10 g vertical and ± 5 g horizontal shock resistant to survive a "soft" landing
- power draw needs to allow sustained operations of 50 h
- telemetry limit of 300 kb/s down and 15 kb/s uplink

These conditions led to the design decisions that will be described in the following.

6.3.1 The COMCUBE Balloon Prototype

The prototype detector is very similar in design to 1U of the baseline design. Its layout differs in that there is only of one instead of two side detectors. Also the bottom calorimeter was redesigned with Gagg-crystals instead of $CeBr_3$. The tracker consists of two commercially available Micron BB7v2 DSSDs with 32 strips in x and y. The active area of each layer is $64 \cdot 64\,mm^2$, resulting in a strip pitch of 2 mm. A mounted DSSD is shown in Figure 6.19.
The side calorimeter is repurposed from the MAUD project [Leb+18]. It consists of a monolithic $CeBr_3$ crystal with a size of $5{\cdot}5{\cdot}1cm^3$. It is enclosed in a 1 mm aluminium case with a quartz window at the bottom. The scintilation signal is read out by 4 4×4 Hamamatsu S14 Silicon Photomultipliers (SiPMs) arrays [Ham]. The emission spectrum of $CeBr_3$ peaks at 380 nm which is close to the peak efficiency at 450 nm of the selected SiPM model. The main advantage of $CeBr_3$ is its high light yield of 60 photons per keV and fast decay time in the order of 20 ns. Handling of the crystal is more challenging due to its highly hygroscopic nature. It has a density of $5.1\,g\,cm^{-3}$ and an effective Z of 45.9. The CAD-drawing and the disassembled detector are shown in Figure 6.20.
The bottom calorimeter is a redesign of the Gamma-ray Module (GMOD) [Mur+22] flown on EIRSAT-1 [Mur18], a 2U university-class CubeSat developed by the University College Dublin for the "Fly your satellite!"(FYS) announcement of opportunity of the European Space Agency (ESA). In contrast to GMOD the calorimeter for the COMCUBE prototype consists of 4 modules with 16 $6.1 \cdot 6.1 \cdot 15mm^3$ cerium doped gadolinium aluminium gallium garnet (GAGG) scintillators coupled to 16 SiPMs each. Each crystal module has a total size of $25 \cdot 25 \cdot 15\,mm^3$. The size of the crystals was chosen to match the size of the SiPMs. The scintillation spectrum

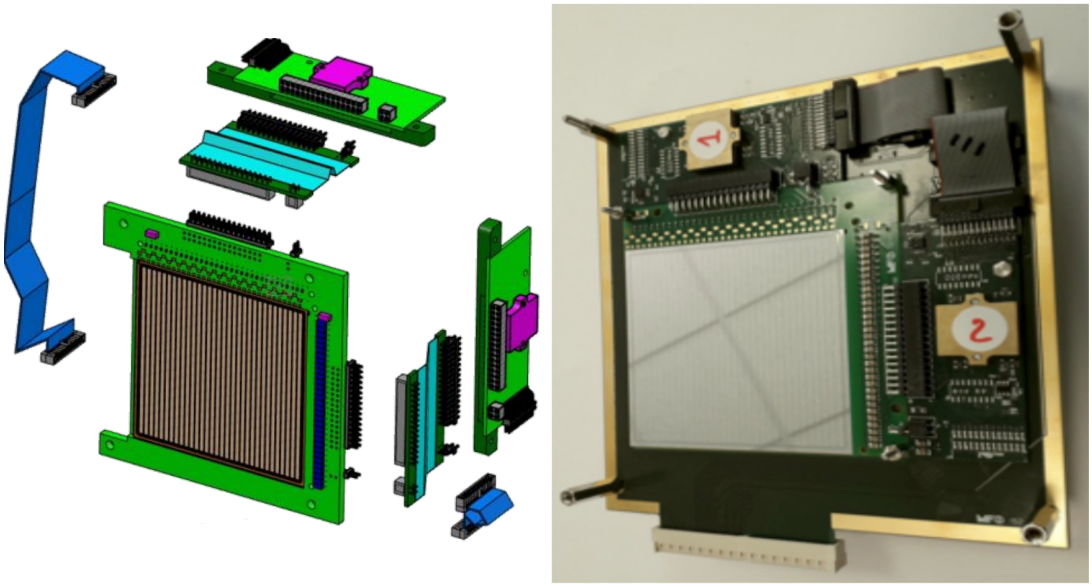


Figure 6.19: CAD drawing and final design of the DSSD detectors. The Micron BB7 DSSD is framed by the front-end electronic boards from two sides. This allows to allign the layers with minimal passive material in between DSSD modules.

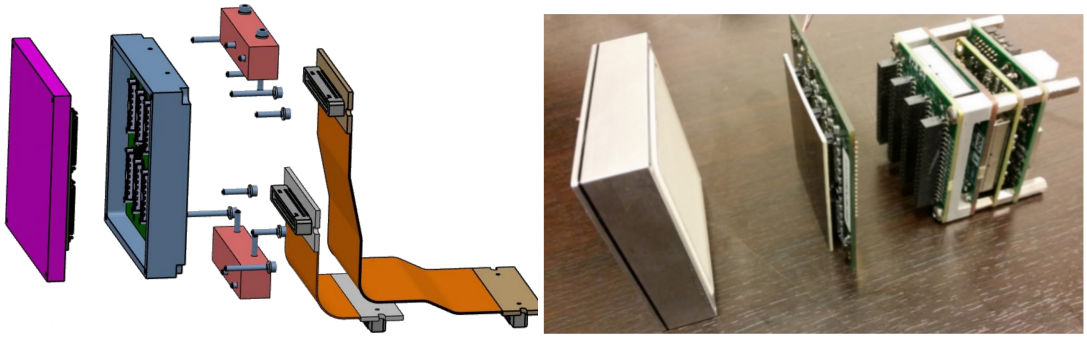


Figure 6.20: CAD drawing and final design of the MAUD detector. The monolithic crystal in its aluminium housing is on the left. The SiPM array is attached to the quartz-window at the bottom. The front-end electronic is connected via two flat-band connectors.

of the Epic GAGG-HL [Epi] crystals peaks at 530 nm, which makes it complicated to match to the efficiency of SiPMs, resulting in a lower quantum efficiency compared to a $CeBr_3$ array. It has a slightly lower light yield at around 54 photons per keV and a slower decay time of 150 ns. The main advantage of GAGG over $CeBr_3$ is the far easier handling of the crystals as the crystals are non-hygroscopic, dropping the need of a hermetically sealed crystal housing. It has a higher density and effective Z than $CeBr_3$ at 6.6 $g\,cm^{-3}$ and 50.5, respectively. A test bench measurement determined an

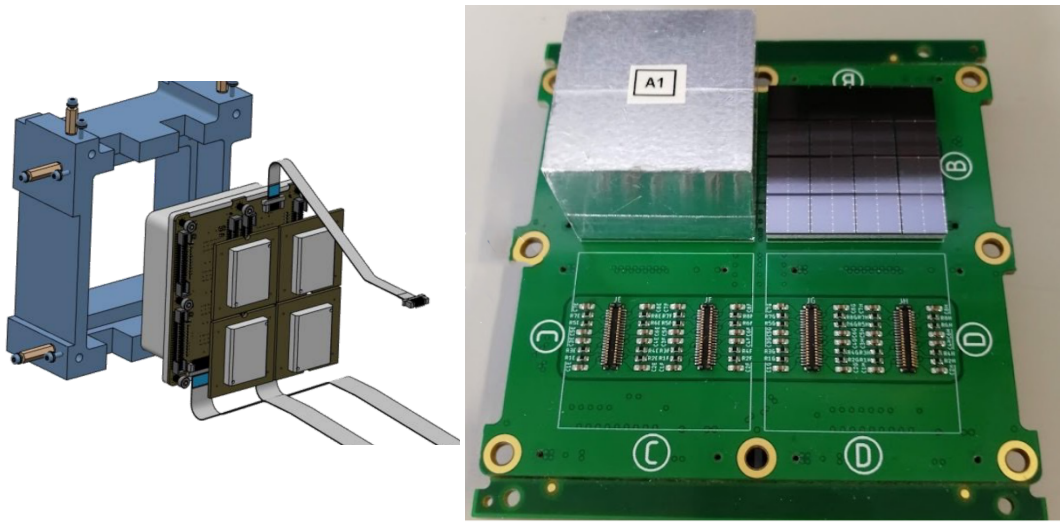


Figure 6.21: CAD drawing and final design of the UCD detector. Four crystal modules with 4·4 SiPMs and crystals each are combined into the D2A detector module. A non-mounted SiPM array is placed in slot B to illustrate the crystal matrix of each module. The whole assembly is housed by an aluminium holding structure, also used to mount D2B. Adapted from [Mur23].

energy resolution of 9.1% at 662 keV, which was used to model the energy resolution analogous to Equation 5.19. Figure 6.21 shows a CAD drawing and the main PCB with one crystal module and one SiPM array mounted for illustrative purposes.
The mass model used for the simulation estimates is shown in Figure 6.22. The setup is housed inside an enclosure made of 5 mm aluminium flushed with nitrogen to protect the setup from humidity induced sparks during ascend and descent and falling ballast from the top.
Figure 6.23 shows the CAD design in its mature stage and the assembly before the flight. The holding structure is a 3D printed design and provides the required support during landing and wind gusts during ascend and descend. The assembly is mounted on a stainless steel flange, on which the enclosure is mounted.

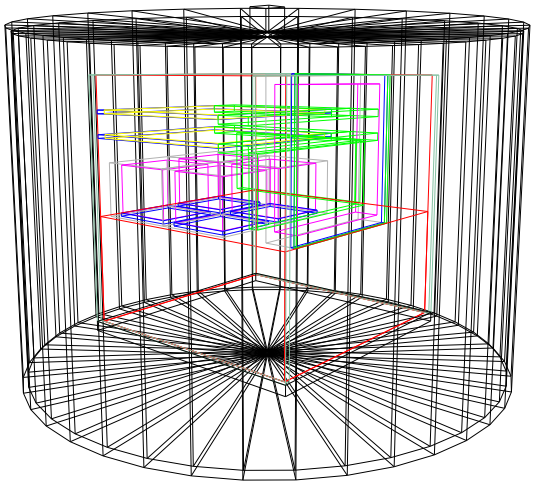


Figure 6.22: COMCUBE balloon prototype mass model. The enclosure is modelled as closed aluminium cylinder with 5 mm wall thickness. In contrast to 1U of the proposed design, the prototype design includes only one side detector.

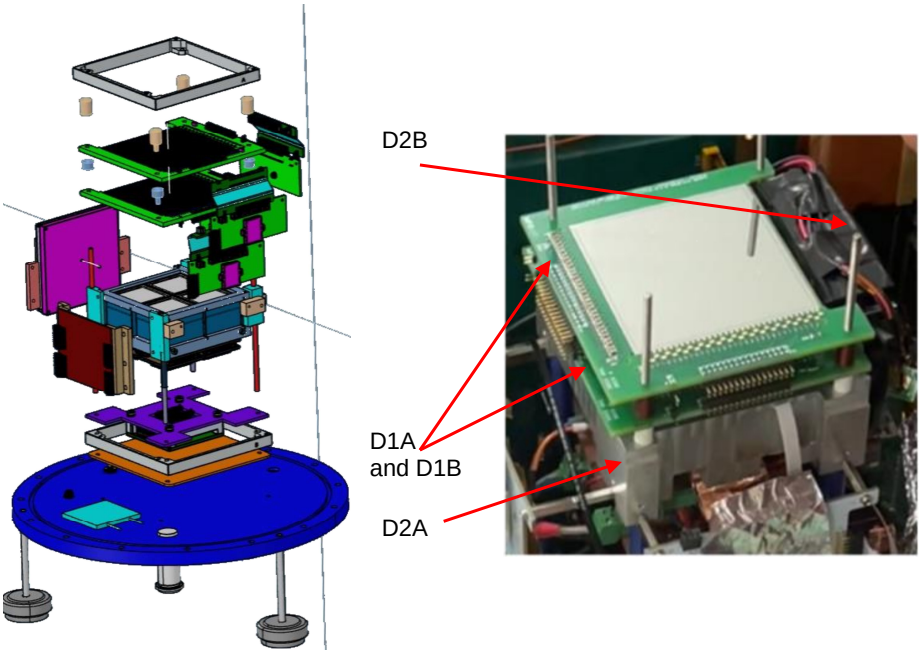


Figure 6.23: CAD drawing and final design of the COMCUBE balloon prototype without enclosure.

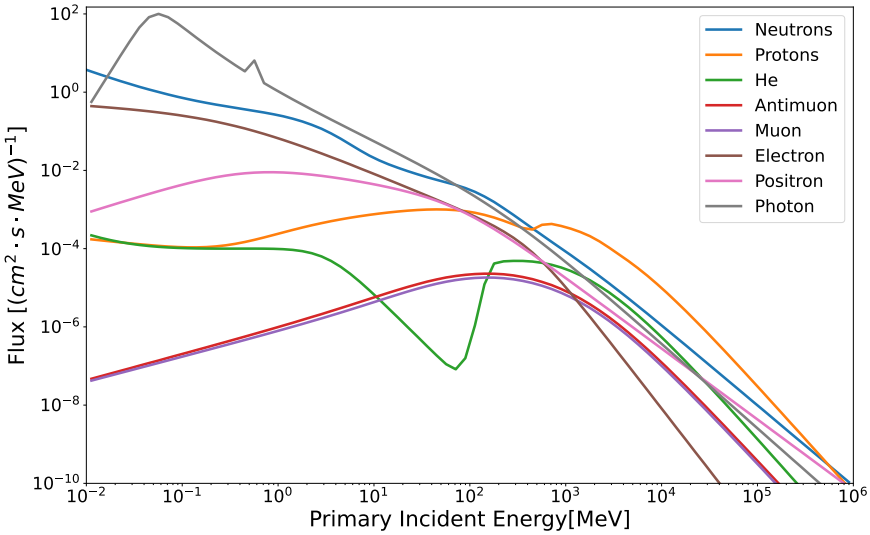


Figure 6.24: Input spectra generated with EXPACS [Sat16] for Timmins (Ontario). The shown spectra are the input for the MEGAlib background simulations.

6.3.2 The radiation environment in Earth's atmosphere

Earth's atmosphere acts as a target for high energetic cosmic particles. The by-products of these interactions consist mainly of photons, leptons, neutrons and protons. They constitute the background environment of the balloon prototype. Even more so than in space, background dominates the generated data rate, because any cosmic source flux is dampened even further by the integrated depth of the atmosphere on top of the detector. The test flight was scheduled to take place in Timmins (Ontario), situated at a geomagnetic cut-off of 1.1 GV. The Excel-based Program for calculating Atmospheric Cosmic-ray Spectrum (EXPACS) [Sat16] is the basis for the used input spectra for the background simulations with MEGAlib. The model predicts the flux of secondary particles for a given location and altitude based on a fit of a gridded set of Monte-Carlo-simulation. It is claimed to be valid for almost all locations and altitudes up to 62 km. Figure 6.24 shows the input spectra for various particles at 33 km altitude. The dominant component are secondary photons, followed closely by neutrons. Neutrons and protons can generate events by either scattering in the sensitive material, mimicking hit patterns of photons, or by being captured in the detector material inducing radioactive decays. Since the prototype is not equipped with an active veto against charged particles, all particles can generate events in the detector.

6.3.3 Trigger Rate Estimates

The trigger rate as function of altitude is derived from the simulated event and trigger rates at the respective altitude. The inspected altitudes range from 20 km to

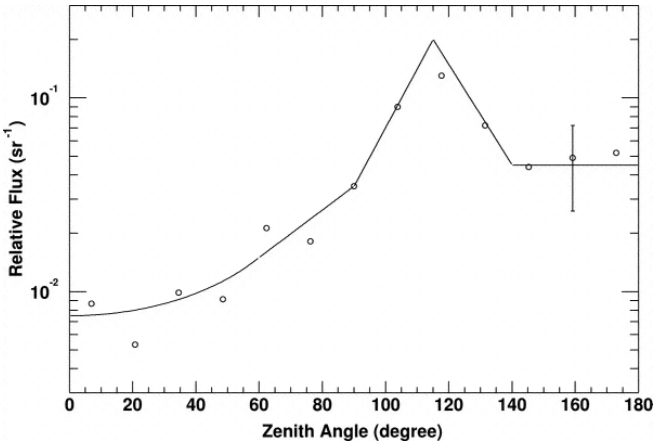


Figure 6.25: Modelled zenith dependence of the secondary photon component. Points represent measurements from a collection of balloon flights, the line represents the used approximation. From [Miz+04].

Table 6.3: Estimated trigger rates in single trigger mode.

Detector	Rate [Hz] @33 km	Rate [Hz] @20 km
Bottom Calorimeter	338	709
Side Calorimeter	252	549
DSSDs	124	207
Total Triggers	714	1465

33 km. For each step, a set of input spectra for photons, electrons, positrons, neutrons, protons, He-cores and Muons is generated with EXPACS. The zenith angle dependency of the photon component is modelled analogous to Mizuno et al. [Miz+04] as shown in Figure 6.25. The remaining components are assumed to be evenly distributed over all incoming directions.
The trigger rates for each component are derived from simulations with 1000 s exposure time and evaluated per detector. Table 6.3 shows the estimated trigger rates at 20 km and 33 km altitude. At 20 km the detector is very close to the muon creation zone at roughly 15 km altitude. This results in very high trigger rates compared to higher altitudes. The main driver of the data rate is the bottom calorimeter, as it is also containing the majority of the detector mass, shortly followed by the thinner side calorimeter. The DSSDs contribute the least amount of rate as their sensitive mass is also the lowest of all detector systems. In total one can expect a roughly 40% decrease in trigger rate at 33 km compared to 20 km.
The majority of the flight was projected to be at roughly 33 km altitude, therefore the coincidence triggers that would tag reconstructible data were only inspected for this altitude.

Table 6.4: Trigger and event rates for coincidence trigger at 33 km.

