UNIVERSITÀ DI SIENA 1240

Department of Physical Sciences, Earth and Environment

PhD in Experimental Physics

XXXIV Cycle

Coordinator: Prof. Riccardo Paoletti

Event classification in MAGIC through Convolutional Neural Networks

Disciplinary Scientific Sector: FIS/01

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

PhD Student:
Stefano Truzzi

DocuSigned by:
Stefano Truzzi
CEEB9ECE025A46B...

Supervisor:
Prof. Riccardo Paoletti

Co-Supervisors:
Dott. Giacomo Bonnoli

Dott. Roberto Cappuccio

AA 2021/2022

Abstract

The Major Atmospheric Gamma Imaging Cherenkov (MAGIC) telescopes are able to detect gamma rays from the ground with energies beyond several tens of GeV emitted by the most energetic known objects, including Pulsar Wind Nebulae, Active Galactic Nuclei, and Gamma-Ray Bursts.

Gamma rays and cosmic rays are detected by imaging the Cherenkov light produced by the charged superluminal leptons in the extended air shower originated when the primary particle interacts with the atmosphere. These Cherenkov flashes brighten the night sky for short times in the nanosecond scale. From the image topology and other observables, gamma rays can be separated from the unwanted cosmic rays, and thereafter incoming direction and energy of the primary gamma rays can be reconstructed.

The standard algorithm in MAGIC data analysis for the gamma/hadron separation is the so-called Random Forest, that works on a parametrization of the stereo events based on the shower image parameters. Until a few years ago, these algorithms were limited by the computational resources but modern devices, such as GPUs, make it possible to work efficiently on the pixel maps information. Most neural network applications in the field perform the training on Monte Carlo simulated data for the gamma-ray sample. This choice is prone to systematics arising from discrepancies between observational data and simulations.

Instead, in this thesis I trained a known neural network scheme with observation data from a giant flare of the bright TeV blazar Mrk421 observed by MAGIC in 2013. With this method for gamma/hadron separation, the preliminary results compete with the standard MAGIC analysis based on Random Forest classification, which also shows the potential of this approach for further improvement.

In this thesis first an introduction to the High-Energy Astrophysics and the Astroparticle physics is given. The cosmic messengers are briefly reviewed, with a focus on the photons, then astronomical sources of γ rays are described, followed by a description of the detection techniques.

In the second chapter the MAGIC analysis pipeline starting from the low level data acquisition to the high level data is described. The MAGIC Instrument Response Functions are detailed. Finally, the most important astronomical sources

used in the standard MAGIC analysis are listed.

The third chapter is devoted to Deep Neural Network techniques, starting from an historical Artificial Intelligence excursus followed by a Machine Learning description. The basic principles behind an Artificial Neural Network and the Convolutional Neural Network used for this work are explained.

Last chapter describes my original work, showing in detail the data selection/manipulation for training the Inception Resnet V2 Convolutional Neural Network and the preliminary results obtained from four test sources.

Contents

Chapter 1

High Energy Astrophysics and Astroparticle Physics

1.1 Astroparticle physics and Cosmic Messengers

The space is permeated by particles of different kind (protons, nuclei, photons, neutrinos, ...) generated in astrophysical sources like stars, novae, supernovae, nebulae, galaxies, pulsars, quasars, neutron star, black holes and their interactions and mergers. Astrophysics and Astroparticle physics study astronomical events and sources through the detection of particles and the electromagnetic waves generated by these events: in general High Energy Astrophysics considers the electromagnetic waves while Astroparticle physics studies the particles.

The particles that reach the Earth from the space can be summarized in three types: (fig.1.1):

- Cosmic Rays
- Photons
- Neutrinos

These three messengers together with the recently detected gravitional waves, are the cosmic messengers that the Astroparticle/Astrophysics consider to study the phenomena (see e.g. Mészáros et al., 2019). These messengers will be introduced in this chapter with a specific focus on photons, the most relevant messengers for the thesis, that deals with an Imaging Atmospheric Cherenkov Telescope (IACT) that is able to detect gamma rays (γ rays) the most energetic part of electromagnetic spectrum.

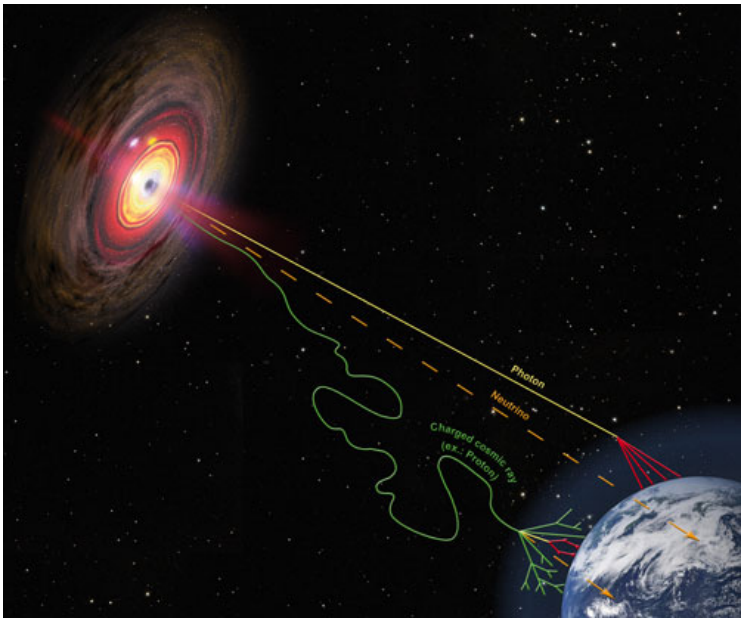


Figure 1.1: Cosmic messengers (credit: HAP/A. Chantelauze)

1.1.1 Cosmic Rays

Cosmic Rays (CR) are high energetic protons and nuclei (>99% of the total particle budget), electrons and positrons (<1%) travelling at relativistic speed that arrive on the Earth from the space. They are generated by the Sun, by other sources within the Milky Way or by extra-galactic sources. Since CR are charged particles they are influenced by the cosmic magnetic fields and for that reason they do not travel along straight trajectories from the sources to the Earth. This makes the task of connecting CR to their celestial accelerators a non trivial one even for CR at the highest energies (see e.g. Pierre Auger Collaboration et al., 2017)).

CR can be divided into two main classes: primary CR, that are the particles generated directly from the sources and secondary CR, that are particles generated by an interaction of primary CR with the environment along the path, for example the Earth atmosphere.