Trigger	Rate [Hz]
Side-Bottom	9
Side-DSSD	265
DSSD-Bottom	57
DSSD-DSSD	64
Total Coincidence Trigger Rate	395
Total Coincidence Event Rate	337
Total Single Trigger Rates	714
Total Single Trigger Event Rate	633

The following conditions were logged:

- Side calorimeter with bottom calorimeter
- Side calorimeter with any DSSD
- Any DSSD with bottom calorimeter
- DSSD with DSSD

The resulting trigger rates are shown in Table 6.4. The trigger rate drops significantly from 714 Hz to 395 Hz with a similar drop in to be formed events from 633 Hz to 337 Hz. If the data rate would prove to be too high, only writing out the coincidence events would be valid fall-back.
The expected trigger rate per detector as function of altitude is shown in Figure 6.26. The rate declines steeply from 20 km to around 30 km and starts flatten out at 33 km. For slightly higher altitudes one can therefore use the last derived values for a conservative estimate.

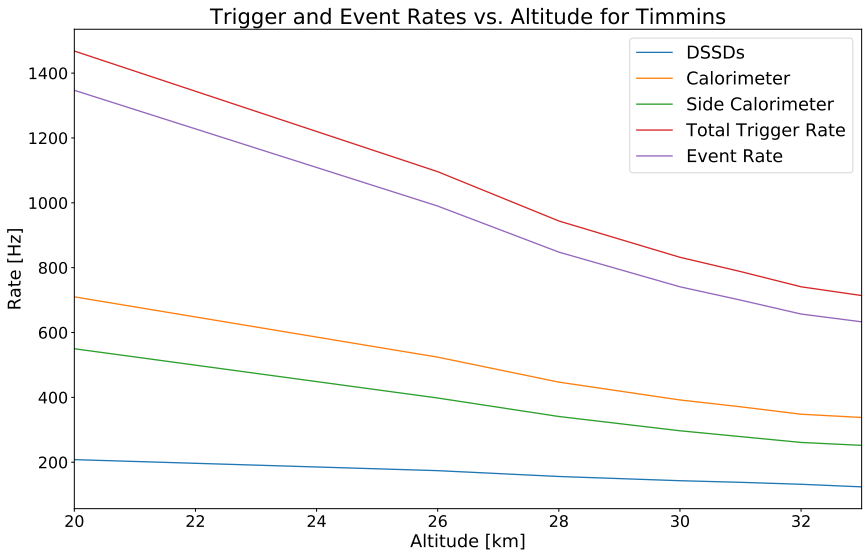


Figure 6.26: Trigger rate of the individual detector systems as function of altitude.

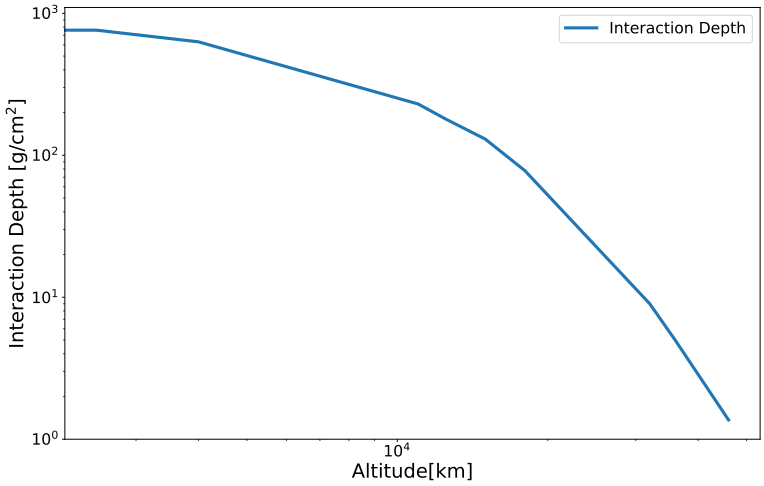


Figure 6.27: Interaction depth as function of altitude. Input parameters are taken from [Nat92].

6.3.4 Detection Sensitivity Estimates for the Crab Nebula

The input spectra generated for the trigger rate estimates can also be used to determine a sensitivity estimate for the Crab nebula. The atmosphere acts as an absorber in this setup. The amount of photons lost during their passage from the top of the atmosphere to the detector can be estimated by the *interaction depth*. It is calculated by integrating the encountered atmospheric density along the path of the incident particle:

$$X = \int \rho dr = \frac{P}{g} \tag{6.13}$$

The interaction depth at a certain altitude can therefore be calculated by the quotient of the pressure at the given altitude and its gravitational acceleration. Figure 6.27 shows the interaction depth obtained using the U.S. Standard Atmosphere [Nat92]. Earth's atmosphere is becoming more transparent the higher the energy of the incident photon. In order to model this characteristic, the attenuation coefficient for dry standard air were obtained with XCOM [Ber+09] for the composition shown in Table 6.5.
The attenuation coefficient a is a measure for the penetrability of a given material. It is the reciprocal to the attenuation length. Figure 6.28 shows its evolution as function of incident energy at sea-level. The component up to 10 keV is heavily suppressed. Photons with energy above several 100 MeV have higher but still low probability to reach the ground.
The altitude corrected spectrum is obtained by applying a correction factor

$$Corr = exp(-a(E) \cdot X(h)) \tag{6.14}$$

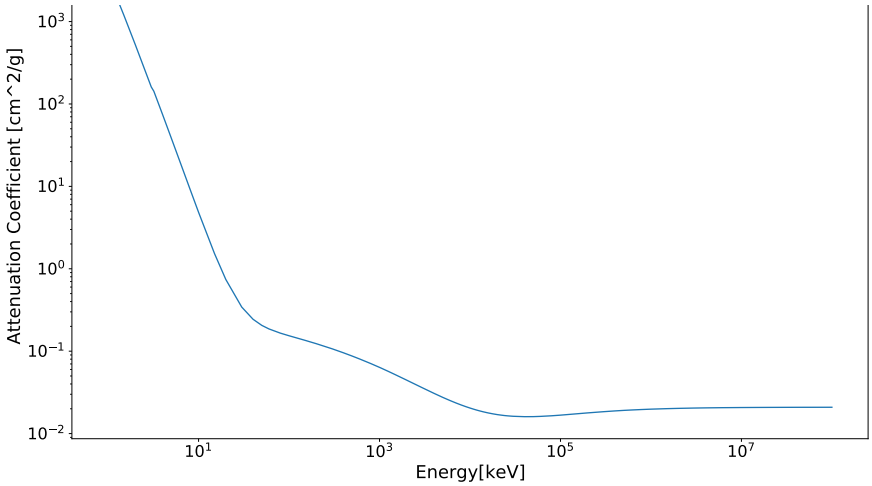


Figure 6.28: Attenuation coefficient as function of incident energy at sea-level. From: [Ber10].

with the resulting flux of

$$F_{corr}(E,h) = Corr(E,h) \cdot F_{input}(E) \tag{6.15}$$

The input flux was taken as before from Jourdain and Roque [JR09]:

$$F_{input} = 3.87 \cdot E^{-(1.79+0.134 \cdot ln(E/E_0))} \frac{1}{keV \cdot cm^2 \cdot s} \tag{6.16}$$

The original spectrum and the attenuated spectrum for 20 km and 33 km are shown in Figure 6.29. The spectrum assumes an on-axis source position. In both cases, the shape and the amplitude of the spectrum are affected by Earth's atmosphere. At 20 km the effect on the spectrum is most drastic, cutting of photons with energy of less than 20 keV almost completely. The spectrum then peaks at around 80 keV and then drops of over the modelled range. At 33 km the shape is closer to the original input spectrum. The effect is mostly a drop-off in flux in the order of 50% at lower energies decreasing to below 20% at the fluxes in the MeV region.

Table 6.5: Composition of dry air used to derive the attenuation coefficients. From [Pic+08].

Component	relative abundance
N	0.78
O_2	0.209
Ar	$9{\cdot}10^{-3}$
CO_2	$4{\cdot}10^{-4}$
Ne	$1.82{\cdot}10^{-5}$

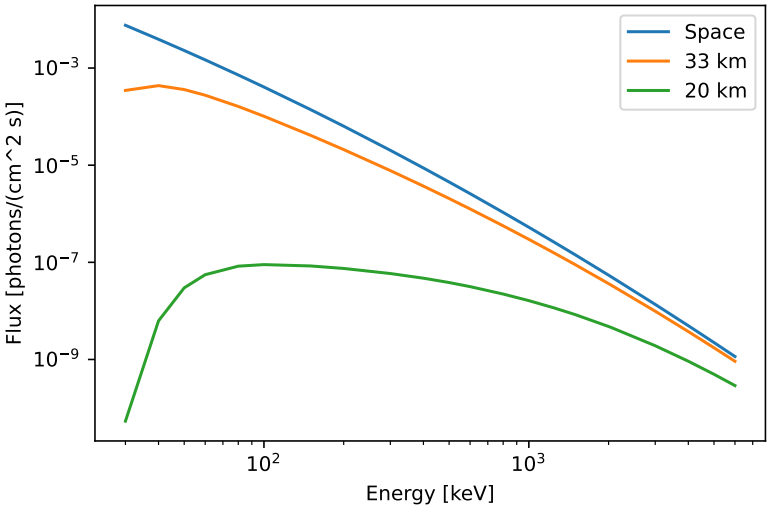


Figure 6.29: Flux vs. Energy for 20 km (green) and 33 km (orange) altitude derived from the spectral model derived from [JR09] using the altitude and energy dependent correction described above.

Figure 6.30 shows the expected background rate as function of measured energy for single-hit events at 33 km. The simulation contained three activating components (neutron, proton, helium), where the neutron component is the most dominant. It shows several activation lines at energies below 1 MeV, neither of which is considerably present in the sum spectrum. The photon as well as the positron spectrum show either a bulge or a line at 511 keV, which is also visible in the sum. Muon, anti-muon and helium play only a minor role, while the photonic and neutron components are dominant.
Figure 6.31 shows the background rate for multi-hit events, which reconstructed as Compton events. Only for the photon component and events induced by deexitation photons of activating components, the reconstruction algorithm tagged genuine Compton events. The summed spectrum rises with increasing instrument efficiency and shows with the exception of the 511 keV line no significant line feature. At around 800 keV the spectrum shows a significant bump. This is due to an increase in (wrongfully) reconstructed charged particle events, that the reconstruction algorithm reconstructed as Compton event. The presence of this bump suggests to only use Compton events up to 800 keV. The spectrum also illustrates a threshold of at least 80 keV for an event to be reconstructed as a Compton event. Since the efficiency at this energy is known to be very low, the lower bound of the region of interest was set to 100 keV. The Compton-spectrum follows similar trends as the single spectrum. Up to 800 keV photons and neutrons are the main component, with the charged particles, especially electrons, positrons and protons taking over.

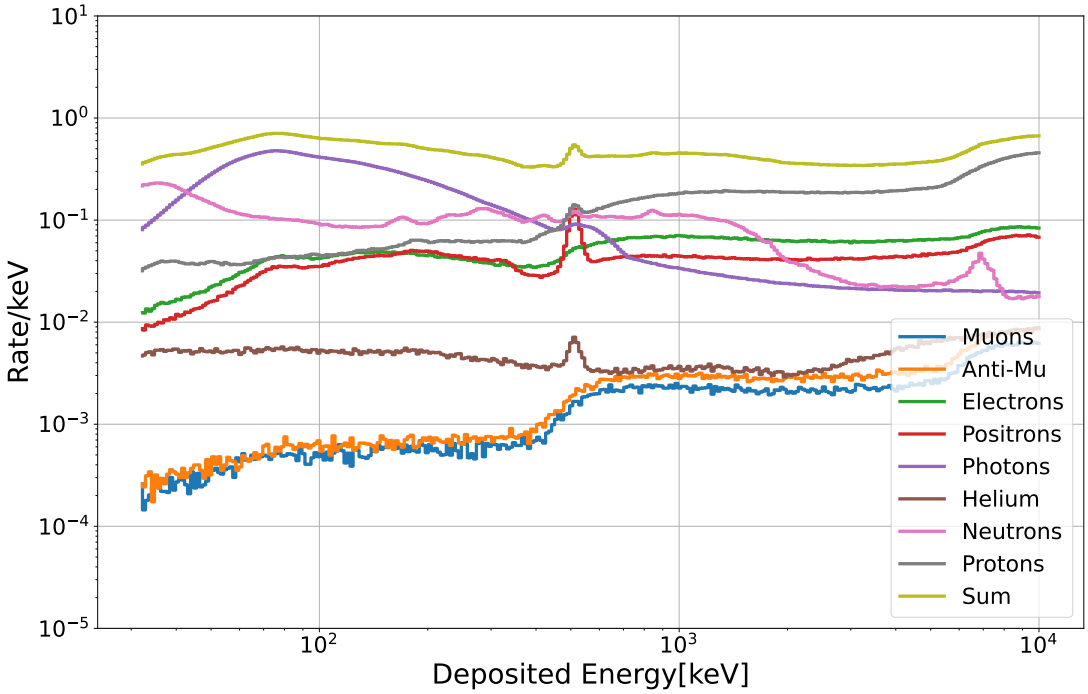


Figure 6.30: Rate as function of measured energy for single hit events.

The source simulation were performed in three steps. First, the effective area for the source was determined by running a simulation with 10 million pre-triggered events for an altitude of 33 km. For the analysis, only Compton events between 100 and 800 keV were considered, due to the reasons stated above. The directional element was set as one FWHM of the ARM and was determined to be 41°. The effective area was calculated as

$$A_{eff} = A_{init} \cdot \frac{N_{measured}}{N_{emitted}} = 0.47\,\mathrm{cm}^2 \tag{6.17}$$

with $A_{init} = 257\,\mathrm{cm}^2$ as starting area and $N_{emitted} = 22\,511\,136$. As second step, the background rates were estimated using the same angular and energy selection as for the source using its location, in this case the zenith. The background rate under the source was determined to 3.2 Hz. From the obtained effective area and background rates, the sensitivity was obtained using [Zog06]

$$S_{min} = \frac{(n^2 + n) \cdot \sqrt{n^2 + b \cdot t_{obs}}}{2 \cdot A_{eff} \cdot t_{obs}} \tag{6.18}$$

where n is the significance level, usually 3, b is the background rate and t_{obs} is the integration time of the measurement. Figure6.32 shows the simulated Crab measurement as well as the ARM-distribution after energy selection.

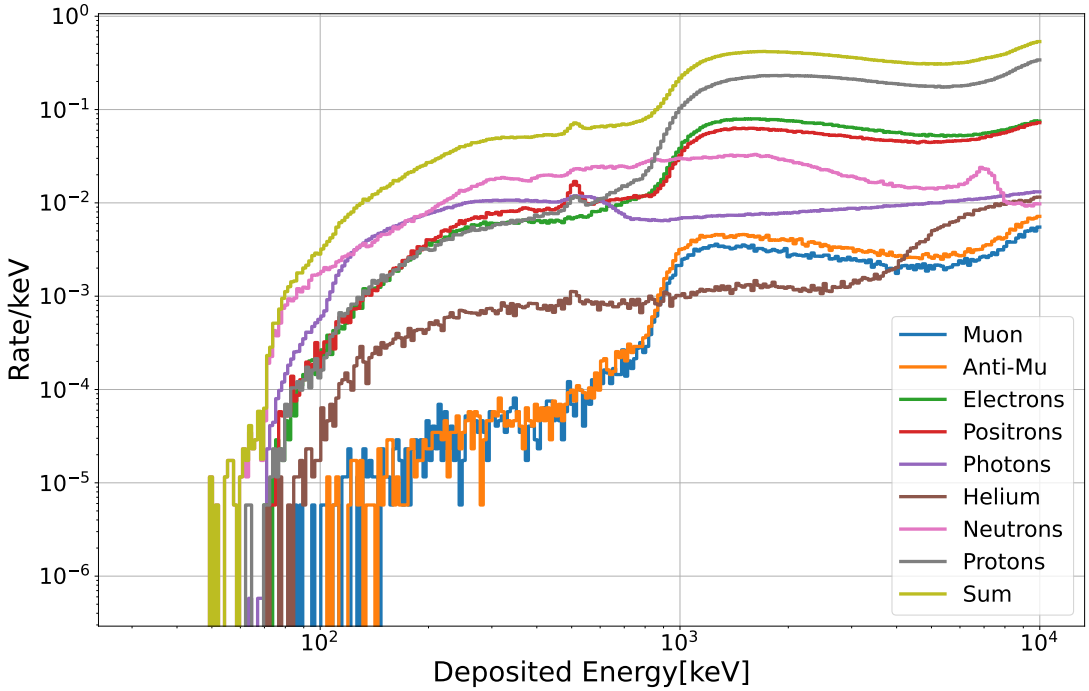


Figure 6.31: Rate as function of measured energy for multi hit events reconstructed as Compton events.

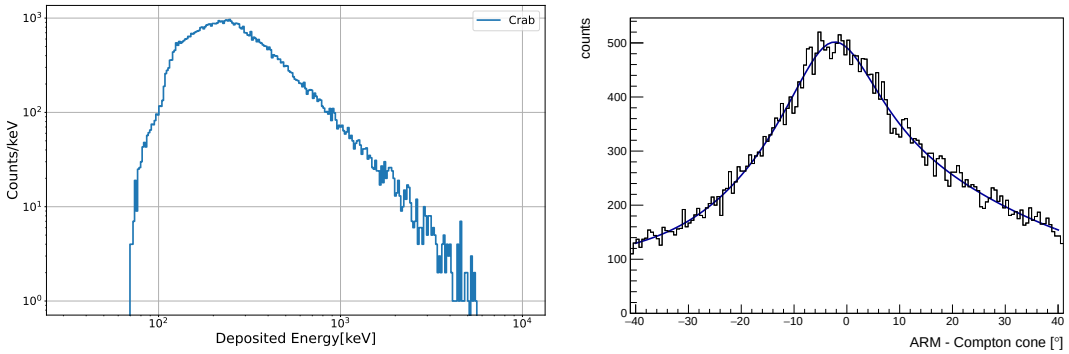


Figure 6.32: Left: Counts as function of measured energy for multi hit events reconstructed as Compton events for an on-axis Crab source. Right: ARM distribution after energy selection between 100 and 800 keV.

Equation 6.18 allows to plot the sensitivity as function of integration time t_{obs} in hours. The resulting curve is shown in Figure 6.33. The sensitivity is improving with the square root of the observation time, but only crosses the estimated source flux after 100 hours of integration time, making a significant detection of the Crab over the expected flight duration very unlikely.

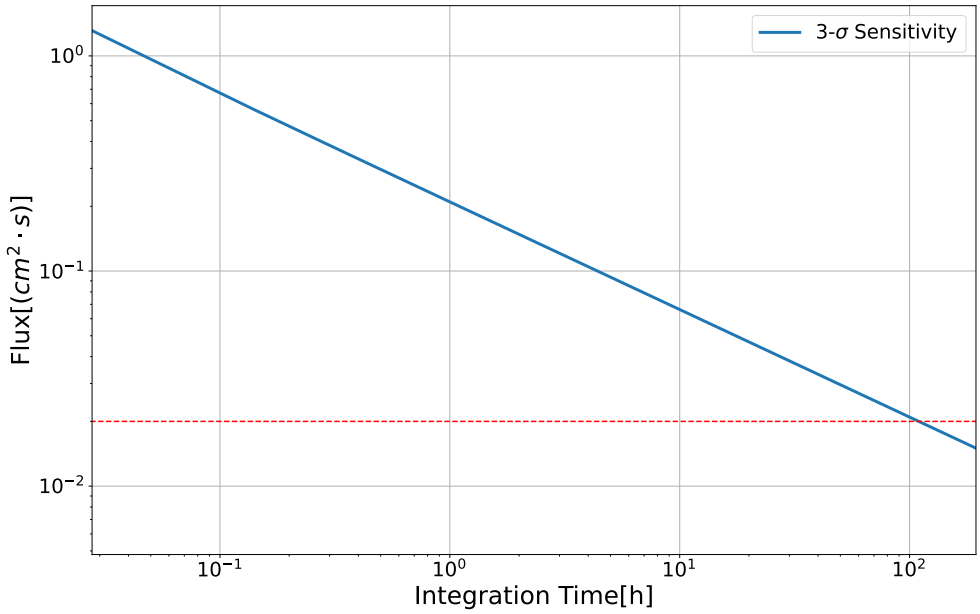


Figure 6.33: Sensitivity as function of integration time. Given is the 3-σ sensitivity. The dashed line corresponds to the flux of the simulated Crab source.

6.3.5 Balloon Flight and Measured Rates

The prototype flight was launched on 27 August 2023 at 07:12 UTC as the final flight of the Strato Flight 2023 campaign carried out by the Centre National d'Études Spatiales (CNES) and the Canadian Space Agency. The gondola housed three experiments. SPECtromètre Infrarouge à lasErs in Situ (SPECIES) [Cat+23] is an infrared spectrometer for the simultanious detection of a variety of atmospheric trace gases. The second large experiment was the Canadian Atmospheric Laser Absorption Spectrometer Experiment Test-bed (CALASET) [Sau+20] developed by the universities of Toronto and Alberta. The infrared laser spectrometer is designed to detect CO_2 and NO_2 in the stratosphere. In addition to the main instrument CALASET also housed several small student run experiments. COMCUBE was the third payload on the CARMENCITA gondola provided by CNES. In addition to the scientific payload, the flight was also used by CNES as test-flight for their new Iridium communication module L-SAT and a retractable solar power array called MEDOR. A sketch of the gondola and the placement of the experiments is depicted on the left of Figure 6.34. The picture on the right shows the gondola during system testing with extended solar panel.

The flight started from RW28 of the Victor M. Power airport close to the city of Timmins designated as NIMBUS-4. The wind conditions on the ground and the lower atmosphere are critical for the start. In order to broaden the range of acceptable launching conditions, CNES employs a two-balloon setup for their flights. The gondola is attached to a smaller auxiliary balloon that is connected to the main balloon

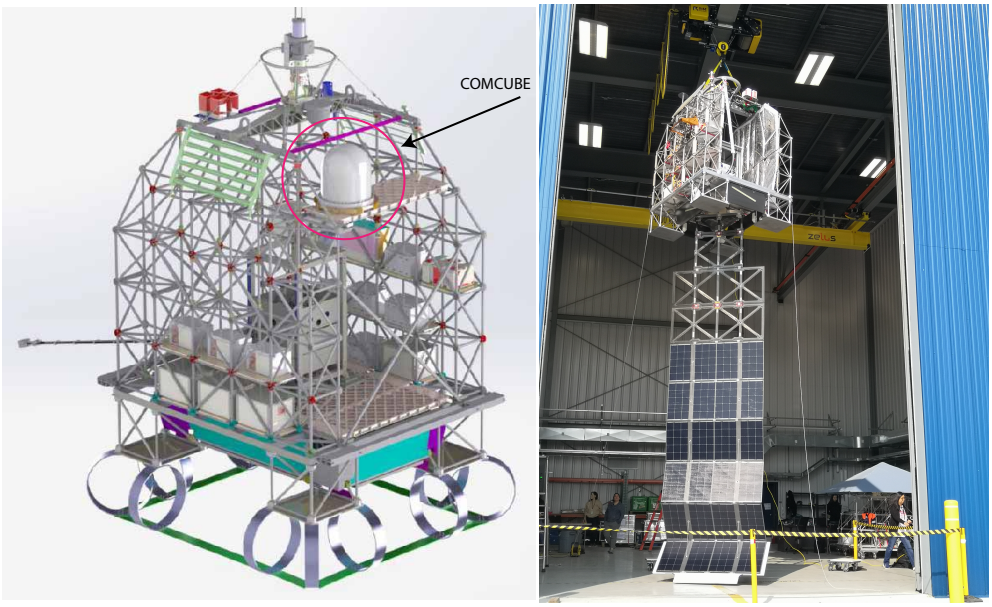


Figure 6.34: Left: Sketch of the CARMENCITA gondola selected for the flight. The location of COMCUBE is marked by the red circle. From: [Cne]. On the right: Gondola hooked up for testing of the MEDOR system and attitude controls during pre-launch checks.

deployed further away. Prior to the launch, the gondola is hovering over the ground, allowing personnel and heavy machinery to evacuate the area. Only then the main balloon is released off the ground and the gondola is launched. Pictures of the gondola on the tarmac and during launch are shown in Figure 6.37.
All flight related data (positions, altitude, timing, data rates) was derived from the house-keeping database running alongside the experiment. Flight tracking data was obtained from the gondola systems and stored in the house keeping database. The flight route is shown in Figure 6.35. The flight route circled the area around Timmins. The gondola landed in a deserted area roughly 30 km west-north-west of the city of Timmins after 14.5 h at 21:42 UTC.
The flight profile is shown in Figure 6.36. After the initial climb phase the gondola reached its designated operation altitude of 36 km. In order to help mitigate thermal issues in the SPECIES experiment, the altitude was later reduced to 32 km. Data tracking was stopped at the top of descent in compliance to CNES regulations. The flight ended with no incidents.
Prior to the launch several issues with the detector-hardware emerged. The main issue was rooted in the necessity to replace the intended power supply unit (PSU). The intended design was not ready for the beginning of the balloon campaign. The replacement design, essentially the prototype of the intended design, proved adequate to power the setup. However, its lack of features like a sufficient amount of filtering capacities made the scientific operation of the DSSDs (D1A and D1B) impossible.

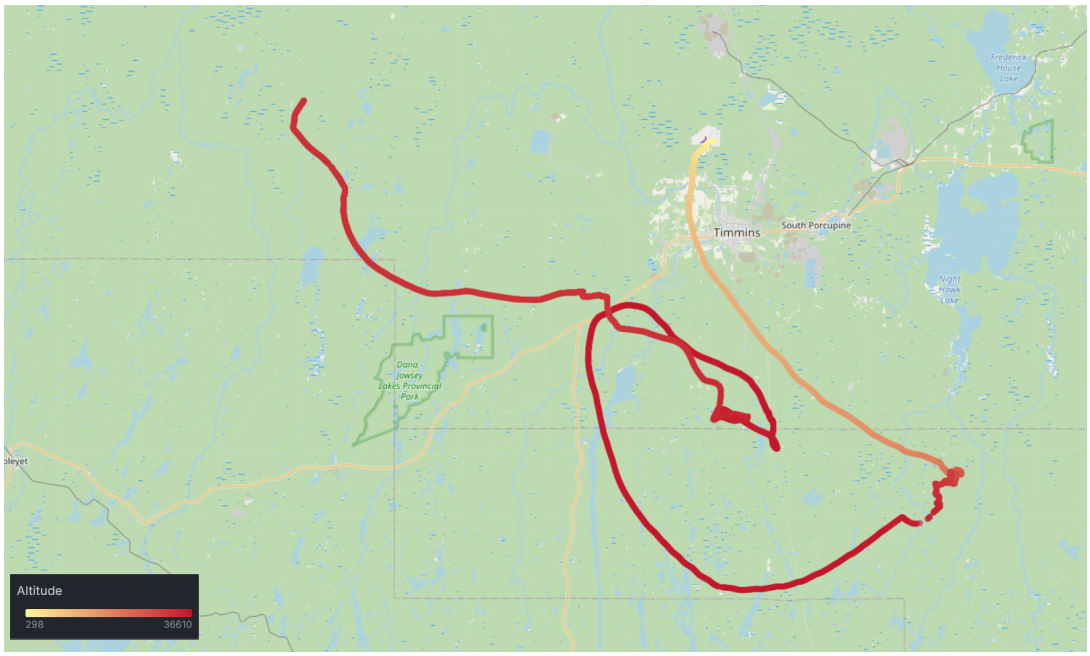


Figure 6.35: Flight route of the balloon gondola. Data taking stopped at descent. Given are the positions taken from the GPS tracker of the gondola and the altitude values as color code.

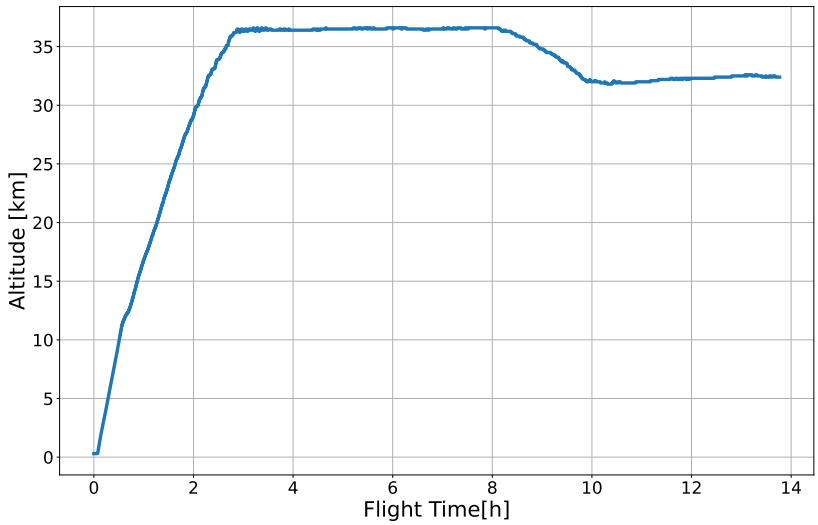
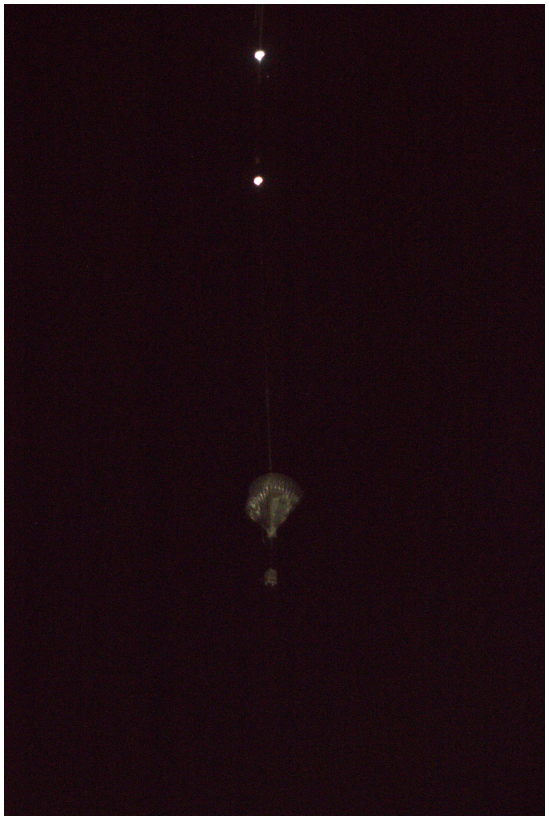


Figure 6.36: Altitude of the gondola as function of flight time. Note that data taking had to stop at the top of descent of the second float phase.

(a) Filling of the ballons

(b) Gondola floating prior launch

(c) Gondola is pulled by main balloon



(d) Gondola after successful launch

Figure 6.37: Launch of the gondola. The two balloons are filled prior to connecting the payload to the smaller auxiliary balloon (a). The connected gondola is able to hover over the ground after disconnecting the harnesses (b). The main balloon is released pulling the gondola upwards (c). The main balloon transports the gondola to its designated altitude. For air-safety at lower altitudes populated by conventional aircrafts, the flight is equipped with strobes signalling its position (d).

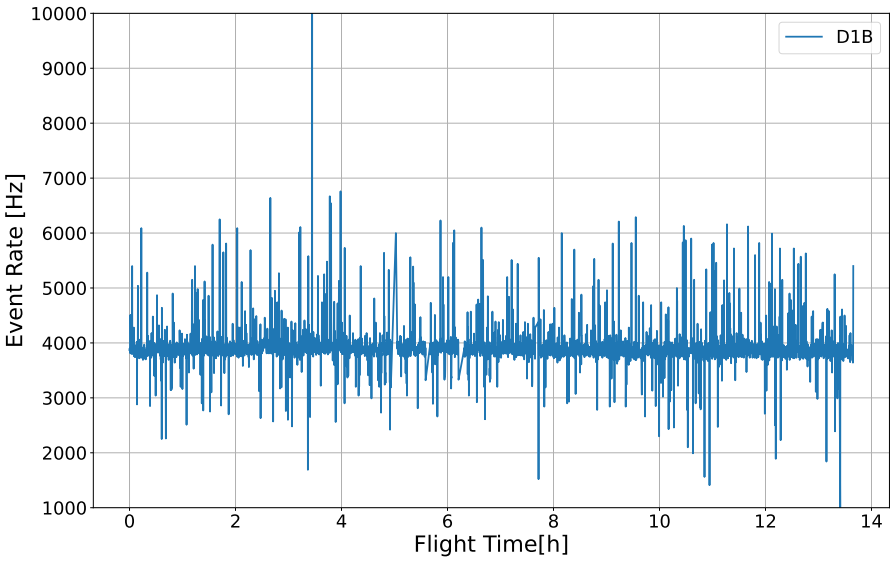


Figure 6.38: D1B trigger rate as function of flight time.

Their front-end electronic proved to be too sensitive to the switching noise of the PSU. Figure 6.38 shows the data rate of the IJC-lab developed DSSD (D1B) as function of time. As can be seen, the DSSD was generating a flat trigger rate of around 4 kHz, lacking any of the expected altitude dependent features.
The side calorimeter (D2B) developed noise issues over the duration of the four week campaign, resulting in a shift of the trigger level from the intended 15 keV to a value larger than 80 keV. Attempts to determine the exact trigger level after the return of the gondola turned out to be futile since the optical grease between quartz-window and SiPM-array deteriorated due to air-bubbles. This makes an increase in the trigger level assumed in the simulations necessary.
The bottom calorimeter (D2A) showed cross-talk behaviour, especially after measuring large energy deposits. The event data would then show additional hits in other regions of the crystal, unrelated to the actual deposit. These issues where later attributed to firmware and wiring issues.
The evolution of the trigger rates of D2A and D2B over the duration of the flight is shown in Figures 6.39a and 6.39b respectively. Both rate curves show the same overall shapes with maxima and minima at similar points in time. The data rate of D2A proved to less stable than for D2B, which was later isolated as a firmware issue leading to a bursty write-out via the Direct Memory Access (DMA). It also led to two significant spikes during the first hour and close to the eighth hour of the flight with 1.5 and 4.4 kHz of data rate respectively. The overall sustained maximum in both rate curves corresponds to the crossing of the muon-creation zone. The rate stabilized at a lower value once the gondola reached its operation altitude.

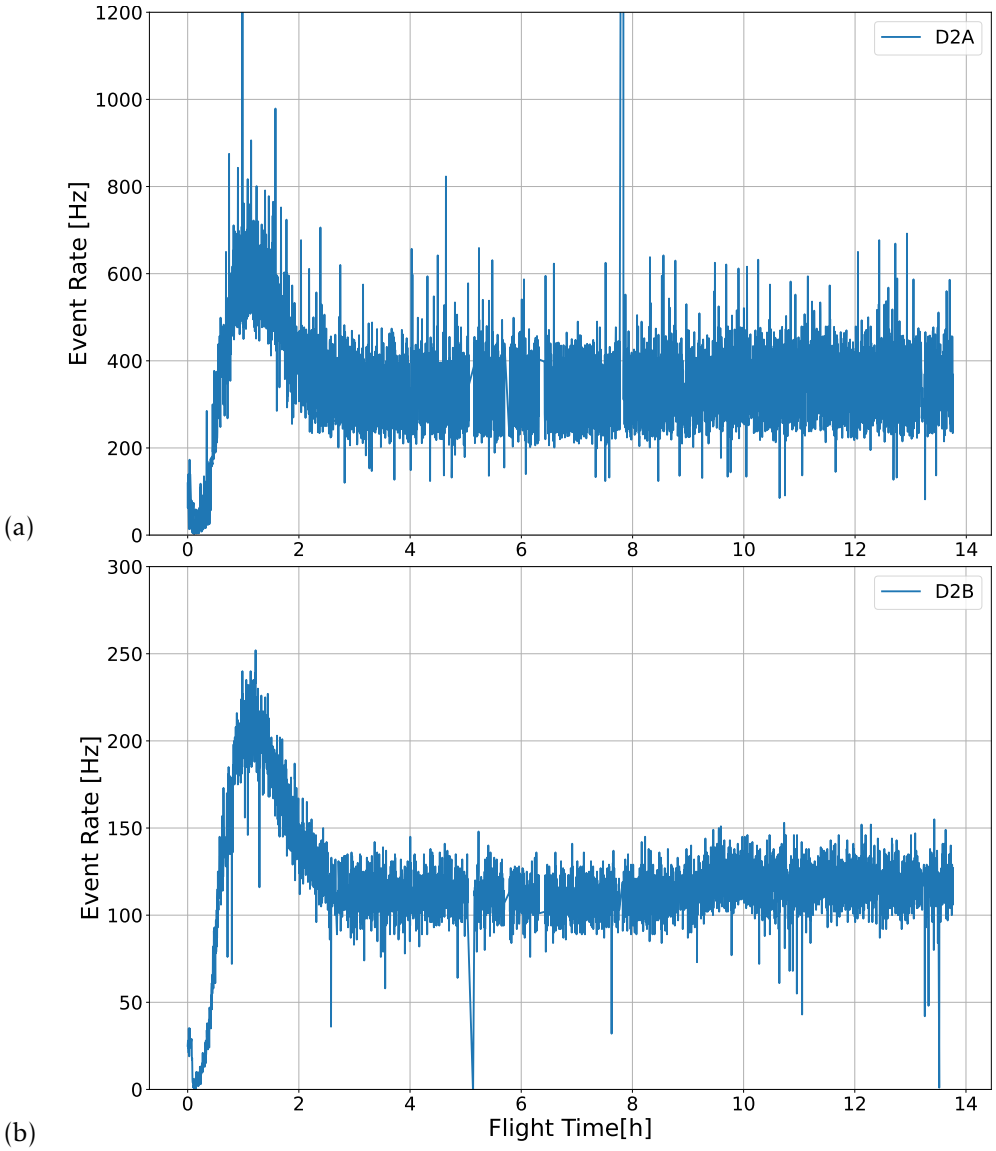


Figure 6.39: Trigger rate vs flight time of D2A and D2B.
(a) D2A trigger rate as function of flight time. The bursty output stream resulted in a large spread in data rate.
(b) D2B trigger rate as function of flight time. The output rate is more stable, but approximately a factor 3 lower compared to D2A.