Cosmic Rays Discovery

CR were discovered by Victor Hess in 1912 (see e.g. Hess (2018) for a commented modern edition). In early 1900 after the discovery of the natural radioactivity by Henri Becquerel (1896) it was believed that CR originated from the rocks. With this assumption moving away from the Earth surface would have implied the lowering of the ionization level induced by this penetrating radiation. In 1909 Theodor

Wulf measured the rate of ionization on the top of the Eiffel tower using a portable electroscope, designed by him, and he found that the decrement with height was less than expected. Afterwards Domenico Pacini (Pacini (1912)) performed in 1911 underwater measurements and he found a decrease of the radioactivity underwater. He concluded that the radiation should come from sources located outside the Earth, so anticipating the result commonly attributed to V. Hess.
In 1911 Hess performed a balloon flight that reached an altitude of about 1100 m, and found no "significant changes" in the amount of radiation with respect to the measurements on the ground. Later, in 1912, Hess flew another balloon up to 5300 m during an almost total solar eclipse. Since he didn't measure an effect of the eclipse on the ionization of the atmosphere, he hypothesized that the radiation did not come from the Sun. Therefore the origin of CR had to be located further out in the space.

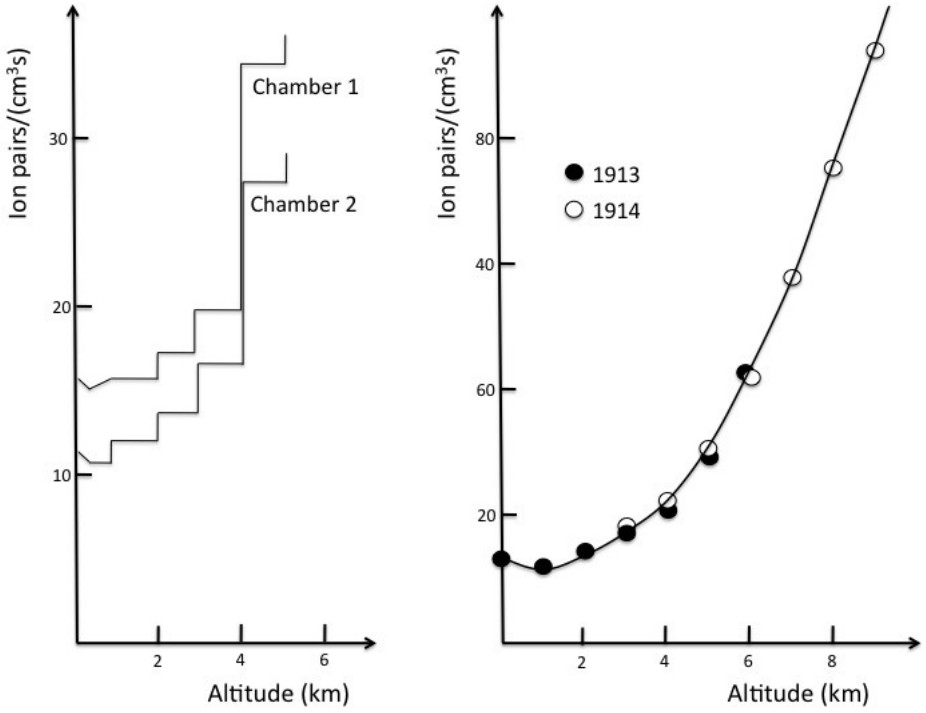


Figure 1.2: Increase of ionization with altitude as measured by Hess in 1912 (left panel) and by Kolhörster 1913-14 (right panel). From De Angelis (2012).

Cosmic Rays Spectrum

The CR spectrum (plotted in Fig. 1.3) covers many decades in energy . It extends to energies beyond $E = 10^{20}$ eV, with particles reaching energies that are more than 10^8 times larger than those achievable in the most powerful terrestrial accelerators such as CERN-LHC. The differential flux as a function of the energy can be approximated with a power law:

$$\Phi(E) \propto E^{-\alpha} \tag{1.1}$$

positive α means that the particle flux decreases significantly with increasing energy.

Some energy intervals related to changes in the slope can be noted, and this feature points to the contribution of different components. In particular:

- $10^9\text{eV} \lesssim E \lesssim 10^{10}\,\text{eV}$. In this region CR are mainly solar.
- $10^{10}\text{eV} \lesssim E \lesssim 10^{15}\,\text{eV}$. In this region the CR origin is generally attributed to galactic sources. The slope of the spectrum in this region is $\alpha \sim 2.7$.
- $5 \times 10^{15}\text{eV} \lesssim E \lesssim 5 \times 10^{18}\,\text{eV}$. In this region the spectrum follows a steeper slope ($\alpha \sim 3.1$). The transition region at $E \sim 3 - 5 \times 10^{15}\,\text{eV}$ is called the "knee". The slope change is widely attributed to the galactic/extra-galactic transition, meaning that in this region a component of CR generated by extra-galactic sources starts to contribute significantly to the total flux.
- Above $E \sim 5 \times 10^{18}$ eV the slope hardens again ($\alpha \sim 2.6$). This region is widely known as the "ankle". At $E \sim 5 \times 10^{19}$ a cutoff appears that is commonly interpreted as the GZK cutoff.

Cosmic ray Acceleration

Which are the mechanisms that can accelerate the particles at those energies? Currently the most accredited acceleration mechanisms are called Fermi I and II and will be mentioned in the two following subsections.

Fermi Second Order Acceleration (or Fermi II)

The Fermi II acceleration was the first, in chronological order (Fermi (1949)), theoretical explanation proposed for the acceleration/formation of the CRs. Fermi theorized the presence of magnetic clouds (modelled as magnetic mirrors) that move randomly through the interstellar medium. During the motion the particles