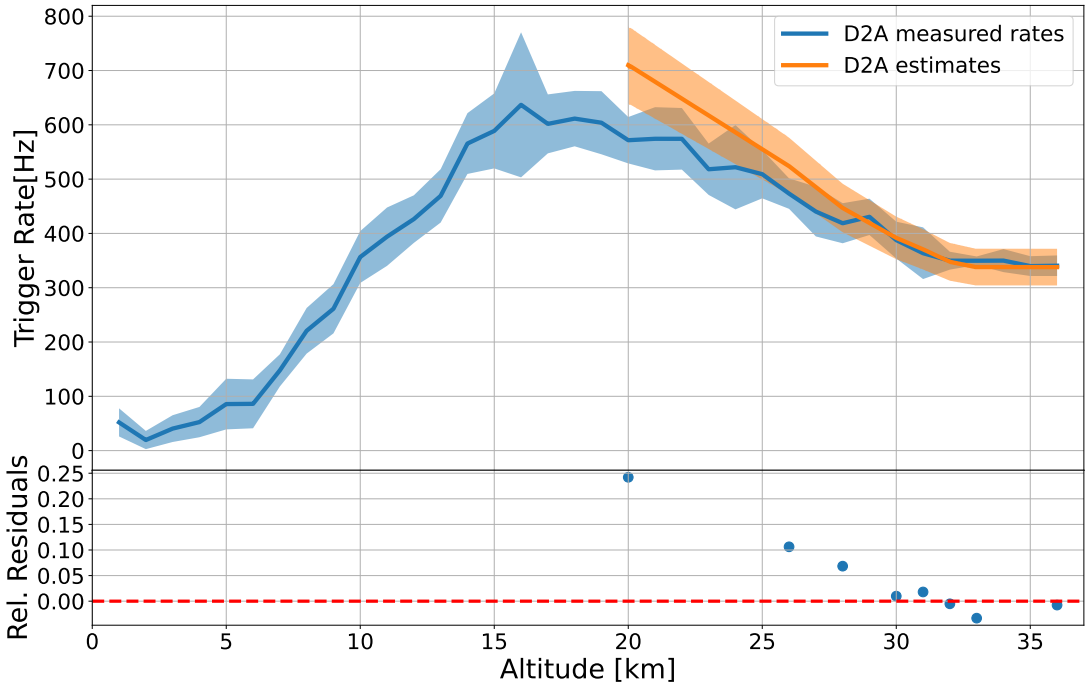


Figure 6.40: Trigger rate in D2A as function of altitude and their simulated estimates for altitudes larger than 20 km. The bands around the measured rates give the 3 σ error, the bands around the estimates correspond to the minimum 10% uncertainty inherent from uncertainties from simulation input and Geant4 physics modelling.

In order to compare the estimation shown in Figure 6.26 the measured rates were treated as follows:

- The altitude was modelled as function of time derived from the flight profile
- The measured rates were assigned altitudes according to their time stamp
- The measured rates were then binned into 1 km bins ranging from 1 km to 36 km and above.
- For each bin, the mean trigger rate and its standard error were determined.

Figure 6.40 and 6.41 show the resulting altitude depending trigger rates for D2A. The error-band around the measured mean trigger rates corresponds to three σ of the calculated standard error for each altitude bin. Here, the unstable trigger rate of D2A is resulting in a wide error-band. The simulation derived estimates for D2A are given in orange. The band around curve gives the inherent minimum uncertainty from simulation inputs and simulated physics and detector modelling of 10%. The main sources of uncertainties are the simulation inputs from the spectral model, the input energy resolutions, modelled trigger thresholds of the detectors and the quality of the simulated physics, here Geant4 version 10.2 patch 3. A comparison between simulated and measured spectra with a focus on activation can be found in [Zog+09]. The authors used the transient gamma-ray spectrometer(TGRS) [Owe+91] as bench-

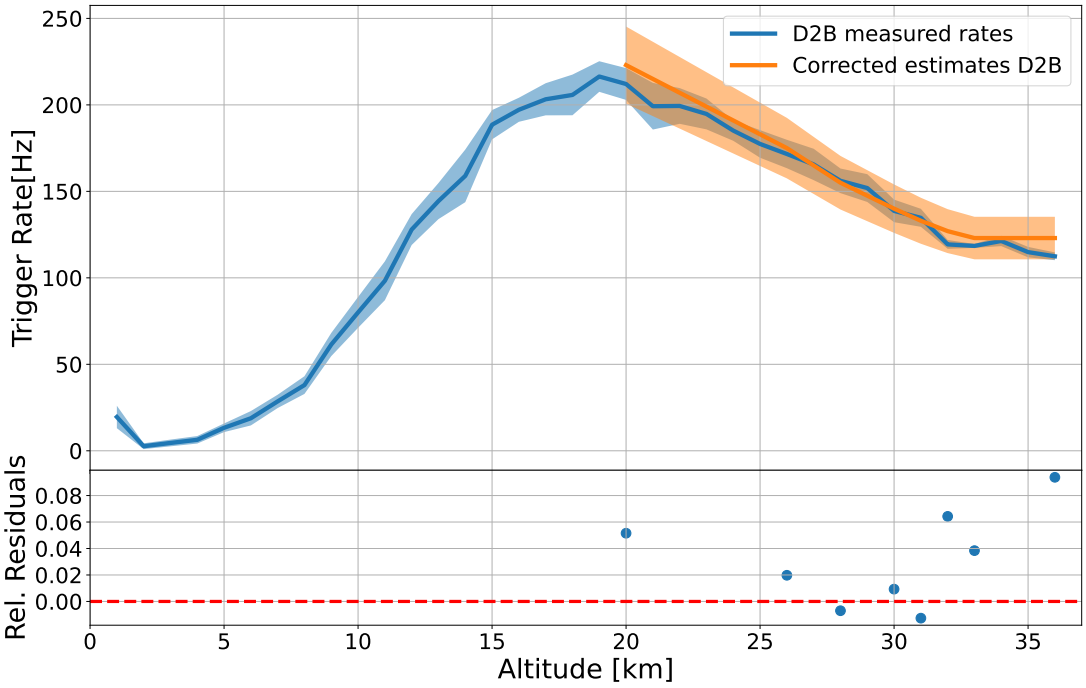


Figure 6.41: Trigger rate in D2B as function of altitude and their simulated estimates for altitudes larger than 20 km. The bands around the measured rates give the 3 σ error, the bands around the estimates correspond to the minimum 10% uncertainty inherent from uncertainties from simulation input and Geant4 physics modelling.

mark example and concluded, that the simulated spectra are in good agreement with the measured spectra. Please note that the MEGAlib version of the authors was using an earlier version of Geant4. A recent study using the COSI balloon flight in 2016 arrived to a similar degree of agreement (10-20% depending on energy) using MEGAlib with Geant4 version 11.1 [Gal+25] and EXPACS for input spectra modelling. Note that the applied phenomenological correction mainly affects the shape of the simulated spectra, not the integrated rates, that are relevant here.
Figure 6.40 shows the measure and estimated rates for D2A. The measured trigger rates of D2A are in good agreement for altitudes larger than 27 km and at operation altitude. For lower altitudes the simulations over-predict the trigger rates by more than 10 %. Since these altitudes are in general passed fairly quickly en rout to the stratosphere, the impact is negligible.
The measured trigger rates for D2B are shown in Figure 6.41. The raise in trigger threshold resulted in the need of a reanalysis of the simulation data, as the estimated rates with the original trigger threshold of 15 keV are naturally a severe overprediction of the actual trigger rate. The new trigger threshold was estimated to 160 keV. The curve obtained from the new estimates follows the measured trigger rates well. For altitudes lower than 25 km and higher than 30 km one can observe

an over-prediction of the simulated rates. As a whole, the measured rates are inside of the minimum 10% uncertainty of the simulated estimates. This makes the determined new threshold at least a very likely value, when assuming the validity of the input model.

6.3.6 Summary

The COMCUBE prototype flight was deemed a partial success. During the 14.5 h of operation, D2A and D2B were able to take data consistently, while the DSSDs were at least operating and validated the robustness of the DMA-write-out. The detector itself was operating stable with respect to thermals over the whole flight.
Sadly, the state of the detector did not permit for an analysis of the spectral data. Since the DSSDs triggered essentially on the switching noise of the PSU, it is unknown for which events the detector actually measured a coincidence between D2B and D2A or single hits. Furthermore, the coincidence between D2A and D2B is expected to generate the lowest coincidence rate of around 9 Hz. This is only around 2% of the total expected coincidence rate for particles coming from all directions. The noise issues turned out to affect D2B quite severely, resulting in an increase of trigger threshold to roughly 160 keV. The later discovered issue of the deteriorating optical grease also have a position dependent negative effect on the energy resolution and add big challenges to the position reconstruction.
The state of D2A was not ideal. The presence of cross-talk for events with especially large energy deposits would introduce a shift the slope in the measured spectrum. These events would have to be removed prior to analysis. The exact energy of the removed events is in general unknown. The absence of a custom detector effects engine modelling this particular behaviour properly, makes it very difficult to perform a meaningful analysis of the spectrum.
The sensitivity for the Crab was determined to be insufficient to detect the Crab for the duration of the flight with all four detectors working nominally. Having only two out of four subsystems available and only one working at a trigger level as assumed for the simulations reduced the probability for a source detection using Compton imaging to zero.
EXPACS proved to be a sufficient basis for deriving expected trigger rates at stratospheric balloon altitudes, albeit it shows tendencies to over predict the trigger rates at altitudes lower than 25 km. This can be either a modelling error inherited from the EXPACS derived input spectra, the quality of the simulated physics of the Geant4 version used in MEGAlib. Nonetheless, this should come at no harm for stratospheric balloon operating altitudes, which are typically higher than 30 km. Here the model derived rates show a good agreement with the measured rates during the flight. Potential caveats are the different shape and dimension of the enclosure and the different final design of the detector assembly. These changes added more passive material, that would have had a shielding effect but could also drive the rates up because of potentially more short term activation.

Chapter 7

Summary, Outlook and Future Applications

Over the course of the thesis, machine learning turned from being a novelty to a standard approach for tackling numerous problems in all parts of physics. This thesis tried to show the applicability of especially deep learning models in the field of gamma-ray astronomy for event classification. The goal here was not just to provide a working model, but also to design it in such a way, that it can be used on the satellite, processing the event data directly at the detector, implementing what is called *edge computing*. The model presented demonstrates an improvement over freely available algorithms or classical machine learning algorithms. It leaves nonetheless a bit of achievable performance on the table due to the restrictions that needed to be applied. The applicability of approaches like the presented one hinges on the development of sufficiently powerful and radiation hardened embedded systems running these models. Future development will allow to revert some of the design decisions and potentially increase the performance even more. Even though e-ASTROGAM was rejected in the M5 proposal, the approach taken is viable for almost any gamma-ray detector with x-y-z resolution power. A future application would be the development of a similar classification model for the upcoming COSI detector. Here the task would be to classify into Compton events and everything else, a significant reduction of complexity compared to the case presented here. Other applications warranting further study are the application of deep learning models for Compton sequence reconstruction. Here a mixture of so-called graph-based neural networks with physics informed loss-functions seem a promising route.
The application of machine learning of course does not stop there. Other applications tried in the local research group with direct or indirect participation are anomaly detection for a visual wire inspection setup [Dei+25; Mit+25], unfolding of PMT-pulses [Bec21] or S1-S2 classification for the Xenon1T experiment [Mer24]. In all three cases, the application of machine learning turned out to be either an improvement over classical methods or in the case of the visual wire inspection very helpful as it allows to automatize the visual inspection with ample accuracy.

This shows, that machine learning is becoming an integral approach for tackling the physics problem of the future.
The second part of the thesis covered the design and testing process of the COMCUBE-proposal. Gamma-ray astronomy is lacking sensitivity especially in the low to medium MeV range. Since the Compton Gamma Ray Observatory, no mission covered the range between 1 to 30 MeV with sufficient sensitivity, creating the so-called MeV-gap. Over the course of the thesis several proposals with personal and group involvement for a larger scale MeV to GeV mission were rejected (e.g. e-ASTROGAM, F-ASTROGAM, M7-ASTROGAM, newASTROGAM). In order to conserve knowledge about how to setup and operate Compton Telescopes, small scale CubeSat missions with relatively low cost and short time-scales may prove to be vital. These small-size satellites have shown their capabilities to perform scientific measurements. In the context of γ-ray astronomy, their main application is in the context of GRB monitoring and observation of low energy steady sources [Blo+22] with mass restrictions preventing their usability for higher energies. COMCUBE is an example for a GRB polarimeter, observing photons up to one MeV. The work in this thesis was instrumental in determining the final design of the 4U-detector design, which prevailed into the Phase-A study of COMCUBE-S[1]. The 4U design shows promising performance characteristics and, given a sufficient number of swarms, would allow for a complete coverage of the whole sky at all times for the duration of the mission. The simulated polarization response of the design is sufficient to reach the performance of larger scale single instruments like POLAR, given a sufficient number of satellites. Of course, these estimates are limited by the simulation's input parameters. The individual detectors must be carefully modelled, input spectra and the modelling of the physical processes need to be accurate to derive representative performance benchmarks.
The prototype flight in Timmins was plagued by teething issues and technical difficulties, albeit it was declared a partial success for the preparation of the following transatlantic flight a year later. EXPACS delivered input spectra that resulted in trigger rate estimates in agreement to the measured trigger rates close to the operation altitude of the flight. This allowed for an accurate planing of the on-board storage and write-out bandwidth for nominal working subdetectors. The weakness of every simulation-driven approach is its dependency on accurate input parameters and correctly modelled physics processes. This is the largest source of systematic errors in the presented results. One way of testing the input parameters is the comparison of simulation data with actual measured data. Sadly, an in-depth spectral analysis of the gathered data was feasible due to the multitude of technical issues before and during the flight.
Using CubeSats as either technology test-beds for science satellites or full replacements for larger detector is a promising idea worth investigating further. NASA together with the Karlsruhe Institute of Technology (KIT) is developing a novel silicon

[1] Note that the proposed COMCUBE-S detector design contains an additional BGO-calorimeter to increase the measurement range to 10 MeV.

pixel detector for gamma-ray astronomy [Str+24] to replace DSSDs as tracking detector. Their main advantage are easier scalability in size (the size of the layer scales as multiples of the chip size, rather than the DSSD size), better yields (the chips size is in the order of cm compared to 10's of cm) and with that a lower price. The technology is still in its infancy with a lot of research and development necessary to meet requirements for, e.g., power draw and energy resolution, but looks promising nonetheless. Combining silicon pixels with more modern, faster solid state scintillators or fast liquid noble gas TPCs could potentially lead to the next medium or large scale MeV mission so desperately needed. Balloon flights and CubeSat missions will be an integral part to demonstrate the feasibility of new concepts and to qualify solutions applicable to larger detectors.
This chapter concludes with a short and incomplete list of lessons learned over the course of the thesis:

1. Lead times of companies are not worth the paper they are printed on, especially when concerning mission critical equipment.

2. Any machine learning project hinges on the quality of training data and its coverage of the underlying data space. Any left over bias can and probably will harm the generalization performance of the model.

3. It is in general beneficial to design deep learning models as complex as necessary, not as complex as possible. Tasks in physics tend to be on the less complex range of the spectrum compared to applications encountered in other fields.

4. Most of the time needed to develop a model is spend designing the data pipeline and generate the data sets not the actual development of the model. A flexible data pipeline reduces the need of time-consuming re-implementations.

5. Any simulation is only as good as its input parameters and simulated physics. It is almost never a waste of time to sanity check the output at least once.

Appendix Nutzung von KI-Werkzeugen

Bei der Erstellung dieser Arbeit sind keine explizit auf KI-basierten Anwendungen verwendet worden. Zur Literatursuche und Erarbeitung von Lösungsansätzen für konkrete Fragestellungen wurden jedoch indirekt KI-Optimierungen der gängigen Suchmaschinen verwendet.

Bibliography

[MTF15] F. Massaro, D. J. Thompson, and E. C. Ferrara. "The extragalactic gamma-ray sky in the Fermi era". In: *The Astronomy and Astrophysics Review* 24, 2 (2015), p. 2. DOI: 10.1007/s00159-015-0090-6. arXiv: 1510.07660 [astro-ph.HE].

[Con71] R. G. Conway. "Radio Polarization of the Crab Nebula". In: *The Crab Nebula*. Ed. by R. D. Davies and F. Graham-Smith. Vol. 46. IAU Symposium. 1971, p. 292.

[Die21] R. Diehl. "Radioactive isotopes in the interstellar medium". In: *Astrophysics and Space Science* 366.11 (2021), p. 104. ISSN: 1572-946X. DOI: 10.1007/s10509-021-04003-8.

[Wie71] R. Wielebinski. "The Effelsberg 100-m radio telescope". In: *Naturwissenschaften* 58.3 (1971), pp. 109–116. ISSN: 1432-1904. DOI: 10.1007/BF00593099.

[Rei+18] A. Reinacher et al. "The SOFIA Telescope in Full Operation". In: *Journal of Astronomical Instrumentation* 07.04 (2018), p. 1840007. DOI: 10.1142/S225117171840007X.

[Neu+84] G. Neugebauer et al. "The Infrared Astronomical Satellite (IRAS) Mission." In: *The Astrophysical Journal Letters* 278 (1984), pp. L1–L6. DOI: 10.1086/184209.

[Gar+23] J. P. Gardner et al. "The James Webb Space Telescope Mission". In: *Publications of the Astronomical Society of the Pacific* 135.1048, 068001 (June 2023), p. 068001. DOI: 10.1088/1538-3873/acd1b5. arXiv: 2304.04869 [astro-ph.IM].

[Sut+15] W. Sutherland et al. "The Visible and Infrared Survey Telescope for Astronomy (VISTA): Design, technical overview, and performance". In: *Astronomy and Astrophysics* 575, A25 (2015), A25. DOI: 10.1051/0004-6361/201424973. arXiv: 1409.4780 [astro-ph.IM].

[Vlt] *The VLT white book – European Southern Observatory*.

[Gre+12] J. C. Green et al. "The Cosmic Origins Spectrograph". In: *The Astrophysical Journal* 744.1, 60 (Jan. 2012), p. 60. DOI: 10.1088/0004-637X/744/1/6010.1086/141956. arXiv: 1110.0462 [astro-ph.IM].

[Wei+00] M. C. Weisskopf et al. "Chandra X-ray Observatory (CXO): overview". In: *X-Ray Optics, Instruments, and Missions III*. Ed. by J. E. Truemper and B. Aschenbach. Vol. 4012. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July 2000, pp. 2–16. DOI: 10.1117/12.391545. arXiv: astro-ph/0004127 [astro-ph].

[Sch+24] N. Schartel et al. "XMM-Newton". In: *Handbook of X-ray and Gamma-ray Astrophysics*. Ed. by C. Bambi and A. Santangelo. Singapore: Springer Nature Singapore, 2024, pp. 1501–1538. ISBN: 978-981-19-6960-7. DOI: 10.1007/978-981-19-6960-7_41.

[Lon12] M. S. Longair. *High Energy Astrophysics*. 3rd. Cambridge: Cambridge University Press, 2012.

[CTA13] CTA Consortium. "Introducing the CTA concept". In: *Astroparticle Physics* 43 (2013). Seeing the High-Energy Universe with the Cherenkov Telescope Array - The Science Explored with the CTA, pp. 3–18. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2013.01.007.

[Aha+06] F. Aharonian et al. "Observations of the Crab nebula with HESS". In: *Astronomy and Astrophysics* 457.3 (Oct. 2006), pp. 899–915. DOI: 10.1051/0004-6361:20065351. arXiv: astro-ph/0607333 [astro-ph].

[Die01] R. Diehl. "Nucleosynthesis Gamma-Rays". Gould Belt Workshop at ESO. 2001.

[Mat+88] S. M. Matz et al. "Gamma-ray line emission from SN1987A". In: *Nature* 331.6155 (1988), pp. 416–418. ISSN: 1476-4687. DOI: 10.1038/331416a0.

[Kur+92] J. D. Kurfess et al. "Oriented Scintillation Spectrometer Experiment Observations of 57Co in SN 1987A". In: *Astrophysical Journal Letters* 399 (1992), p. L137. DOI: 10.1086/186626.

[Iyu+94] A. F. Iyudin et al. "COMPTEL observations of 44Ti gamma-ray line emission form CAS A." In: *Astronomy and Astrophysics* 284 (1994), pp. L1–L4.

[Mah+82] W. A. Mahoney et al. "Diffuse galactic gamma-ray line emission from nucleosynthetic Fe-60, Al-26, and Na-22 - Preliminary limits from HEAO 3." In: *The Astrophysical Journal* 262 (1982), pp. 742–748. DOI: 10.1086/160469.

[Smi04] D. M. Smith. "Gamma-Ray Line Observations with RHESSI". In: *5th INTEGRAL Workshop on the INTEGRAL Universe*. Ed. by V. Schoenfelder, G. Lichti, and C. Winkler. Vol. 552. ESA Special Publication. 2004, p. 45. DOI: 10.48550/arXiv.astro-ph/0404594. arXiv: astro-ph/0404594 [astro-ph].

[JH73a] I. Johnson W. N. and R. C. Haymes. "Detection of a Gamma-Ray Spectral Line from the Galactic-Center Region". In: *The Astrophysical Journal* 184 (1973), pp. 103–126. DOI: 10.1086/152309.

[Kno+25] J. Knodlseder et al. "Detection of positron in-flight annihilation from the Galaxy". In: *Astronomy and Astrophysics* 700 (2025). DOI: 10.1051/0004-6361/202556046.

[Die22] R. Diehl. "Radioactive Decay". In: *Handbook of X-ray and Gamma-ray Astrophysics*. Springer Nature Singapore, Sept. 2022, 1–35. DOI: 10.1007/978-981-16-4544-0_86-1.

[Die+21a] R. Diehl et al. "The radioactive nuclei 26Al and 60Fe in the Cosmos and in the solar system". In: *Publications of the Astronomical Society of Australia* 38, e062 (Dec. 2021), e062. DOI: 10.1017/pasa.2021.48. arXiv: 2109.08558 [astro-ph.HE].

[Obe+96] U. Oberlack et al. "The COMPTEL 1.809MeV all-sky image." In: *Astronomy and Astrophysics Supplement* 120 (Dec. 1996), pp. 311–314.

[Plu+01] S. Plueschke et al. *The Comptel 1.809 Mev Survey*. 2001. arXiv: astro-ph/0104047 [astro-ph].

[BJR15] L. Bouchet, E. Jourdain, and J.-P. Roques. "The Galactic ^{26}Al Emission Map as Revealed by INTEGRAL SPI". In: *The Astrophysical Journal* 801.2, 142 (2015), p. 142. DOI: 10.1088/0004-637X/801/2/142. arXiv: 1501.05247 [astro-ph.HE].

[Kre+13] K. Kretschmer et al. "Kinematics of massive star ejecta in the Milky Way as traced by 26Al". In: *Astronomy and Astrophysics* 559, A99 (2013), A99. DOI: 10.1051/0004-6361/201322563. arXiv: 1309.4980 [astro-ph.HE].

[DHP18] R. Diehl, D. H. Hartmann, and N. Prantzos. "Distributed Radioactivities". In: *Astrophysics with Radioactive Isotopes*. Ed. by R. Diehl, D. H. Hartmann, and N. Prantzos. Cham: Springer International Publishing, 2018, pp. 427–497. ISBN: 978-3-319-91929-4. DOI: 10.1007/978-3-319-91929-4_7.

[Ple20] M. M. Pleintinger. "Star Groups and their Nucleosynthesis". PhD thesis. Max-Planck-Institute for Extraterrestrial Physics, Garching, 2020.

[Die+06] R. Diehl et al. "Radioactive 26Al from massive stars in the Galaxy". In: *Nature* 439.7072 (2006), pp. 45–47. ISSN: 1476-4687. DOI: 10.1038/nature04364.

[Gre19] D. A. Green. "A revised catalogue of 294 Galactic supernova remnants". In: *Journal of Astrophysics and Astronomy* 40.4, 36 (Aug. 2019), p. 36. DOI: 10.1007/s12036-019-9601-6. arXiv: 1907.02638 [astro-ph.GA].

[KSP18] J. R. Knies, M. Sasaki, and P. P. Plucinsky. "Suzaku observations of the Monogem Ring and the origin of the Gemini H α ring". In: *Monthly Notices of the Royal Astronomical Society* 477.4 (July 2018), pp. 4414–4422. DOI: 10.1093/mnras/sty915.

[CP11] L. Chomiuk and M. S. Povich. "Toward a unification of star formation rate determination in the Milky Way and other galaxies". In: *The Astronomical Journal* 142.6 (2011), p. 197. DOI: 10.1088/0004-6256/142/6/197.

[Rug+09] G. Rugel et al. "New Measurement of the ^{60}Fe Half-Life". In: *Phys. Rev. Lett.* 103 (7 2009), p. 072502. DOI: 10.1103/PhysRevLett.103.072502.

[Hef+15] T. Heftrich et al. "Thermal neutron capture cross section of the radioactive isotope Fe 60". In: *Physical Review C* 92 (July 2015). DOI: 10.1103/PhysRevC.92.015806.

[Die+21b] R. Diehl et al. "The radioactive nuclei 26Al and 60Fe in the Cosmos and in the solar system". In: *Publications of the Astronomical Society of Australia* 38, e062 (Dec. 2021), e062. DOI: 10.1017/pasa.2021.48. arXiv: 2109.08558 [astro-ph.HE].

[Wan+20] W. Wang et al. "Gamma-Ray Emission of 60Fe and 26Al Radioactivity in Our Galaxy". In: *The Astrophysical Journal* 889.2 (2020), p. 169. DOI: 10.3847/1538-4357/ab6336.

[Gao+21] B. Gao et al. "New ^{59}Fe Stellar Decay Rate with Implications for the ^{60}Fe Radioactivity in Massive Stars". In: *Phys. Rev. Lett.* 126 (15 2021), p. 152701. DOI: 10.1103/PhysRevLett.126.152701.

[Spy+24] A. Spyrou et al. "Enhanced production of 60Fe in massive stars". In: *Nature Communications* 15.1 (2024), p. 9608. ISSN: 2041-1723. DOI: 10.1038/s41467-024-54040-4.

[JHH72] W. N. Johnson III, F. R. Harnden Jr., and R. C. Haymes. "The Spectrum of Low-Energy Gamma Radiation from the Galactic-Center Region." In: *Astrophysical Journal* 172 (Feb. 1972), p. L1. DOI: 10.1086/180878.

[JH73b] W. N. Johnson III and R. C. Haymes. "Detection of a Gamma-Ray Spectral Line from the Galactic-Center Region". In: *Astrophysical Journal* 184 (Aug. 1973), pp. 103–126. DOI: 10.1086/152309.

[LMS78] M. Leventhal, C. J. MacCallum, and P. D. Stang. "Detection of 511 keV positron annihilation radiation from the galactic center direction." In: *Astrophysical Journal* 225 (Oct. 1978), pp. L11–L14. DOI: 10.1086/182782.

[Alb+81] F. Albernhe et al. "Detection of the positron annihilation gamma ray line from the Galactic Center region". In: *Astronomy and Astrophysics* 94.2 (Feb. 1981), pp. 214–218.

[Pur+97] W. R. Purcell et al. "OSSE Mapping of Galactic 511 keV Positron Annihilation Line Emission". In: *The Astrophysical Journal* 491.2 (1997), p. 725. DOI: 10.1086/304994.

[Chu+20] E. Churazov et al. "INTEGRAL results on the electron-positron annihilation radiation and X-ray and Gamma-ray diffuse emission of the Milky Way". In: *New Astronomy Reviews* 90 (2020), p. 101548. ISSN: 1387-6473. DOI: https://doi.org/10.1016/j.newar.2020.101548.

[Cox05] D. P. Cox. "The Three-Phase Interstellar Medium Revisited". In: *Annual Review of Astronomy and Astrophysics* 43.Volume 43, 2005 (2005), pp. 337–385. ISSN: 1545-4282. DOI: https://doi.org/10.1146/annurev.astro.43.072103.150615.

[Jea+09] P. Jean et al. "Positron transport in the interstellar medium". In: *Astronomy and Astrophysics* 508.3 (Dec. 2009), pp. 1099–1116. DOI: 10.1051/0004-6361/200809830. arXiv: 0909.4022 [astro-ph.HE].

[Sie+16] Siegert, Thomas et al. "Gamma-ray spectroscopy of positron annihilation in the Milky Way". In: *Astronomy and Astrophysics* 586 (2016), A84. DOI: 10.1051/0004-6361/201527510.

[OP49] A. Ore and J. L. Powell. "Three-Photon Annihilation of an Electron-Positron Pair". In: *Phys. Rev.* 75 (11 1949), pp. 1696–1699. DOI: 10.1103/PhysRev.75.1696.

[Sie+19] T. Siegert et al. "Constraints on positron annihilation kinematics in the inner Galaxy". In: *Astronomy and Astrophysics* 627, A126 (July 2019), A126. DOI: 10.1051/0004-6361/201833856. arXiv: 1906.00498 [astro-ph.HE].

[CL93] K.-W. Chan and R. E. Lingenfelter. "Positrons from Supernovae". In: *Astrophysical Journal* 405 (Mar. 1993), p. 614. DOI: 10.1086/172393.

[Ale+14] A. Alexis et al. "Monte Carlo modelling of the propagation and annihilation of nucleosynthesis positrons in the Galaxy". In: *Astron. Astrophys.* 564 (2014), A108. DOI: 10.1051/0004-6361/201322393. arXiv: 1402.6110 [astro-ph.HE].

[Nav+24] S. Navas et al. "Review of Particle Physics". In: *Phys. Rev. D* 110 (3 2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[Bel99] A. M. Beloborodov. "Electron-positron outflows from gamma-ray emitting accretion discs". In: *Monthly Notices of the Royal Astronomical Society* 305.1 (May 1999), pp. 181–189. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.1999.02384.x. eprint: https://academic.oup.com/mnras/article-pdf/305/1/181/3083643/305-1-181.pdf.

[Sve82b] R. Svensson. "The pair annihilation process in relativistic plasmas". In: *Astrophysical Journal* 258 (July 1982), pp. 321–334. DOI: 10.1086/160081.

[Sve82a] R. Svensson. "Electron-Positron Pair Equilibria in Relativistic Plasmas". In: *Astrophysical Journal* 258 (July 1982), p. 335. DOI: 10.1086/160082.

[Sve83] R. Svensson. "The thermal pair annihilation spectrum - A detailed balance approach". In: *Astrophysical Journal* 270 (July 1983), pp. 300–304. DOI: 10.1086/161122.

[RM81] R Ramaty and P Meszaros. *Annihilation radiation from a hot e (+)-e (-) plasma*. Tech. rep. 1981.

[Sie23] T. Siegert. "The Positron Puzzle". In: *Astrophysics and Space Science* 368.4 (2023), p. 27. ISSN: 1572-946X. DOI: 10.1007/s10509-023-04184-4.

[Mar+12] P. Martin et al. "Galactic annihilation emission from nucleosynthesis positrons". In: *Astronomy and Astrophysics* 543, A3 (July 2012), A3. DOI: 10.1051/0004-6361/201118721. arXiv: 1205.1194 [astro-ph.HE].

[ZC97] L. Zhang and K. S. Cheng. "High-Energy Radiation from Rapidly Spinning Pulsars with Thick Outer Gaps". In: *The Astrophysical Journal* 487.1 (1997), p. 370. DOI: 10.1086/304589.

[CHR86] K. S. Cheng, C. Ho, and M. Ruderman. "Energetic Radiation from Rapidly Spinning Pulsars. I. Outer Magnetosphere Gaps". In: *The Astrophysical Journal* 300 (Jan. 1986), p. 500. DOI: 10.1086/163829.

[Bou+91] L. Bouchet et al. "SIGMA Discovery of Variable E +- E - Annihilation Radiation from the Near Galactic Center Variable Compact Source 1E 1740.7-2942". In: *Astrophysical Journal Letters* 383 (Dec. 1991), p. L45. DOI: 10.1086/186237.

[Gol+92] A. Goldwurm et al. "SIGMA/GRANAT Soft Gamma-Ray Observations of the X-Ray Nova in Musca: Discovery of Positron Annihilation Emission Line". In: *Astrophysical Journal Letters* 389 (Apr. 1992), p. L79. DOI: 10.1086/186353.

[Sie+16b] T. Siegert et al. "Positron annihilation signatures associated with the outburst of the microquasar V404 Cygni". In: *Nature* 531.7594 (2016), pp. 341–343. ISSN: 1476-4687. DOI: 10.1038/nature16978.

[Die01] R. Diehl. "Gamma-Ray Production and Absorption Processes". In: *The Universe in Gamma Rays*. Ed. by V. Schönfelder. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 9–25. ISBN: 978-3-662-04593-0. DOI: 10.1007/978-3-662-04593-0_2.

[SS84] W. Sacher and V. Schoenfelder. "The bremsstrahlung component of the diffuse galactic gamma-ray emission at MeV energies". In: *Astrophysical Journal* 279 (Apr. 1984), pp. 817–826. DOI: 10.1086/161952.

[Fun15] S. Funk. "Ground- and Space-Based Gamma-Ray Astronomy". In: *Annual Review of Nuclear and Particle Science* 65.1 (2015), pp. 245–277. DOI: 10.1146/annurev-nucl-102014-022036. eprint: https://doi.org/10.1146/annurev-nucl-102014-022036.

[AAK97] F. A. Aharonian, A. M. Atoyan, and T. Kifune. "Inverse Compton gamma radiation of faint synchrotron X-ray nebulae around pulsars". In: *Monthly Notices of the Royal Astronomical Society* 291.1 (Oct. 1997), pp. 162–176. DOI: 10.1093/mnras/291.1.162.

[BG70] G. R. BLUMENTHAL and R. J. GOULD. "Bremsstrahlung, Synchrotron Radiation, and Compton Scattering of High-Energy Electrons Traversing Dilute Gases". In: *Rev. Mod. Phys.* 42 (2 1970), pp. 237–270. DOI: 10.1103/RevModPhys.42.237.

[Lev18] A. Levan. *Gamma-Ray Bursts*. 2514-3433. IOP Publishing, 2018. ISBN: 978-0-7503-1502-9. DOI: 10.1088/2514-3433/aae164.

[KSO73] R. W. Klebesadel, I. B. Strong, and R. A. Olson. "Observations of Gamma-Ray Bursts of Cosmic Origin". In: *The Astrophysical Journal* 182 (June 1973), p. L85. DOI: 10.1086/181225.

[Bri+96] M. S. Briggs et al. "BATSE Observations of the Large-Scale Isotropy of Gamma-Ray Bursts". In: *The Astrophysical Journal* 459 (Mar. 1996), p. 40. DOI: 10.1086/176867. arXiv: astro-ph/9509078 [astro-ph].