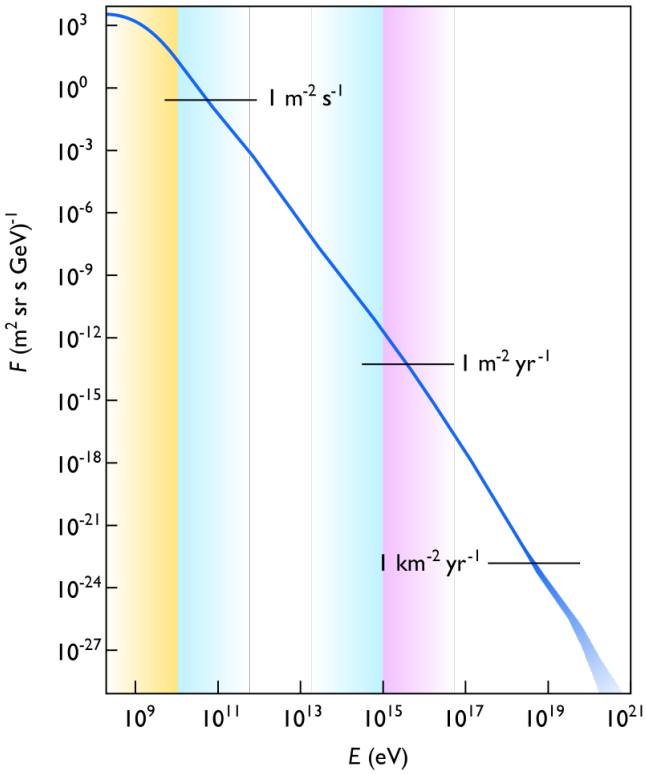


Figure 1.3: CR flux as a function of energy arriving on Earth (Swordy (2001))

and the mirror collide. When a particle hits the magnetic mirror that is moving towards the particle, the particle gains energy in the reflection, vice versa the particles that travel in the same direction of the clouds lose energy in the collision. Fermi proposed that the collision probability favors energy gains over energy losses, meaning that on the average the particles will be accelerated. The process name (second order) is due to the dependence of the rate of energy gain on the squared mirror speed:

$$\frac{\Delta E}{\Delta t} = (\frac{8U_c^2}{lv_p})E \tag{1.2}$$

where U_c is the speed of the cloud, v_p is the speed of the particle, l is the particle mean free path and E is the energy before the collision. The limit of this model is that the speed of the clouds in the galaxy is small ($v/c \leq 10^{-4}$, $U_c \leq 3*10^6$ cm/s). For this reason the Fermi II process is not considered capable to accelerate the particles at the very high energies.

Fermi First Order Acceleration (or Fermi I)

The Fermi first order (Fermi I) acceleration process is more efficient than Fermi II. It is modelled taking into account two considerations: the first one (same as Fermi II) is that the clouds collide with the particles and the second is that the clouds approach each other. In that way the particles bounce back and forth from one cloud to the other more and more energetically. This "bouncing" phenomenon can be produced in the shock waves (for example the waves produced in Supernova Remnant (SNR)) where the particles can cross the shock front forward and backward gaining energy in both directions. In this case the formula for the energy gain will be:

$$\frac{\Delta E}{\Delta t} \simeq 4\frac{U_c}{l}E \tag{1.3}$$

Again U_c is the speed of the clouds, l is the particle mean free path and E is the energy before the collision. Comparing with Eq. 1.2 there is not the particle speed v_p at denominator, so the process is more energy efficient than Eq. 1.2. A problem behind this model regards the injection of the particles, because only the particles that have overcome the thermal energy are able to cross the shock front and start the "bouncing" acceleration. Today the mechanism that allows particles to reach very high energies is not clear and remains under investigation.

The Larmor Radius

The Larmor radius it used to establish if the coming particle arrive from the Milky Way or from an extra-galactic source. Starting from the Lorentz force for a magnetic field B on a particle of charge q, mass m, speed v and Lorentz factor Γ (Fig.1.4):

$$\Gamma m\frac{d\boldsymbol{v}}{dt} = \frac{q}{c}\boldsymbol{v} \times \boldsymbol{B} \tag{1.4}$$

Considering that the acceleration is normal to both B and v. With B uniform and static, the orbit is a circle around B (see the Fig. 1.4).

Integrating Eq. 1.4:

$$\boldsymbol{v} = \boldsymbol{r_L} \times \frac{q\boldsymbol{B}}{\Gamma mc} \tag{1.5}$$

$$\boldsymbol{\omega_l} = \frac{2\pi}{T_L} = \frac{qB}{\Gamma mc} \qquad \boldsymbol{v} = \boldsymbol{r} \times \boldsymbol{\omega_l}$$

Where r_L is the radius of the orbit, T_L is the orbital period. ω_L is the circular frequency of motion and the Larmor radius for a particle with charge Ze (where e

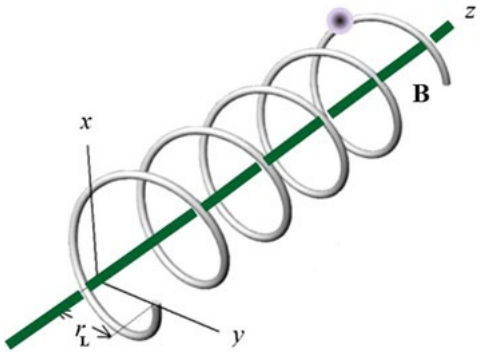


Figure 1.4: Trajectory of charged particle in an uniform magnetic field (B) oriented along $\hat{z}$

is the electron charge and Z is the atomic number) is given by:

$$r_L = \frac{v}{\omega_L} = \frac{\Gamma mvc}{ZeB} = \frac{pc}{ZeB} \approx \frac{E}{ZeB} \tag{1.6}$$

The last substitution holds in the relativistic limit, suitable for the majority of the CR.

Now consider that the galaxy is filled with an average magnetic field of intensity B $\sim$ 4 µG and using Eq. 1.6 for particle with energy E:

$$r_L = (E = 10^{12} eV) \cong 10^{15} cm = 3 \times 10^{-4}\, pc$$
$$r_L = (E = 10^{15} eV) \cong 10^{18} cm = 0.3\, pc$$
$$r_L = (E = 10^{18} eV) \cong 10^{21} cm = 300\, pc$$

If one compare the pc scale with the Galaxy dimension [1] it can be inferred that all the particles with E $\leq 10^{18}$ eV are confined in the Milky Way by its magnetic field.

The GZK effect and the Energy Cutoff

The Greisen–Zatsepin–Kuzmin limit (GZK) limit is a proposed cutoff for the CR (fig 1.3) energing at $\geq 5 \times 10^{19}$ for protons. This limit is due to the interaction of protons with the cosmic microwave background (CMB) throught the Δ resonance:

$$p + \gamma_{CMB} \rightarrow \Delta^+ \rightarrow p + \pi^0$$
$$p + \gamma_{CMB} \rightarrow \Delta^+ \rightarrow n + \pi^+$$

[1] the Milky Way can be approximated as a cilinder with radius $\sim$ 150-200 pc and half height $\sim$ 15 pc

Given the high density of CMB the mean free path for CR protons of the highest energy is rather short. Therefore the maximum distances traveled by the CR are constrained within about UHECR (defined in table 1.2) at 50 Mpc.

An experimental confirmation of the GZK effect is one of the main scientific are being under investigation by HiRes and Auger collaborations (Bahcall and Waxman (2003)).