[Mee+98] C. A. Meegan et al. "The 4B BATSE gamma-ray burst catalog". In: *Gamma-Ray Bursts, 4th Hunstville Symposium*. Ed. by C. A. Meegan, R. D. Preece, and T. M. Koshut. Vol. 428. American Institute of Physics Conference Series. AIP, May 1998, pp. 3–9. DOI: 10.1063/1.55355.

[Fro19] F. Frontera. "The key role of BeppoSAX in the GRB history". In: *Rendiconti Lincei. Scienze Fisiche e Naturali* 30.1 (2019), pp. 171–184. ISSN: 1720-0776. DOI: 10.1007/s12210-019-00766-z.

[Ama+99] L. Amati et al. "BeppoSAX observations of GRB970508: first evidence of bursting activity continuing on very long time scale". In: *Nuclear Physics B - Proceedings Supplements* 69.1 (1999). The Active X-ray Sky Results from BeppoSAX and RXTE, pp. 656–659. ISSN: 0920-5632. DOI: https://doi.org/10.1016/S0920-5632(98)00316-8.

[Met+97] M. R. Metzger et al. "Spectral constraints on the redshift of the optical counterpart to the gamma-ray burst of 8 May 1997". In: *Nature* 387.6636 (1997), pp. 878–880. ISSN: 1476-4687. DOI: 10.1038/43132.

[Bur+05] D. N. Burrows et al. "The Swift X-Ray Telescope". In: *Space Science Reviews* 120.3–4 (Oct. 2005), 165–195. ISSN: 1572-9672. DOI: 10.1007/s11214-005-5097-2.

[Hor91] J. Horacek. *Development of the Burst and Transient Source Experiment (BATSE)*. Tech. rep. 19910022692. 1991.

[Kou+93a] C. Kouveliotou et al. "Identification of Two Classes of Gamma-Ray Bursts". In: *The Astrophysical Journal Letters* 413 (Aug. 1993), p. L101. DOI: 10.1086/186969.

[Ban+93] D. Band et al. "BATSE Observations of Gamma-Ray Burst Spectra. I. Spectral Diversity". In: *The Astrophysical Journal* 413 (Aug. 1993), p. 281. DOI: 10.1086/172995.

[Bri+99] M. S. Briggs et al. "Observations of GRB 990123 by the Compton Gamma Ray Observatory". In: *The Astrophysical Journal* 524.1 (1999), p. 82. DOI: 10.1086/307808.

[Atw+09] W. B. Atwood et al. "The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission". In: *The Astrophysical Journal* 697 (June 2009), pp. 1071–1102. DOI: 10.1088/0004-637X/697/2/1071. arXiv: 0902.1089 [astro-ph.IM].

[Mee+09] C. Meegan et al. "The Fermi Gamma-ray Burst Monitor". In: *The Astrophysical Journal* 702.1 (Sept. 2009), pp. 791–804. DOI: 10.1088/0004-637X/702/1/791. arXiv: 0908.0450 [astro-ph.IM].

[Abd+09] A. A. Abdo et al. "Fermi Observations of High-Energy Gamma-Ray Emission from GRB 080916C". In: *Science* 323.5922 (2009), pp. 1688–1693. DOI: 10.1126/science.1169101. eprint: https://www.science.org/doi/pdf/10.1126/science.1169101.

[Fer05] D. Ferenc. "The MAGIC gamma-ray observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 553.1 (2005). Proceedings of the fifth International Workshop on Ring Imaging Detectors, pp. 274–281. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2005.08.085.

[Nat] "Teraelectronvolt emission from the -ray burst GRB 190114C". In: *Nature* 575.7783 (Nov. 2019), 455–458. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1750-x.

[Acc+19] V. A. Acciari et al. "Observation of inverse Compton emission from a long γ-ray burst". In: *Nature* 575.7783 (2019), pp. 459–463. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1754-6.

[Kou+93b] C. Kouveliotou et al. "Identification of Two Classes of Gamma-Ray Bursts". In: *The Astrophysical Journal* 413 (Aug. 1993), p. L101. DOI: 10.1086/186969.

[Pir99] L. Piro. "GRB990123, BeppoSAX WFC detection and NFI planned follow-up." In: *GRB Coordinates Network* 199 (Jan. 1999), p. 1.

[Gis+99] G. Gisler et al. "The ROTSE detection of early optical light from GRB990123". In: *AIP Conference Proceedings* 499.1 (Dec. 1999), pp. 82–89. ISSN: 0094-243X. DOI: 10.1063/1.1302223. eprint: https://pubs.aip.org/aip/acp/article-pdf/499/1/82/12139774/82_1_online.pdf.

[Fru+23] A. S. Fruchter et al. "Hubble Space Telescope and Palomar Imaging of GRB 990123: Implications for the Nature of Gamma-Ray Bursts and Their Hosts". In: (Aug. 2023). DOI: 10.1086/312094.

[WB06] S. Woosley and J. Bloom. "The Supernova–Gamma-Ray Burst Connection". In: *Annual Review of Astronomy and Astrophysics* 44.1 (Sept. 2006), 507–556. ISSN: 1545-4282. DOI: 10.1146/annurev.astro.43.072103.150558.

[Pod13] P. Podsiadlowski. "Supernovae and Gamma-Ray Bursts". In: *Planets, Stars and Stellar Systems: Volume 4: Stellar Structure and Evolution*. Ed. by T. D. Oswalt and M. A. Barstow. Dordrecht: Springer Netherlands, 2013, pp. 693–733. ISBN: 978-94-007-5615-1. DOI: 10.1007/978-94-007-5615-1_14.

[CT+99] A. J. Castro-Tirado et al. "Decay of the GRB 990123 Optical Afterglow: Implications for the Fireball Model". In: *Science* 283.5410 (1999), pp. 2069–2073. DOI: 10.1126/science.283.5410.2069. eprint: https://www.science.org/doi/pdf/10.1126/science.283.5410.2069.

[Woo93] S. E. Woosley. "Gamma-Ray Bursts from Stellar Mass Accretion Disks around Black Holes". In: *The Astrophysical Journal* 405 (Mar. 1993), p. 273. DOI: 10.1086/172359.

[Whe+00] J. C. Wheeler et al. "Asymmetric Supernovae, Pulsars, Magnetars, and Gamma-Ray Bursts". In: *The Astrophysical Journal* 537.2 (July 2000), 810–823. ISSN: 1538-4357. DOI: 10.1086/309055.

[Nag18] S. Nagataki. "Theories of central engine for long gamma-ray bursts". In: *Reports on Progress in Physics* 81.2 (2018), p. 026901. DOI: 10.1088/1361-6633/aa97a8.

[BZ77] R. D. Blandford and R. L. Znajek. "Electromagnetic extraction of energy from Kerr black holes". In: *Monthly Notices of the Royal Astronomical Society* 179.3 (July 1977), pp. 433–456. ISSN: 0035-8711. DOI: 10.1093/mnras/179.3.433. eprint: https://academic.oup.com/mnras/article-pdf/179/3/433/9333653/mnras179-0433.pdf.

[Cio18] R. Ciolfi. "Short gamma-ray burst central engines". In: *International Journal of Modern Physics D* 27.13 (2018), p. 1842004. DOI: 10.1142/S021827181842004X.

[FIH89] M. Franx, G. Illingworth, and T. Heckman. "Multicolor Surface Photometry of 17 Ellipticals". In: *Astronomical Journal* 98 (Aug. 1989), p. 538. DOI: 10.1086/115157.

[GF10] G. J. Graves and S. M. Faber. "DISSECTING THE RED SEQUENCE. III. MASS-TO-LIGHT VARIATIONS IN THREE-DIMENSIONAL FUNDAMENTAL PLANE SPACE". In: *The Astrophysical Journal* 717.2 (June 2010), 803–824. ISSN: 1538-4357. DOI: 10.1088/0004-637x/717/2/803.

[Nak07] E. Nakar. "Short-hard gamma-ray bursts". In: *Physics Reports* 442.1 (2007). The Hans Bethe Centennial Volume 1906-2006, pp. 166–236. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2007.02.005.

[Ber11] E. Berger. "The environments of short-duration gamma-ray bursts and implications for their progenitors". In: *New Astronomy Reviews* 55.1 (2011), pp. 1–22. ISSN: 1387-6473. DOI: https://doi.org/10.1016/j.newar.2010.10.001.

[NPP92] R. Narayan, B. Paczynski, and T. Piran. "Gamma-Ray Bursts as the Death Throes of Massive Binary Stars". In: *The Astrophysical Journal* 395 (Aug. 1992), p. L83. DOI: 10.1086/186493. arXiv: astro-ph/9204001 [astro-ph].

[Eic+89] D. Eichler et al. "Nucleosynthesis, neutrino bursts and gamma-rays from coalescing neutron stars". In: *Nature* 340.6229 (1989), pp. 126–128. ISSN: 1476-4687. DOI: 10.1038/340126a0.

[Aa17] B. P. Abbott and et al. "Multi-messenger Observations of a Binary Neutron Star Merger*". In: *The Astrophysical Journal Letters* 848.2 (Oct. 2017), p. L12. ISSN: 2041-8213. DOI: 10.3847/2041-8213/aa91c9.

[Lyu04] M. Lyutikov. "Electromagnetic (versus Fireball) Model of GRBs". In: *AIP Conference Proceedings*. AIP, 2004. DOI: 10.1063/1.1810906.

[DDR04] A. Dar and A. De Rújula. "Towards a complete theory of gamma-ray bursts". In: *Physics Reports* 405.4 (Dec. 2004), 203–278. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2004.09.008.

[Dar05] A. Dar. "Fireball and Cannonball Models of Gamma-Ray Bursts Confront Observations". In: *Chinese Journal of Astronomy and Astrophysics* 6 (Nov. 2005). DOI: 10.1088/1009-9271/6/S1/39.

[Goo86] J. Goodman. "Are gamma-ray bursts optically thick?" In: *The Astrophysical Journal Letters* 308 (Sept. 1986), p. L47. DOI: 10.1086/184741.

[Pac86] B. Paczynski. "Gamma-ray bursters at cosmological distances". In: *The Astrophysical Journal Letters* 308 (Sept. 1986), pp. L43–L46. DOI: 10.1086/184740.

[DDDR22] S. Dado, A. Dar, and A. De Rújula. "Critical Tests of Leading Gamma Ray Burst Theories". In: *Universe* 8.7 (2022). ISSN: 2218-1997. DOI: 10.3390/universe8070350.

[LB02] M. Lyutikov and R. Blandford. *Electromagnetic Outflows and GRBs*. 2002. arXiv: astro-ph/0210671 [astro-ph].

[GKG21] R. Gill, M. Kole, and J. Granot. "GRB Polarization: A Unique Probe of GRB Physics". In: *Galaxies* 9.4 (2021). ISSN: 2075-4434. DOI: 10.3390/galaxies9040082.

[CB03] W. Coburn and S. E. Boggs. "Polarization of the prompt -ray emission from the -ray burst of 6 December 2002". In: *Nature* 423.6938 (May 2003), 415–417. ISSN: 1476-4687. DOI: 10.1038/nature01612.

[Wig+04] C. Wigger et al. "Gamma-ray burst polarization: limits from rhessi measurements". In: *Astrophys. J.* 613 (2004), pp. 1088–1100. DOI: 10.1086/423163. arXiv: astro-ph/0405525.

[RF04] R. E. Rutledge and D. B. Fox. "Re-analysis of polarization in the -ray flux of GRB 021206". In: *Monthly Notices of the Royal Astronomical Society* 350.4 (June 2004), pp. 1288–1300. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2004.07665.x. eprint: https://academic.oup.com/mnras/article-pdf/350/4/1288/3086901/350-4-1288.pdf.

[McC17] M. L. McConnell. "High energy polarimetry of prompt GRB emission". In: *New Astronomy Reviews* 76 (Feb. 2017), 1–21. ISSN: 1387-6473. DOI: 10.1016/j.newar.2016.11.001.

[Pro+18] N. Produit et al. "Design and construction of the POLAR detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 877 (2018), pp. 259–268. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2017.09.053.

[Kol+20] M. Kole et al. "The POLAR gamma-ray burst polarization catalog". In: *Astronomy and Astrophysics* 644, A124 (Dec. 2020), A124. DOI: 10.1051/0004-6361/202037915. arXiv: 2009.04871 [astro-ph.HE].

[Pro+23] N. Produit et al. "POLAR-2, the next generation of GRB polarization detector". In: *Proceedings of 38th International Cosmic Ray Conference — PoS(ICRC2023)*. ICRC2023. Sissa Medialab, Aug. 2023. DOI: 10.22323/1.444.0550.

[De +17a] A. De Angelis et al. "The e-ASTROGAM mission - Exploring the extreme Universe with gamma rays in the MeV - GeV range". In: *Experimental Astronomy* (June 2017). DOI: 10.1007/s10686-017-9533-6. arXiv: 1611.02232 [astro-ph.HE].

[McE+19] J. McEnery et al. *All-sky Medium Energy Gamma-ray Observatory: Exploring the Extreme Multimessenger Universe*. 2019. arXiv: 1907.07558 [astro-ph.IM].

[RFT80] V. C. Rubin, W. K. Ford Jr., and N. Thonnard. "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc)." In: *Astrophysical Journal* 238 (June 1980), pp. 471–487. DOI: 10.1086/158003.

[Clo+06] D. Clowe et al. "A Direct Empirical Proof of the Existence of Dark Matter". In: *Astrophysical Journal Letters* 648.2 (Sept. 2006), pp. L109–L113. DOI: 10.1086/508162. arXiv: astro-ph/0608407 [astro-ph].

[Pla+16] Planck Collaboration et al. "Planck 2015 results - XIII. Cosmological parameters". In: *Astronomy and Astrophysics* 594 (2016), A13. DOI: 10.1051/0004-6361/201525830.

[BGW17] R. Bartels, D. Gaggero, and C. Weniger. "Prospects for indirect dark matter searches with MeV photons". In: *Journal of Cosmology and Astroparticle Physics* 2017.05 (May 2017), 001–001. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2017/05/001.

[Cir+11] M. Cirelli et al. "PPPC 4 DM ID: a poor particle physicist cookbook for dark matter indirect detection". In: *Journal of Cosmology and Astroparticle Physics* 2011.03 (Mar. 2011), 051–051. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2011/03/051.

[Car+10] B. J. Carr et al. "New cosmological constraints on primordial black holes". In: *Physical Reviews D* 81.10, 104019 (May 2010), p. 104019. DOI: 10.1103/PhysRevD.81.104019. arXiv: 0912.5297 [astro-ph.CO].

[Dol02] A. D. Dolgov. *Cosmological Matter-Antimatter Asymmetry and Antimatter in the Universe*. 2002. arXiv: hep-ph/0211260 [hep-ph].

[Ste08] G. Steigman. "When clusters collide: constraints on antimatter on the largest scales". In: *Journal of Cosmology and Astroparticle Physics* 2008.10 (Oct. 2008), p. 001. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2008/10/001.

[EE55] R. D. Evans and R. Evans. *The atomic nucleus*. Vol. 582. McGraw-Hill New York, 1955.

[De +17b] A. De Angelis et al. "The e-ASTROGAM mission. Exploring the extreme Universe with gamma rays in the MeV - GeV range". In: *Experimental Astronomy* 44.1 (Oct. 2017), pp. 25–82. DOI: 10.1007/s10686-017-9533-6. arXiv: 1611.02232 [astro-ph.HE].

[Ber+09] M. Berger et al. "XCOM: Photon cross sections database". In: *NIST Standard Reference Database* 8 (Jan. 2009), pp. 87–3597.

[KN29] O. Klein and T. Nishina. "Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac". In: *Zeitschrift fur Physik* 52 (Nov. 1929), pp. 853–868. DOI: 10.1007/BF01366453.

[Com23] W. Commons. *Klein-Nishina distribution*. File: Klein-Nishina$_d$*istribution–en.svg*. 2023. URL: https://https://commons.wikimedia.org/wiki/File:Klein-Nishina_distribution-en.svg.

[ZAS06] A. Zoglauer, R. Andritschke, and F. Schopper. "MEGAlib The Medium Energy Gamma-ray Astronomy Library". In: *New Astronomy Reviews* 50.7-8 (Oct. 2006), pp. 629–632. DOI: 10.1016/j.newar.2006.06.049.

[DM29] J. W. M. Du Mond. "Compton Modified Line Structure and its Relation to the Electron Theory of Solid Bodies". In: *Phys. Rev.* 33 (5 1929), pp. 643–658. DOI: 10.1103/PhysRev.33.643.

[RB82] R. Ribberfors and K. F. Berggren. "Incoherent-x-ray-scattering functions and cross sections $(\frac{d\sigma}{d\Omega'})_{\text{incoh}}$ by means of a pocket calculator". In: *Phys. Rev. A* 26 (6 1982), pp. 3325–3333. DOI: 10.1103/PhysRevA.26.3325.

[ZK03] A. Zoglauer and G. Kanbach. "Doppler broadening as a lower limit to the angular resolution of next-generation Compton telescopes". In: *X-Ray and Gamma-Ray Telescopes and Instruments for Astronomy*. Ed. by J. E. Truemper and H. D. Tananbaum. Vol. 4851. Proceedings of SPIE. Mar. 2003, pp. 1302–1309. DOI: 10.1117/12.461177.

[CGG16] M. J. Cieślak, K. A. Gamage, and R. Glover. "Coded-aperture imaging systems: Past, present and future development – A review". In: *Radiation Measurements* 92 (2016), pp. 59–71. ISSN: 1350-4487. DOI: https://doi.org/10.1016/j.radmeas.2016.08.002.

[Car+87] E. Caroli et al. "Coded aperture imaging in X- and gamma-ray astronomy". In: *Space Science Reviews* 45.3 (1987), pp. 349–403. ISSN: 1572-9672. DOI: 10.1007/BF00171998.

[FC78] E. E. Fenimore and T. M. Cannon. "Coded aperture imaging with uniformly redundant arrays". In: *Appl. Opt.* 17.3 (1978), pp. 337–347. DOI: 10.1364/AO.17.000337.

[GG22] A. Goldwurm and A. Gros. "Coded Mask Instruments for Gamma-Ray Astronomy". In: *Handbook of X-ray and Gamma-ray Astrophysics*. Ed. by C. Bambi and A. Santangelo. Springer Nature Singapore, 2022, pp. 1–57. ISBN: 978-981-16-4544-0. DOI: 10.1007/978-981-16-4544-0_44-1.

[Bou+01] L. Bouchet et al. "The SIGMA/Granat Telescope: Calibration and Data Reduction". In: *The Astrophysical Journal* 548.2 (2001), p. 990. DOI: 10.1086/318997.

[Win+03a] C. Winkler et al. "The INTEGRAL mission". In: *Astronomy and Astrophysics* 411 (Nov. 2003), pp. L1–L6. DOI: 10.1051/0004-6361:20031288.

[Ama+18] L. Amati et al. "The THESEUS space mission concept: science case, design and expected performances". In: *Advances in Space Research* 62.1 (July 2018), 191–244. ISSN: 0273-1177. DOI: 10.1016/j.asr.2018.03.010.

[SGD82] V. Schoenfelder, U. Graser, and R. Diehl. "Properties and performance of the MPI balloon-borne Compton telescope." In: *Astronomy and Astrophysics* 110 (June 1982), pp. 138–151.

[Sch+93] V Schönfelder et al. "Instrument description and performance of the Imaging Gamma-Ray Telescope COMPTEL aboard the Compton Gamma-Ray Observatory". In: 86 (May 1993), pp. 657–692.

[Obe+00] U. G. Oberlack et al. "Compton scattering sequence reconstruction algorithm for the liquid xenon gamma-ray imaging telescope (lxegrit)". In: *Proc. SPIE Int. Soc. Opt. Eng.* 4141 (2000), p. 168. DOI: 10.1117/12.407578. arXiv: astro-ph/0012296 [astro-ph].

[BJ00] S. E. Boggs and P. Jean. "Event reconstruction in high resolution Compton telescopes". In: *Astronomy and Astrophysics Supplement Series* 145 (Aug. 2000), pp. 311–321. DOI: 10.1051/aas:2000107. arXiv: astro-ph/0005250 [astro-ph].

[Orl+22] E. Orlando et al. "Exploring the MeV sky with a combined coded mask and Compton telescope: the Galactic Explorer with a Coded aperture mask Compton telescope (GECCO)". In: *Journal of Cosmology and Astroparticle Physics* 2022.07 (2022), p. 036. DOI: 10.1088/1475-7516/2022/07/036.

[Apr+08] E. Aprile et al. "Compton imaging of MeV gamma-rays with the Liquid Xenon Gamma-Ray Imaging Telescope (LXeGRIT)". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 593.3 (2008), pp. 414–425. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2008.05.039.

[KL24] S. M. Kim and J. S. Lee. "A comprehensive review on Compton camera image reconstruction: from principles to AI innovations". In: *Biomedical Engineering Letters* 14.6 (2024), pp. 1175–1193. ISSN: 2093-985X. DOI: 10.1007/s13534-024-00418-8.

[KTK24] C. Kierans, T. Takahashi, and G. Kanbach. "Compton Telescopes for Gamma-Ray Astrophysics". In: *Handbook of X-ray and Gamma-ray Astrophysics*. Ed. by C. Bambi and A. Santangelo. Singapore: Springer Nature Singapore, 2024, pp. 1711–1782. ISBN: 978-981-19-6960-7. DOI: 10.1007/978-981-19-6960-7_46.

[Die+92] R. Diehl et al. "Data analysis of the COMPTEL instrument on the NASA Gamma Ray Observatory." In: *NASA Conference Publication*. Ed. by C. R. Shrader, N. Gehrels, and B. Dennis. Vol. 3137. NASA Conference Publication. 1992, pp. 95–101.

[Tak09] S. Takeda. "Experimental Study of a Si/CdTe Semiconductor Compton Camera for the Next Generation of Gamma-ray Astronomy." PhD thesis. , The Univer- sity of Tokyo, Japan, 2009.

[Ols63] H. Olsen. "Opening Angles of Electron-Positron Pairs". In: *Phys. Rev.* 131 (1 1963), pp. 406–415. DOI: 10.1103/PhysRev.131.406.

[Atw+13] W. Atwood et al. "Pass 8: Toward the Full Realization of the Fermi-LAT Scientific Potential". In: Mar. 2013. arXiv: 1303.3514 [astro-ph.IM].

[TM22] D. J. Thompson and A. A. Moiseev. "Pair Production Detectors for Gamma-Ray Astrophysics". In: *Handbook of X-ray and Gamma-ray Astrophysics*. Ed. by C. Bambi and A. Santangelo. Singapore: Springer Nature Singapore, 2022, pp. 1–33. ISBN: 978-981-16-4544-0. DOI: 10.1007/978-981-16-4544-0_159-1.

[Hun+14] S. D. Hunter et al. "A pair production telescope for medium-energy gamma-ray polarimetry". In: *Astroparticle Physics* 59 (2014), pp. 18–28. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2014.04.002.

[Blo+06] P. F. Bloser et al. "The MEGA project: Science goals and hardware development". In: *New Astronomy Reviews* 50.7-8 (Oct. 2006), pp. 619–623. DOI: 10.1016/j.newar.2006.06.001.

[Cap+24] R. Caputo et al. "ComPair-2: a next-generation medium-energy gamma-ray telescope prototype". In: *Space Telescopes and Instrumentation 2024: Ultraviolet to Gamma Ray*. Ed. by J.-W. A. den Herder, S. Nikzad, and K. Nakazawa. Vol. 13093. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Aug. 2024, 130932L, p. 130932L. DOI: 10.1117/12.3017619. arXiv: 2412.02562 [astro-ph.IM].

[De +21] A. De Angelis et al. "Gamma-ray astrophysics in the MeV range". In: *Experimental Astronomy* 51.3 (June 2021), pp. 1225–1254. DOI: 10.1007/s10686-021-09706-y. arXiv: 2102.02460 [astro-ph.IM].

[SLL24] E. Stevenson, F. Letizia, and S. Lemmens. *ESA's Annual Space Environment Report*. Tech. rep. GEN-DB-LOG-00288-OPS-SD. ESA Space Debris Office, 2024.

[LSY19] Y. Lu, Q. Shao, and H. Yue. "A Review of the Space Environment Effects on Spacecraft in Different Orbits". In: *IEEE Access* PP (July 2019), pp. 1–1. DOI: 10.1109/ACCESS.2019.2927811.

[OS13] E. Orlando and A. Strong. "Cosmic rays: interstellar gamma-ray and radio emission". In: *Nuclear Physics B Proceedings Supplements* 239 (June 2013), pp. 64–69. DOI: 10.1016/j.nuclphysbps.2013.05.010. arXiv: 1303.5488 [astro-ph.HE].

[Hun+97] S. D. Hunter et al. "EGRET Observations of the Diffuse Gamma-Ray Emission from the Galactic Plane". In: *The Astrophysical Journal* 481 (May 1997), pp. 205–240. DOI: 10.1086/304012.

[Ack+15] M. Ackermann et al. "The spectrum of isotropic diffuse gamma-ray emission between 100 MeV and 820 GeV". In: *Astrophys. J.* 799 (2015), p. 86. DOI: 10.1088/0004-637X/799/1/86. arXiv: 1410.3696 [astro-ph.HE].

[Aje+15] M. Ajello et al. "The Origin of the Extragalactic Gamma-Ray Background and Implications for Dark-Matter Annihilation". In: *Astrophys. J.* 800.2 (2015), p. L27. DOI: 10.1088/2041-8205/800/2/L27. arXiv: 1501.05301 [astro-ph.HE].

[And+07] S. Ando et al. "Dark matter annihilation or unresolved astrophysical sources? Anisotropy probe of the origin of the cosmic gamma-ray background". In: *Phys. Rev. D* 75 (6 2007), p. 063519. DOI: 10.1103/PhysRevD.75.063519.

[Sie+22] Siegert, Thomas et al. "Diffuse Galactic emission spectrum between 0.5 and 8.0 MeV". In: *Astronomy and Astrophysics* 660 (2022), A130. DOI: 10.1051/0004-6361/202142639.

[Cum+19] P. Cumani et al. "Background for a gamma-ray satellite on a low-Earth orbit". In: *Experimental Astronomy* 47 (Mar. 2019), pp. 1–30. DOI: 10.1007/s10686-019-09624-0.

[Tue+10] M. Tuerler et al. "INTEGRAL hard X-ray spectra of the cosmic X-ray background and Galactic ridge emission". In: *Astronomy and Astrophysics* 512, A49 (Mar. 2010), A49. DOI: 10.1051/0004-6361/200913072. arXiv: 1001.2110 [astro-ph.CO].

[Dru12] L. O. Drury. "Origin of cosmic rays". In: *Astroparticle Physics* 39-40 (2012). Cosmic Rays Topical Issue, pp. 52–60. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2012.02.006.

[TBF22] N. Tomassetti, B. Bertucci, and E. Fiandrini. "Temporal evolution and rigidity dependence of the solar modulation lag of Galactic cosmic rays". In: *Phys. Rev. D* 106 (10 2022), p. 103022. DOI: 10.1103/PhysRevD.106.103022.

[Sto55] C. Stoermer. *The polar aurora*. International monographs on radio. Clarendon Press, 1955.

[Sch01] V. Schönfelder, ed. *The Universe in Gamma Rays*. Berlin Heidelberg: Springer, 2001.

[Dea+03] A. Dean et al. "The Modelling of Background Noise in Astronomical Gamma Ray Telescopes". In: *Space Science Reviews* 105.1 (2003), pp. 285–376. ISSN: 1572-9672. DOI: 10.1023/A:1023995803108.

[Zom07] M. Zombeck. *Handbook of Space Astronomy and Astrophysics: Third Edition*. Cambridge University Press, 2007.

[Abd+09] A. A. Abdo et al. "Fermi large area telescope observations of the cosmic-ray induced gamma-ray emission of the Earth's atmosphere". In: *Physical Review D* 80.12, 122004 (Dec. 2009), p. 122004. DOI: 10.1103/PhysRevD.80.122004. arXiv: 0912.1868 [astro-ph.HE].

[HSL03] M. J. Harris, G. H. Share, and M. D. Leising. "Spatial and temporal variability of the gamma radiation from Earth's atmosphere during a solar cycle". In: *Journal of Geophysical Research: Space Physics* 108.A12 (2003). 1435. ISSN: 2156-2202. DOI: 10.1029/2003JA009958.

[Chu+07] E. Churazov et al. "INTEGRAL observations of the cosmic X-ray background in the 5-100 keV range via occultation by the Earth". In: *AStronomy and Astrophysics* 467 (May 2007), pp. 529–540. DOI: 10.1051/0004-6361:20066230. eprint: astro-ph/0608250.

[Saz+07] S. Y. Sazonov et al. "Hard X-ray emission of the Earth's atmosphere: Monte Carlo simulations". In: *Mon. Not. Roy. Astron. Soc.* 377 (2007), pp. 1726–1736. DOI: 10.1111/j.1365-2966.2007.11746.x. arXiv: astro-ph/0608253 [astro-ph].

[Miz+04] T. Mizuno et al. "Cosmic-ray background flux model based on a gamma-ray large-area space telescope balloon flight engineering model". In: *Astrophys. J.* 614 (2004), pp. 1113–1123. DOI: 10.1086/423801. arXiv: astro-ph/0406684 [astro-ph].

[KPM15] M. Kole, M. Pearce, and M. Muñoz Salinas. "A model of the cosmic ray induced atmospheric neutron environment". In: *Astroparticle Physics* 62 (Mar. 2015), pp. 230–240. DOI: 10.1016/j.astropartphys.2014.10.002. arXiv: 1410.1364 [astro-ph.IM].

[Adr+15] O. Adriani et al. "Reentrant albedo proton fluxes measured by the PAMELA experiment". In: *Journal of Geophysical Research (Space Physics)* 120 (May 2015), pp. 3728–3738. DOI: 10.1002/2015JA021019. arXiv: 1504.06253 [astro-ph.HE].

[MSIV77] D M. Sawyer and J I. Vette. "AP8 Trapped Proton Environment for Solar Maximum and Solar Minimum". In: (Jan. 1977).

[ARB18] P. Anderson, F. Rich, and S. Borisov. "Mapping the South Atlantic Anomaly continuously over 27 years". In: *Journal of Atmospheric and Solar-Terrestrial Physics* 177 (2018). Dynamics of the Sun-Earth System: Recent Observations and Predictions, pp. 237–246. ISSN: 1364-6826. DOI: https://doi.org/10.1016/j.jastp.2018.03.015.

[Spe] *Space Environment Information System*. https://www.spenvis.oma.be.

[Wei+03] G. Weidenspointner et al. "First identification and modelling of SPI background lines". In: *Astronomy and Astrophysics* 411 (Nov. 2003). DOI: 10.1051/0004-6361:20031209.

[Wei+01] G. Weidenspointner et al. "The COMPTEL instrumental line background". In: *Astronomy and Astrophysics* 368 (Mar. 2001), pp. 347–368. DOI: 10.1051/0004-6361:20000489. arXiv: astro-ph/0012332 [astro-ph].

[Kni+88] D Kniffen et al. "The gamma ray observatory science plan". In: *NASA/-Goddard Space Flight Center* (1988).

[Joh+89] W. N. Johnson et al. "The Oriented Scintillation Spectrometer Experiment (OSSE) instrument description". In: *Proceedings of the GRO Science Workshop*. 1989.

[Str+97] A. Strong et al. "Whole Sky Maps Using Five Years of COMPTEL Data". In: *The Transparent Universe*. Ed. by C. Winkler, T. J.-L. Courvoisier, and P. Durouchoux. Vol. 382. ESA Special Publication. 1997, p. 533.

[Obe97] U. Oberlack. "Über die Natur der galaktischen 26 Al-Quellen - Untersuchung des 1.8 MeV Himmels mit COMPTEL". PhD thesis. -, Jan. 1997.

[Kan+89] G. Kanbach et al. "The project EGRET (energetic gamma-ray experiment telescope) on NASA's Gamma-Ray Observatory GRO". In: *Space Science Reviews* 49.1 (1989), pp. 69–84. ISSN: 1572-9672. DOI: 10.1007/BF00173744.

[CG08] J.-M. Casandjian and I. A. Grenier. "A revised catalogue of EGRET gamma-ray sources". In: *Astronomy and Astrophysics* 489.2 (2008), pp. 849–883. DOI: 10.1051/0004-6361:200809685. arXiv: 0806.0113 [astro-ph].

[Ube+03] P. Ubertini et al. "IBIS: The Imager on-board INTEGRAL". In: *Astronomy and Astrophysics* 411 (Nov. 2003), pp. L131–L139. DOI: 10.1051/0004-6361:20031224.

[Ved+03] G. Vedrenne et al. "SPI: The spectrometer aboard INTEGRAL". In: *Astronomy and Astrophysics* 411 (Nov. 2003), pp. L63–L70. DOI: 10.1051/0004-6361:20031482.

[Lun+03] N. Lund et al. "JEM-X: The X-ray monitor aboard INTEGRAL". In: *Astronomy and Astrophysics* 411 (Nov. 2003), pp. L231–L238. DOI: 10.1051/0004-6361:20031358.

[Mas+03] J. M. Mas-Hesse et al. "OMC: An Optical Monitoring Camera for INTEGRAL. Instrument description and performance". In: *Astronomy and Astrophysics* 411 (Nov. 2003), pp. L261–L268. DOI: 10.1051/0004-6361:20031418.

[Mal+16] A. Malizia et al. "The INTEGRAL/IBIS AGN catalogue: an update". In: *Mon. Not. Roy. Astron. Soc.* 460.1 (2016), pp. 19–29. DOI: 10.1093/mnras/stw972. arXiv: 1604.06303 [astro-ph.HE].

[Kno+05] J. Knodlseder et al. "The All-sky distribution of 511 keV electron-positron annihilation emission". In: *Astron. Astrophys.* 441 (2005), pp. 513–532. DOI: 10.1051/0004-6361:20042063. arXiv: astro-ph/0506026 [astro-ph].

[Sie+16a] T. Siegert et al. "Gamma-ray spectroscopy of Positron Annihilation in the Milky Way". In: *Astron. Astrophys.* 586 (2016), A84. DOI: 10.1051/0004-6361/201527510. arXiv: 1512.00325 [astro-ph.HE].

[Ack+15] M. Ackermann et al. "The Spectrum of Isotropic Diffuse Gamma-Ray Emission between 100 MeV and 820 GeV". In: *The Astrophysical Journal* 799.1, 86 (Jan. 2015), p. 86. DOI: 10.1088/0004-637X/799/1/86. arXiv: 1410.3696 [astro-ph.HE].

[Bal+23] J. Ballet et al. "Fermi Large Area Telescope Fourth Source Catalog Data Release 4 (4FGL-DR4)". In: arXiv:2307.12546 (2023). DOI: 10.48550/arXiv.2307.12546. arXiv: 2307.12546 [astro-ph.HE].

[von+20] A. von Kienlin et al. “The Fourth Fermi-GBM Gamma-Ray Burst Catalog: A Decade of Data”. In: *The Astrophzsical Journal* 893.1, 46 (Apr. 2020), p. 46. DOI: 10.3847/1538-4357/ab7a18. arXiv: 2002.11460 [astro-ph.HE].

[Bur+23] E. Burns et al. “GRB 221009A: The BOAT”. In: *The Astrophysical Journal Letters* 946.1 (2023), p. L31. DOI: 10.3847/2041-8213/acc39c.

[Tom+23] J. Tomsick et al. “The Compton Spectrometer and Imager”. In: *PoS* ICRC2023 (2023), p. 745. DOI: 10.22323/1.444.0745.

[Bee+22] J. Beechert et al. “Measurement of Galactic 26Al with the Compton Spectrometer and Imager”. In: *The Astrophysical Journal* 928.2 (2022), p. 119. DOI: 10.3847/1538-4357/ac56dc.

[Kie+20] C. A. Kierans et al. “Detection of the 511 keV Galactic Positron Annihilation Line with COSI”. In: *The Astrophysical Journal* 895.1 (2020), p. 44. DOI: 10.3847/1538-4357/ab89a9.

[Sie+20] T. Siegert et al. “Imaging the 511 keV Positron Annihilation Sky with COSI”. In: *The Astrophysical Journal* 897.1 (2020), p. 45. DOI: 10.3847/1538-4357/ab9607.

[Kar+23] C. M. Karwin et al. “Probing the Galactic Diffuse Continuum Emission with COSI”. In: *The Astrophysical Journal* 959.2 (2023), p. 90. DOI: 10.3847/1538-4357/ad04df.