1.1.2 Photons

Photons are the elementary particles of the electromagnetic radiation. They are electrically neutral, therefore they can travel straight through the cosmic space without interacting with magnetic fields. This property is very important because differently from charged CR that are deflected by cosmic magnetic fields, one can trace back photons to their sources once the incoming direction is determined.

Electromagnetic spectrum

The Electromagnetic (EM) spectrum represent all the light waves: from radio waves to gamma-ray waves including the visible light and many other classes. Fig. 1.5 shows the whole EM spectrum. It covers many decades in energy and includes waves with very different properties. Here we will focus EM γ rays with energies from $\sim$ 30 GeV up to several TeVs. This band is defined as the Very High Energy (VHE) gamma-ray band and it is where Imaging Atmospheric Cherenkov Telescopes (like MAGIC 2) operate.

Band	Wavelength (λ)	Frequency (ν)	Energy (E)
	[m]	[Hz]	[eV]
Radio	$\geq$ 1m	$\leq$ 300MHz	$\leq$ 1.24 μeV
MicroWave	1m - 1mm	300MHz - 300GHz	1.24μeV - 1.24meV
Infrared	1mm - 750nm	300GHz - 400THz	1.24meV - 1.7eV
Visible	750nm - 400nm	400THz - 750THz	1.7eV - 3eV
UltraViolet	400nm - 10nm	750THz - 30PHz	3eV -124 eV
X-ray	10nm - 10pm	30PHz - 30EHz	124eV - 124KeV
Gamma-ray	$\leq$ 10pm	$\geq$30EHz	$\geq$ 124KeV

Table 1.1: Classification of the electromagnetic spectrum in bands as function of wavelength (λ), frequency (ν) and photon energy (E) from Lide (2004)

Detection techniques for photons depend on the energy for many reasons: for instance the interaction process involved in the detection can be photoelectric process, Compton or pair production depending on photon energy. Also the opacity

of the Earth's atmosphere is crucial: the atmosphere absorbs photons in most of the EM spectrum (Fig. 1.5) and only the visible light, the near infrared (IR) and a wide portion of the radio spectrum (described in sec: 1.1.2) are transparent. In all the other bands energy regions detection must be either from a satellite outside the atmosphere, or performed from the ground but by means of an indirect method.

In this work we will focus on this messenger, in particular we will treat photons in the gamma-ray domain (see table: 1.2).

The energy range of the gamma-ray domain covers many decades; for this reason is is commonly divided in narrower intervals. This further division is mainly related to the different detection techniques that are most suitable in each band. For instance space-borne Compton detectors are most effective in the LE band, space-borne pair conversion detectors in the HE band, and different kinds of ground detectors are the best instruments in the bands at higher energies (VHE, UHE and EHE). In table 1.2 is reported the classification of the gamma-ray spectrum.

Band	Abbreviation	Energy range
Low Energy	LE	100 keV - 30 MeV
High Energy	HE	30 MeV - 30 GeV
Very-High Energy	VHE	30 - GeV 30 TeV
Ultra-High Energy	UHE	30 TeV - 30 PeV
Extremely-High Energy	EHE	$\geq$ 30 PeV

Table 1.2: gamma-ray electromagnetic spectrum description. This table can be considered an extension of table 1.1. The spectrum is not contiguous with respect table: 1.1 because the boundary between the X-ray band and gamma-ray band is not defined sharply. The lower limit of gamma-ray band is taken from: Kanbach, Schonfelder, and Zehnder (2010)

Production of gamma rays

The main processes that produce γ rays in astrophysical environments are in general divided in two broader classes: leptonic processes and hadronic processes. In leptonic processes photons are generated by leptons, most commonly electrons or positrons. The following list summarizes these processes:

- **Bremsstrahlung**: when a charged particle moves through an electric field, for example an electron interacting with an atomic nucleus, part of its kinetic energy is converted into electromagnetic radiation.
- **Electron synchrotron emission**: An electron accelerated in a magnetic

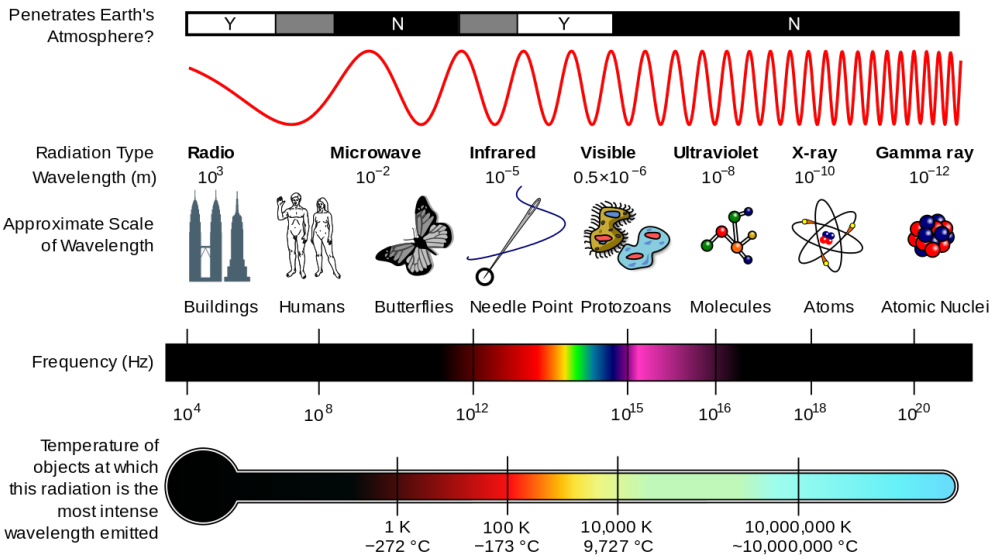


Figure 1.5: Electromagnetic spectrum diagram showing various properties: atmosphere opacity, scale, radiation type, wavelength, frequency, temperature. Credit by NASA

field generates photons.

$$e^{\pm} + \vec{B} \rightarrow \gamma \tag{1.7}$$

- **Inverse Compton (IC) scattering**: a process in which a photon gains energy from the collision with a high energy electron. This mechanisms is important in regions where relativistic electrons coexist with a population of photons of adequate density.

$$e^{\pm} + \gamma \rightarrow e^{\pm} + \gamma \tag{1.8}$$

  In particular there is a specific kind of IC process where the same high energy electrons are responsible of both generating the photons (via synchrotron emission) and of scattering them ahead. This is the so-called "Synchrotron Self-Compton (SSC)" process and plays a mayor role in some contexts, for instance in leptonic models of blazar emission.