[Rob+24] J. Roberts et al. “Imaging and Spectral Fitting of Bright Gamma-ray Sources with the COSI Balloon Payload”. In: (Dec. 2024). DOI: 10.48550/arXiv.2412.04721.

[Ber+25] D. Berge et al. “newASTROGAM – The New MeV to GeV Gamma-ray Observatory”. In: arXiv:2507.08133 (July 2025), arXiv:2507.08133. DOI: 10.48550/arXiv.2507.08133. arXiv: 2507.08133 [astro-ph.IM].

[RN21] S. Russell and P. Norvig. *Artificial Intelligence, Global Edition A Modern Approach*. Pearson Deutschland, 2021, p. 1168. ISBN: 9781292401133.

[Ang+25] J. R. Angevaare et al. *XENONnT/straxen: v3.2.3*. Version v3.2.3. July 2025. DOI: 10.5281/zenodo.16278327.

[Nuz15] R. Nuzzo. “How scientists fool themselves – and how they can stop”. In: *Nature* 526.7572 (2015), pp. 182–185. ISSN: 1476-4687. DOI: 10.1038/526182a.

[Cor+22] A. S. Cornell et al. “Boosted decision trees in the era of new physics: a smuon analysis case study”. In: *Journal of High Energy Physics* 2022.4 (2022), p. 15. ISSN: 1029-8479. DOI: 10.1007/JHEP04(2022)015.

[WZ07] M. Wolter and A. Zemla. “Implementation of the SVM algorithm for high energy physics data analysis”. In: *Ann. UMCS Informatica* 6.1 (2007), pp. 149–160.

[Kro+21] B. Kronheim et al. "Bayesian neural networks for fast SUSY predictions". In: *Physics Letters B* 813 (2021), p. 136041. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2020.136041.

[Mit97] T. M. Mitchell. *Machine learning*. Vol. 1. 9. McGraw-hill New York, 1997.

[Ped+11] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[GBC16] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[Jen+23] J. Jenis et al. "Engineering Applications of Artificial Intelligence in Mechanical Design and Optimization". In: *Machines* 11.6 (2023). ISSN: 2075-1702. DOI: 10.3390/machines11060577.

[BB23] C. M. Bishop and H. Bishop. *Deep Learning - Foundations and Concepts*. Ed. by S. Cham. 1st ed. 2023. ISBN: 978-3-031-45468-4. DOI: https://doi.org/10.1007/978-3-031-45468-4.

[SMT23] L. Sanchez, E. Merényi, and C. D. Tunnell. "SOM-based Classification and a Novel Stopping Criterion for Astroparticle Applications." In: *ESANN*. 2023.

[Fin+21] T. Finke et al. "Autoencoders for unsupervised anomaly detection in high energy physics". In: *Journal of High Energy Physics* 2021.6 (2021), p. 161. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2021)161.

[Dro+24] Drozdova, M. et al. "Radio-astronomical image reconstruction with a conditional denoising diffusion model". In: *Astronomy and Astrophysics* 683 (2024), A105. DOI: 10.1051/0004-6361/202347948.

[MP43] W. Mcculloch and W. Pitts. "A Logical Calculus of Ideas Immanent in Nervous Activity". In: *Bulletin of Mathematical Biophysics* 5 (1943), pp. 127–147.

[Ros58] F. Rosenblatt. "The perceptron: A probabilistic model for information storage and organization in the brain [J]". In: *Psychol. Review* 65 (Dec. 1958), pp. 386 –408. DOI: 10.1037/h0042519.

[Agg18] C. Aggarwal. *Neural Networks and Deep Learning: A Textbook*. Springer, Jan. 2018. ISBN: 978-3-319-94462-3. DOI: 10.1007/978-3-319-94463-0.

[Sej18] T. Sejnowski. *The Deep Learning Revolution*. Oct. 2018. ISBN: 9780262346825. DOI: 10.7551/mitpress/11474.001.0001.

[HSW89] K. Hornik, M. Stinchcombe, and H. White. "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5 (1989), pp. 359–366. ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(89)90020-8.

[KK24] V. Kunc and J. Kléma. *Three Decades of Activations: A Comprehensive Survey of 400 Activation Functions for Neural Networks*. 2024. arXiv: 2402.09092 [cs.LG].

[VD18] S. Varma and S. Das. *Deep Learning*. https://srdas.github.io/DLBook/. 2018.

[Zei12] M. D. Zeiler. "Adadelta: an adaptive learning rate method". In: *arXiv preprint arXiv:1212.5701* (2012).

[Tie12] T. Tieleman. "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude". In: *COURSERA: Neural networks for machine learning* 4.2 (2012), p. 26.

[KB17] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[RHW86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323.6088 (1986), pp. 533–536. ISSN: 1476-4687. DOI: 10.1038/323533a0.

[Bay+17] A. G. Baydin et al. "Automatic differentiation in machine learning: a survey". In: *J. Mach. Learn. Res.* 18.1 (Jan. 2017), 5595–5637. ISSN: 1532-4435.

[Yam+90] K. Yamaguchi et al. "A neural network for speaker-independent isolated word recognition". Proc. First International Conference on Spoken Language Processing (ICSLP 1990). 1990. DOI: 10.21437/ICSLP.1990-282.

[SSC24] Y. Sun, Z. Sun, and W. Chen. "The evolution of object detection methods". In: *Engineering Applications of Artificial Intelligence* 133 (2024), p. 108458. ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2024.108458.

[Bra+25] K. K. Brar et al. "Image segmentation review: Theoretical background and recent advances". In: *Information Fusion* 114 (2025), p. 102608. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2024.102608.

[Win+03b] C. Winkler et al. "The INTEGRAL mission". In: *Astronomy and Astrophysics* 411 (Nov. 2003), pp. L1–L6. DOI: 10.1051/0004-6361:20031288.

[Zog06] A. C. Zoglauer. "First light for the next generation of Compton and pair telescopes : Development of new techniques for the data analysis of combined Compton and pair telescopes and their application to the MEGA prototype". PhD thesis. PhD Thesis, Garching: Max-Planck-Institut für Extraterrestrische Physik, 2006, MPE Report, No. 289, 2006.

[Mor08] N Moriya. "Noise-related multivariant optimal joint-analysis in longitudinal". In: *Progress in applied mathematical modeling* (2008), p. 223.

[Zog] A. Zoglauer. personal conversation at AHEAD WP9.

[Bog06] S. E. Boggs. "The Advanced Compton Telescope mission". In: *New Astronomy Reviews* 50.7-8 (Oct. 2006), pp. 604–607. DOI: 10.1016/j.newar.2006.06.076. arXiv: astro-ph/0608532 [astro-ph].

[Leo12] W. Leo. *Techniques for Nuclear and Particle Physics Experiments: A How-to Approach*. Springer Berlin Heidelberg, 2012. ISBN: 9783642579202.

[Fan47] U. Fano. "Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions". In: *Phys. Rev.* 72 (1 1947), pp. 26–29. DOI: 10.1103/PhysRev.72.26.

[Bre+84] L. Breiman et al. *Classification and Regression Trees*. Taylor & Francis, 1984. ISBN: 9780412048418.

[Sal94] S. L. Salzberg. "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993". In: *Machine Learning* 16.3 (1994), pp. 235–240. ISSN: 1573-0565. DOI: 10.1007/BF00993309.

[Tan+19] P.-N. Tan et al. *Introduction to Data Mining Global Edition*. Pearson Deutschland, 2019, p. 864. ISBN: 9780273769224.

[Gin55] C. Gini. *Memorie di metodologia statistica*. Vol. 1. Libr. goliardica., 1955.

[Sha48] C. E. Shannon. "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[Zho+15] B. Zhou et al. "Learning Deep Features for Discriminative Localization". In: *CoRR* abs/1512.04150 (2015). arXiv: 1512.04150.

[Sel+16] R. R. Selvaraju et al. "Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization". In: *CoRR* abs/1610.02391 (2016). arXiv: 1610.02391.

[SVZ14] K. Simonyan, A. Vedaldi, and A. Zisserman. *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*. 2014. arXiv: 1312.6034 [cs.CV].

[Bre01] L. Breiman. "Random Forests". In: *Machine Learning* 45.1 (2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324.

[Aur+16] A. Aurisano et al. "A convolutional neural network neutrino event classifier". In: *Journal of Instrumentation* 11.09 (2016), P09001–P09001. ISSN: 1748-0221. DOI: 10.1088/1748-0221/11/09/p09001.

[Sze+14] C. Szegedy et al. "Going Deeper with Convolutions". In: *arXiv e-prints*, arXiv:1409.4842 (Sept. 2014), arXiv:1409.4842. DOI: 10.48550/arXiv.1409.4842. arXiv: 1409.4842 [cs.CV].

[Ian+16] F. N. Iandola et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size". In: *CoRR* abs/1602.07360 (2016). arXiv: 1602.07360.

[Gsc16] D. Gschwend. "ZynqNet: An FPGA-Accelerated Embedded Convolutional Neural Network". Aug. 2016.

[He+14] K. He et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition". In: *CoRR* abs/1406.4729 (2014). arXiv: 1406.4729.

[Sri+14] N. Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15 (June 2014), pp. 1929–1958.

[Aba+15] M. Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015.

[Hen17] Y. Henon. *keras-spp*. https://github.com/yhenon/keras-spp. 2017.

[Sif14] L. Sifre. "Rigid-motion scattering for image classification". PhD Thesis. Ecole polytechnique, CMAP, 2014.

[Kos+20] L. Kosmidis et al. "GPU4S: Embedded GPUs in space - Latest project updates". In: *Microprocessors and Microsystems* 77 (2020), p. 103143. ISSN: 0141-9331. DOI: https://doi.org/10.1016/j.micpro.2020.103143.

[Ort+23] F. Ortiz et al. "Onboard Processing in Satellite Communications Using AI Accelerators". In: *Aerospace* 10.2 (2023). ISSN: 2226-4310. DOI: 10.3390/aerospace10020101.

[HL13] D. W. Hosmer and S. Lemeshow. *Applied logistic regression*. 3rd ed. Hoboken: Wiley-Interscience Publication, 2013.

[JR09] E. Jourdain and J. P. Roques. "The High-Energy Emission of the Crab Nebula from 20 keV TO 6 MeV with Integral SPI". In: *apj* 704.1 (Oct. 2009), pp. 17–24. DOI: 10.1088/0004-637X/704/1/17. arXiv: 0909.3437 [astro-ph.HE].

[Rev22] C. D. S. Rev. *CubeSat Design Specification, rev. 14.1*. 2022.

[Bus+15] S. Busch et al. "UWE-3, In-Orbit Performance and Lessons Learned of a Modular and Flexible Satellite Bus for Future Pico-Satellite Formations". In: *Acta Astronautica* 117 (Aug. 2015). DOI: 10.1016/j.actaastro.2015.08.002.

[KBS20] A. Kramer, P. Bangert, and K. Schilling. "UWE-4: First Electric Propulsion on a 1U CubeSat—In-Orbit Experiments and Characterization". In: *Aerospace* 7.7 (2020). ISSN: 2226-4310. DOI: 10.3390/aerospace7070098.

[BT18] D. Bock and M. Tajmar. "Highly miniaturized FEEP propulsion system (NanoFEEP) for attitude and orbit control of CubeSats". In: *Acta Astronautica* 144 (2018), pp. 422–428. ISSN: 0094-5765. DOI: https://doi.org/10.1016/j.actaastro.2018.01.012.

[Mye+24] A. Myers et al. "BurstCube: A CubeSat for Gravitational Wave Counterparts". In: Aug. 2024. DOI: 10.1117/12.3020456.

[Lav+21] A. Laviron et al. "COMCUBE: A constellation of CubeSats to measure the GRB prompt emission polarization". In: *SF2A-2021: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*. Ed. by A. Siebert et al. Dec. 2021, pp. 105–108.

[Kom+17] S. Komura et al. "Imaging Polarimeter for a Sub-MeV Gamma-Ray All-sky Survey Using an Electron-tracking Compton Camera". In: *The Astrophysical Journal* 839.1 (2017), p. 41. DOI: 10.3847/1538-4357/aa68dc.

[WEO10] M. C. Weisskopf, R. F. Elsner, and S. L. O'Dell. "On understanding the figures of merit for detection and measurement of x-ray polarization". In: *Space Telescopes and Instrumentation 2010: Ultraviolet to Gamma Ray*. Ed. by M. Arnaud, S. S. Murray, and T. Takahashi. Vol. 7732. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July 2010, 77320E, 77320E. DOI: 10.1117/12.857357. arXiv: 1006.3711 [astro-ph.IM].

[Uly20] A. Ulyanov. "COMCUBE: GRB polarization sensitivity". Coimbra AHEAD 2020 Progress Meeting on Space Experiments for HE Astrophysics Multi-messenger Astronomy. 2020.

[Fra+25] N. Franel et al. "Gamma-Ray Burst Polarimetry with the COMCUBE-S CubeSat Swarm". In: *arXiv e-prints*, arXiv:2510.24549 (Oct. 2025), arXiv:2510.24549. DOI: 10.48550/arXiv.2510.24549. arXiv: 2510.24549 [astro-ph.HE].

[Leb+18] S. Leblond et al. "MAUD Project - development of a new portable detector for alpha and beta surface contamination mapping". In: *DEM 2018 - Dismantling Challenges: Industrial Reality, Prospects and Feedback Experience*. Avignon, France, Oct. 2018.

[Ham] *MPPC S14160/14161*. Hamamatsu.

[Mur+22] D. Murphy et al. "A compact instrument for gamma-ray burst detection on a CubeSat platform II". In: *Experimental Astronomy* 53.3 (2022), pp. 961–990. ISSN: 1572-9508. DOI: 10.1007/s10686-022-09842-z.

[Mur18] F. J. T. J. e. a. Murphy D. "EIRSAT-1 - The Educational Irish Research Satellite." In: *2nd Symposium in Space Educational Activities*. 2018.

[Epi] *GAGG(Ce) Crystal*. Epic.

[Mur23] D. Murphy. "COMCUBE: A Compact Telescope in a 6U CubeSat". Advances in Space AstroParticle Physics, Perugia. 2023.

[Sat16] T. Sato. "Analytical Model for Estimating the Zenith Angle Dependence of Terrestrial Cosmic Ray Fluxes". In: *PLOS ONE* 11.8 (Aug. 2016), pp. 1–22. DOI: 10.1371/journal.pone.0160390.

[Nat92] C. G. National Data. "U.S. standard atmosphere (1976)". In: *Planetary and Space Science* 40.4 (Apr. 1992), pp. 553–554. DOI: 10.1016/0032-0633(92)90203-Z.

[Pic+08] A Picard et al. "Revised formula for the density of moist air (CIPM-2007)". In: *Metrologia* 45.2 (2008), p. 149. DOI: 10.1088/0026-1394/45/2/004.

[Ber10] M. Berger. "XCOM: photon cross sections database". In: *http://www.nist.gov/pml/data/xcom/index.cfm* (2010).

[Cat+23] V. Catoire et al. "SPECIES: a balloon-borne and airborne instrument coupling infrared lasers with Optical Feedback Cavity Enhanced Absorption Spectroscopy technique for atmospheric in-situ trace-gas measurements". In: *EGU General Assembly Conference Abstracts*. EGU General Assembly Conference Abstracts. May 2023, EGU-12978, EGU–12978. DOI: 10.5194/egusphere-egu23-12978.

[Sau+20] J. E. Saunders et al. "Canadian Atmospheric Laser Absorption Spectrometer Experiment Test-bed (CALASET): Field Measurements of CO_2 and N_2O with a Stratospheric Balloon-Borne Laser Spectrometer". In: *AGU Fall Meeting Abstracts*. Vol. 2020. AGU Fall Meeting Abstracts. Dec. 2020, P050-02, P050–02.

[Cne] *Strato Science 2023: Payload Gondola Overview, howpublished = internal communication*.

[Zog+09] A. Zoglauer et al. "Cosima — The cosmic simulator of MEGAlib". In: *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*. 2009, pp. 2053–2059. DOI: 10.1109/NSSMIC.2009.5402128.

[Owe+91] A. Owens et al. "The Transient Gamma-Ray Spectrometer". In: *IEEE Transactions on Nuclear Science* 38.2 (1991), pp. 559–567. DOI: 10.1109/23.289357.

[Gal+25] S. Gallego et al. "Bottom-up Background Simulations of the 2016 COSI Balloon Flight". In: *The Astrophysical Journal* 986.2 (2025), p. 116. DOI: 10.3847/1538-4357/add6a0.

[Dei+25] A. Deisting et al. *GRANITE: Mechanical Characterization and Optical Inspection of Large-Area TPC Electrodes*. Submitted to JINST. 2025. arXiv: 2511.11400 [physics.ins-det].

[Mit+25] S. A. Mitra et al. *GRANITE: High-Resolution Imaging and Electrical Qualification of Large-Area TPC Electrodes*. Submitted to JINST. 2025. arXiv: 2511.11401 [physics.ins-det].

[Bec21] B. H.-L. Beckmann. "Entfaltung von PMT-Signalen mit eindimensionalen Convolutional Neural Networks (CNNs)". Available at https://cms.zdv.uni-mainz.de/fb08-xenon-physik/wp-content/uploads/sites/107/2025/11/oberlack-group-thesis_2021_BSc_Benjamin-Beckmann.pdf. Bachelor's thesis. Johannes Gutenberg-Universität Mainz, 2021.

[Mer24] J. Merz. "Event Reconstruction with Machine Learning Techniques". Available at https://cms.zdv.uni-mainz.de/fb08-xenon-physik/wp-content/uploads/sites/107/2025/07/oberlack-group-thesis_2024_MSc_Merz-Johannes.pdf. Master's thesis. Johannes Gutenberg-Universität Mainz, 2024.

[Blo+22] P. Bloser et al. "CubeSats for Gamma-Ray Astronomy". In: *Handbook of X-ray and Gamma-ray Astrophysics*. Ed. by C. Bambi and A. Sangangelo. 2022, 19, p. 19. DOI: 10.1007/978-981-16-4544-0_53-1.

[Str+24] N. Striebig et al. "AstroPix4 — a novel HV-CMOS sensor developed for space based experiments". In: *Journal of Instrumentation* 19.04 (2024), p. C04010. DOI: 10.1088/1748-0221/19/04/C04010.