- **Electron positron annihilation**: this collision can generate a couple of photons (gamma photon).

The hadronic processes are those emission mechanisms processes where gamma-ray photons are generated by hadrons.. The main hadronic processes are of three

types. The first two are able to generate a neutral pion, which further decays into two photons:

$$\pi^0 \rightarrow \gamma\gamma \tag{1.9}$$

- **Photo-production**: During propagation, protons can interact with photons of the Cosmic Microwave Background (CMB). If the proton has enough energy a resonant hadron Δ_+ is produced only to decay immediately.

  $$p + \gamma_\epsilon \rightarrow \Delta_+ \tag{1.10}$$

  The resonant Δ_+ can decay in two branches:

  $$\Delta_+ \rightarrow \pi^0 p \tag{1.11}$$

  The π^0 is the particle that generates the photons.

  $$\Delta_+ \rightarrow \pi^+ n \tag{1.12}$$

  This second branch for the Δ_+ decay is important for the neutrinos emission because:

  $$\pi^+ \rightarrow \mu^+ \nu_\mu \quad \mu^+ \rightarrow e^+ \nu_e \overline{\nu}_\mu \tag{1.13}$$

- **Proton-proton inelastic collision:**

  $$p + p \rightarrow \pi^\pm, \pi^0, K\pm, K^0, p, n, ... \tag{1.14}$$

  As in the photo-production this effect can generate photons from π^0 (Eq. 1.9). Furthermore it can be very important for generating astrophysical neutrinos both for the π^+ (Eq. 1.13) and for the π^-:

  $$\pi^- \rightarrow \mu^- \, \overline{\nu}_\mu \quad \mu^- \rightarrow e^- \, \overline{\nu}_e \nu_\mu \tag{1.15}$$

- **Proton synchrotron emission**: Similar to electron synchrotron; due to the heavier mass, syncrotron emission from protons requires more intense magnetic fields to contribute significanly.

  $$p + \vec{B} \rightarrow \gamma \tag{1.16}$$

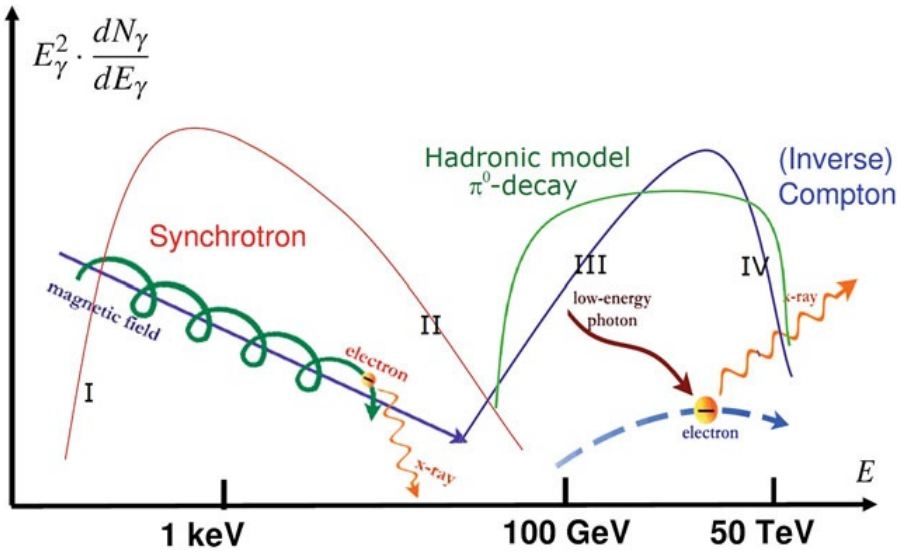


Figure 1.6: From Spurio (2015): sketch of a theoretical SED, Comparison of the prediction of leptonic and hadronic model for the high energy component.

Leptonic and Hadronic models

In the previous Sec. 1.1.2 all the main processes that can generate high energy photons are listed. These processes are used to infer the theoretical models for astrophysical sources.

One of the most powerful methods for challenging these models consists of merging the data from multi-wavelength observations obtained by different telescopes (e.g. optical, X-ray and gamma-ray data) in a single broad-band Spectral Energy Distribution (SED). Then in the framework of each model, the physical parameters of the emission model that fit the experimental data can be determined.

The models are divided in leptonic if the considered processes are based on leptons, hadronic if the processes are based on hadrons and lepto-hadronic if the considered processes include a mixture of leptons and hadrons.

In Fig. 1.6 a comparison sketch of a prediction of two competing emission models, one hadronig and one leptonic, is shown. While the low energy emission can be generally attributed to the electron synchrotron models, in the gamma-ray domain both hadronic models (e.g. π_0 decays) and leptonic models (e.g. inverse Compton scattering of relativistic electrons on low energy photons) can produce an emission bump in the SED, although with discrepancies between the respective spectral distributions. Accurate observations can in principle discriminate these models.

1.1.3 Neutrinos and Gravitational waves

Neutrino is a fundamental particle, with no electrical charge and very small mass that is a very abundant. Neutrinos can interact through weak interaction and gravity.

The neutrino interactions are very rare due to their very low cross section. A neutrino can cross the Earth without interacting and continue its path through the space, the mean free path is very long and this property makes it a good candidate to study astrophysical phenomena because it can travel straight directly from very far away (more than photons).

On the other side these properties makes the neutrino detection very problematic. These are the general reason why experimental physics in the field of neutrinos are less advanced compared to the gamma-rays. The neutrinos flux on the surface of Earth is shown in Fig: 3.1.

The neutrino study is very important to completely understand all the CR spectrum together with photons.

On September 14th 2015 the first gravitational wave was detected by Laser Interferometer Gravitational Wave Observatory (LIGO) (Abbott et al. (2016)). The gravitational waves are perturbations of the space-time caused by an acceleration of masses. This type of messenger is very different from the CR, photons and neutrinos. The GW open a very new way of studying the astrophysical events that together with the other messengers can be very useful to deepen our knowledge of the universe. For all these reasons these messengers are very important.

1.1.4 Multimessenger and Multiwavelength astronomy

The four messengers are used to study and understand the astrophysical events. All of them are important but combining together the scientific information coming from all of them is the most powerful tool to better understand and modeling the physical phenomena that are generated in the astrophysical object. This is so called multimessenger astronomy (Abbott et al. (2017)).

Today with the modern telescopes is possible to observe the same source in different ways, for example the same source can be observed with a radio telescope and with an optical telescope, when this is done with a set of telescopes that detect the same messenger the observation is called multiwavelength (fig: 1.8 credit from eCUIP University of Chicago). Furthermore is possible to observe the same object with two different messengers like for example TXS 0506+056 IceCube Collaboration et al. (2018) case, where from the same source both a neutrino and gamma ray was detected, in this case the observation is called multimessenger. The multiwavelength astronomy is widely used today meanwhile the multimessenger is at the beginning because observations of the neutrino and of the Gravitational

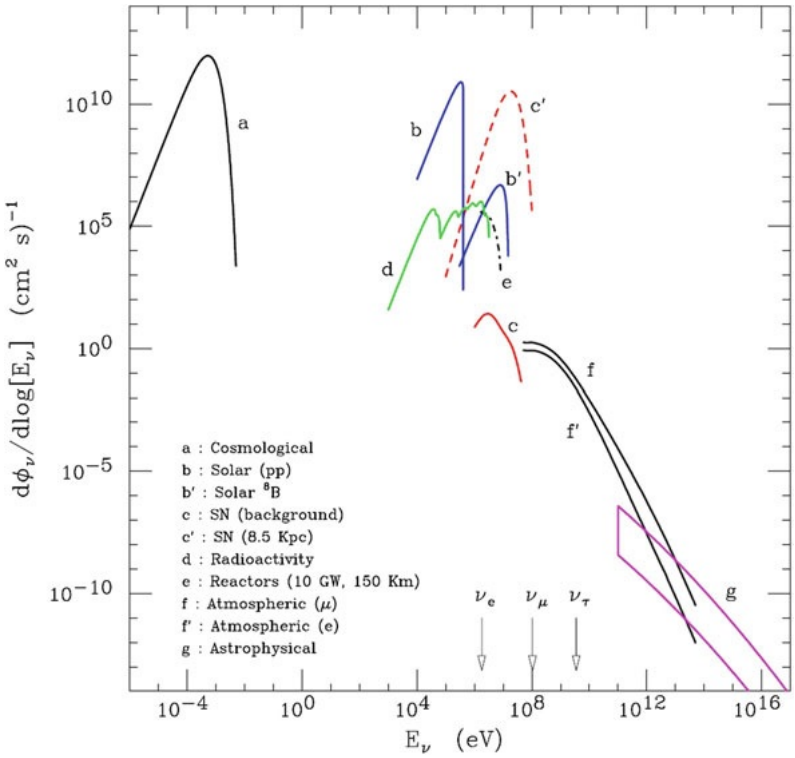


Figure 1.7: Neutrino's flux on the surface of Earth (Lipari (2006):): a are cosmological residual from Big Bang, b are the solar neutrinos, SN c and c' are neutrinos from supernova explosion, d are neutrinos generated from the Earth radioactive materials, e are neutrinos generated from nuclear plant, f and f' are the neutrinos generated by the interaction of CR with the atmosphere and g are generated by Astrophysical Active Galactic Nuclei (AGN), Supernova Remnants (SNR) and various other events

waves are still underdeveloped.

1.2 Astronomical Sources of gamma rays

The astrophysical sources of VHE gamma rays can be in general divided in Galactic (the sources that are observed within the Milky Way) and Extra-Galactic. Table 1.3 taken from Spurio (2015) shows the TeV sources included in the TeVCat[2] a catalogue comprising all the sources able to emit gamma-ray photons in the VHE band. In the following sections some of these sources will described in more detail.

[2]http://tevcat.uchicago. edu/

Figure 1.8: Multiwavelength pictures of Whirlpool Galaxy (M51A) (from `https://ecuip.lib.uchicago.edu/multiwavelength-astronomy/astrophysics/05.html`): Each image shows a narrow band of wavelengths of radiation across the electromagnetic spectrum. Low energy radiation comes from cool regions of molecular gas, and high energy radiation comes from hot spots where atoms are fully ionized. The combined information provides insight into the structure, temperature, and chemical composition of the Whirlpool Galaxy. The stars in the infrared image represent most of the mass of the galaxy, excluding dark matter. The optical image represents a slightly smaller amount of mass and the other three images represent only traces of mass in molecules (radio image) massive hot stars, (ultraviolet image) and hot plasma (X-ray image).

1.2.1 Pulsars and Pulsar wind nebulae

A pulsar is a neutron star (see e.g. Lorimer and Kramer, 2004), a star composed mainly of neutrons, with an enormous density $\rho \simeq 10^{14}\,\mathrm{g/cm^{-3}}$ and a high magnetic field ($B \simeq 10^{15}\,\mathrm{G}$), spinning very quickly (with period $P \sim 10^{-3}\,\mathrm{s}$ in some cases). The magnetic axis of the pulsar rotates together with the pulsar and in case the magnetic axis is pointed towards the Earth during the rotation one can observe the "lighthouse" effect, an effect that appears like a pulsed flash of light at extremely regular time intervals.

The Pulsar Wind Nebula (PWN) is a nebula where plasma is powered by a pulsar embedded in a supernova remnant. These two components are not spatially resolved by gamma-ray telescopes because of their arcmin-scale angular resolution. The first detection of a Pulsar Wind Nebula class with a ground-based Cherenkov telescope was obtained in 1989 by Whipple (Weekes et al., 1989), with the detection of the Crab Nebula, also the first VHE gamma-ray source detected ever.

Type	Designator	Objects	Example
Galactic			
Pulsar wind nebula	PWN	34	Crab, Geminga
1SNR + shell	Shell	16	RXJ1713, IC443
SNR + mol clouds	SNR/Mol. Cloud	11	W28, W51
Binary systems	Binary	11	
Massive star clust.	-	3	
Globular clust.	-	1	
Extragalactic			
HBL Blazar	HBL	55	Mrk421, Mrk501
IBL Blazar	IBL	10	BL-Lac , W Comae
9LBL Blazar	LBL	2	-
FSRQ Blazar	FSQR		3C279
FRI Blazar	FRI	4	-

Table 1.3: The known sources of VHE gamma rays included in TeVCat (http://tevcat.uchicago.edu/),divided by class. Adapted from Spurio (2015) and updated to 2022.

Afterwards the telescopes MAGIC and VERITAS (Ansoldi et al. (2016), VERITAS Collaboration et al. (2011)) have also detected the pulsed emission of VHE gamma rays from the Crab Pulsar, by means of photon timing analysis.

Crab Nebula has a preminent role among VHE gamma-ray sources also because it is used as the standard candle in the field; as a steady emitter, it is a good flux and spectral calibrator. Until 10 years ago the Crab Nebula was considered a good calibrator in X-ray and HE gamma-ray astronomy as well, but in the last years mainly in the X-ray astronomy the Crab Nebula is not anymore used as standard candle due to small-scale variations in the emitted flux (Wilson-Hodge et al. (2011)), while in HE gamma-rays intense flares (with flux raising up to few times the standard one) of debated origin have been detected by the AGILE and Fermi satellites. No counterpart of these flares has been detected in the VHE gamma-ray domain, therefore the Crab Nebula remains a reliable calibrator for IACTs. Other famous gamma-ray PWN sources are Geminga (Acciari et al. (2020)) and Vela (Abdo et al. (2009)). The mechanisms responsible of the gamma-ray emission in pulsars are still under debate.

1.2.2 Supernova Remnants

Supernova Remnants are another class of VHE gamma-ray emitters with peaked emission in X rays. Remnants are what remains after a supernova explosion; these objects are detected within our galaxy. The SNR are classified based on their morphology (Vink (2011)):

- **shell-type:** the Supernova explosion causes a shock wave that during the expansion through the space hits the interstellar medium warming and shaking it. This cause the creation of a hot shell that appears like a ring structure from the Earth.
- **Plerion (Crab-like) remnants:** this type of nebulae is powered by an internal pulsar that accelerate a flux of high-energy electrons. The magnetic field bends the trajectory of the electrons that produce synchrotron radiation and emit in the radio, visible and X-ray bands. Differently from the ring-shape of shell-like remnants this type of SNR shows a less symmetric morphology, like a "blob".
- **Composite remnants:**

  The third type of remnants shows a mixed shape, intermediate between a plerion and a shell-like remnant. The observed morphology depends on the investigated wavelength too. In general these remnants can be further divided in two subclasses:

  - **Thermal composites:** at radio wavelengths, where synchrotron radiation dominates, these SNR show a shell-type shape, while in the X rays these SNR appear plerion-like but with a special feature: X-ray emission lines that are a signature of the presence of hot gas.
  - **Plerionic composites:** These type of SNR show plerion-like appearance in radio and X-ray wavebands but they have a shell. Their X-ray spectrum shows spectral lines close to the shell and no spectral lines in the center.

For all these objects the emission mechanisms are under debate: leptonic models seem prevailing but an hadronic component is not excluded, even if no compelling evidence for it has been reported so far (see e.g. Gabici, Gaggero, and Zandanel (2016)). Perhaps, the strongest claim for a hadronic component in SNRs so far has been proposed by Morlino and Caprioli (2012) for the Tycho SNR. An example where the leptonic model fits better the gamma-ray SED can be found in Abdo et al. (2011) for the young SNR RXJ 1713.7-3946 (see Fig. 1.9).

However, also hadronic models reproducing correctly the HE gamma ray points have been proposed, and results of X-ray observations as well hint to CR acceleration in this source. Thus it is expected that some additional evidence (e.g. arising from the observation of astrophysical neutrinos coming from the direction of this source) is needed to solve the riddle (Gabici and Aharonian (2016)).

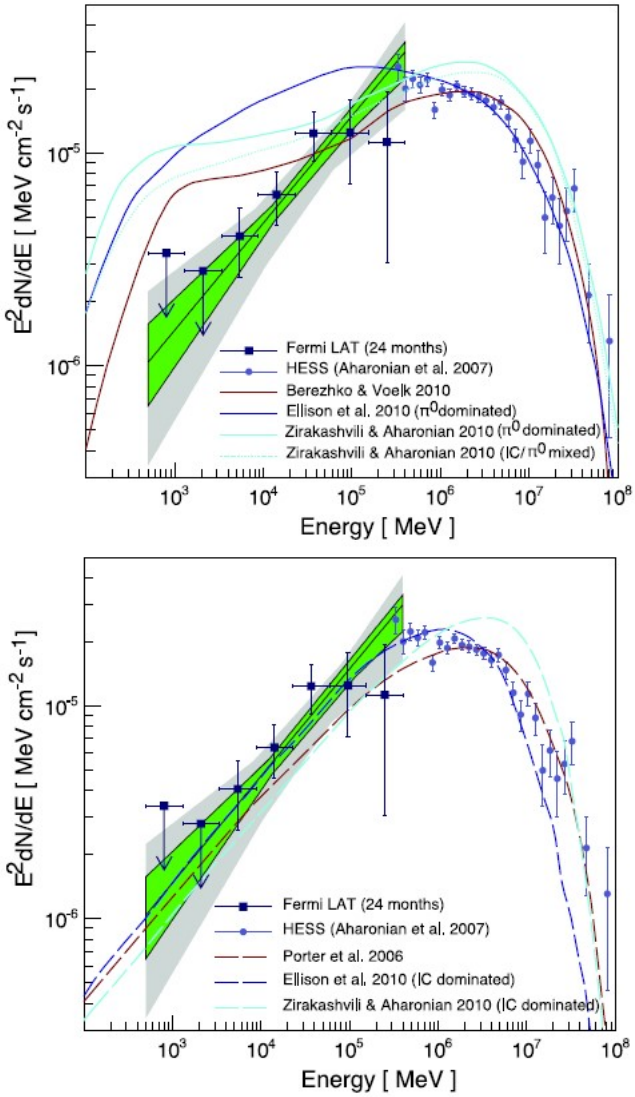


Figure 1.9: From Abdo et al. (2011), the gamma-ray SED for the SNR RXJ 1713.7-3946. While a leptonic model fits adaequately the HE gamma-ray SED (bottom panel) hadronic ones fail to reproduce it.

1.2.3 Active galactic nuclei

A good definition and description of Active Galactic Nuclei can be found in Dermer and Giebels (2016). In the current picture every galaxy hosts a supermassive black hole (SMBH) with mass ranging from $3 \times 10^6 M_\odot$ to $3 \times 10^9 M_\odot$. The mass is correlated with the galaxy velocity dispersion, brightness and hence with the mass of the host galaxy. The majority of known galaxies (99%) are not active and the central SMBH is quiescent. The remaining 1% of galaxies show activity and are called Active Galactic Nuclei (AGN), extragalactic sources that are among the most powerful emitters in the sky, and the most powerful persistent ones, with luminosities up to $L \simeq 10^{48}$erg/s.

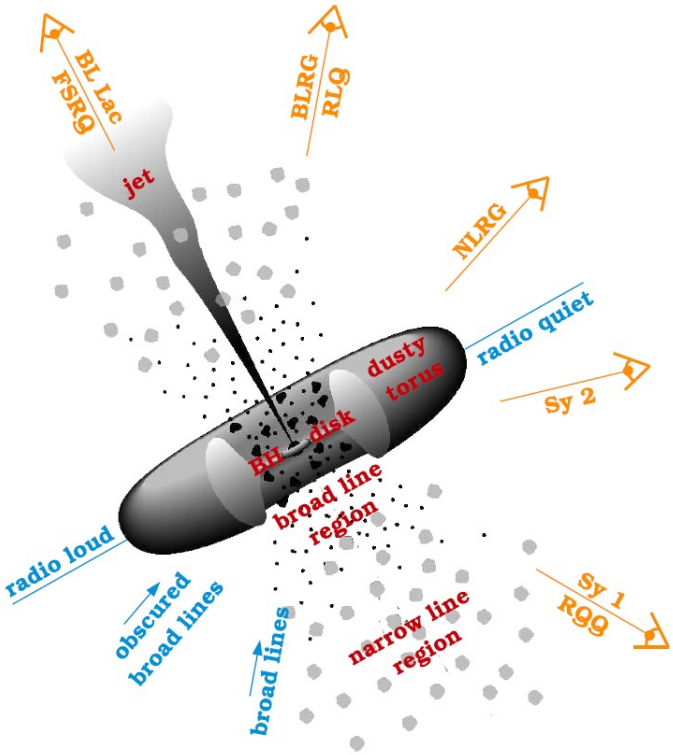


Figure 1.10: Observational classification of active galaxies from Biteau (2013)

AGN model

In this section the AGN unification model is described (see e.g. Antonucci (1993) and Urry and Padovani (1995)). It is believed that the AGNs share a common structure, with most of the observational variety (see Fig. 1.10) being induced by

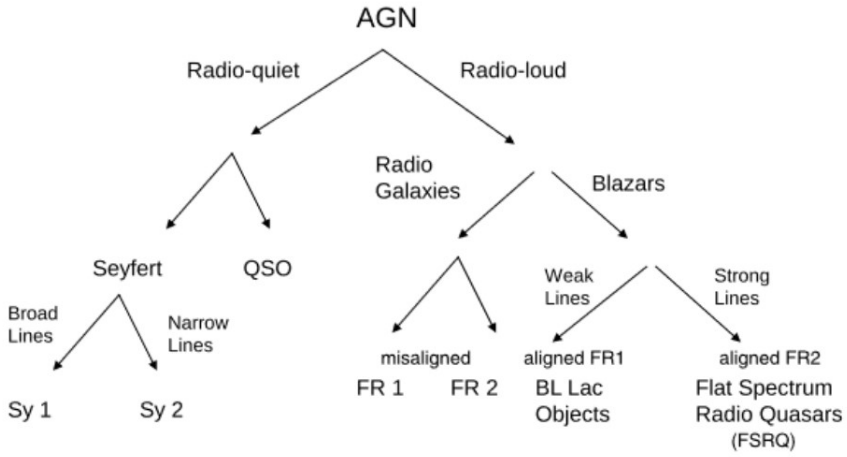


Figure 1.11: Observational classification of active galaxies from Dermer and Giebels (2016)

differences in orientation. Following e.g. Ghisellini (2013) the structure of AGNs (also sketched in Fig. 1.11) can be described as built up by these components:

- **Central SMBH:** in the center of every AGN there is a SMBH, of debated origin, possibly deriving from merging of primordial BHs. The BH causes the rotation of the surrounding matter; a fraction of this matter orbits and condensates into a disk, while an infalling fraction feeds the SMBH.
- **Accretion disk:** a disk of matter in rotation close to the SMBH; the matter is attracted towards the central body spiralling. A compression is caused by friction and gravitational forces, the compression turn causes an increase in the temperature of the matter that eventually produces EM emission, mainly in the UV and X-ray bands. This is a major source of power that can reach a luminosity $L = 10^{44} - 10^{47}$ erg/s.
- **X-ray corona:** hot gas enveloping the accretion disk, either in a hot layer, or in clumpy regions.
- **Jet:** a fraction of AGN jets are produced. These are a pair of collimated flows of relativistic plasma, propagating from the central engine in opposite directions. Jets are launched in the innermost region of the AGN and extend up to several Mpcs. About 10% of AGNs are jetted.
- **Torus:** At few parsec from the BH there is an obscuring torus. A fraction of the radiation emitted from the disk is intercepted and re-emitted by the torus in the form of infrared radiation. The torus blocks the view of the

central region for those AGNs that are observed perpendicularly (or almost perpendicularly) to the axis of the torus.

- **Broad line region (BLR):** Near the center composed by many small clouds in fast motion ($v \sim 3000$ km$\times$s^{-1}) close to the central singularity (at distances ($D \sim 10^{17}$ - 10^{18} cm). This region re-emits as optical/UV emission lines the 10% of the ionizing radiation received from the disk. These lines are broadened due to the Doppler shift, for this reason the region is called "broad line region".

- **Narrow line region (NLR):** Further away with respect to BLR ($D \sim 100$ pc) there is another region composed by clouds that move slower and are less dense than BLR.

AGN classification

The AGN classification (see Fig.1.10 and Fig.1.11) is mainly based on the observation direction and on few other properties (see e.g. Dermer and Giebels (2016) and Biteau (2013)). The main intrinsic separation is based on the power of the radio emission component, relative to the optical. AGNs are thus divided into "radio-loud" and "radio-quiet". This separation is given in terms of the ratio of the radio on optical bands (Eq.1.17).

$$R = \frac{F_{radio}}{F_B} \tag{1.17}$$

Where F_{radio} and F_B are the fluxes in the radio and in the optical B band respectively.

Radio-loud AGNs are $\sim 10\%$ of the AGNs, and are those associated to jets. They are divided according to the direction of observation; if the direction is aligned with the jet the AGN are called "Blazars" (treated in more detail 1.2.3), otherwise we have the misaligned AGN called "Radio Galaxies". These are in turn divided in "Fanaroff-Riley I and II" (FRI and FRII) based on the radio power, with FRII being the most powerful ones.

In the radio-quiet branch (without the jet) the AGNs are divided in:

- **Quasi-stellar object QSO** that are in general very distant object and are detected as point-like sources;

- **Seyfert** galaxies with the host galaxies clearly detectable, generally associated to spiral host galaxies. The Seyfert galaxies are in turn split in "Seyfert I" where the broad lines dominate the narrow lines and "Seyfert II" where the narrow lines are dominant.

Blazar classification

The radio-loud AGNs observed along the axis of their jet are called blazars. These objects are studied deeply as they are among the most energetic objects in the universe; they emit in all the EM spectrum and they are easily observable also due to relativistic beaming effects.

Historically different classifications for blazars have been introduced and coexist; they can depend, for instance, on the peak frequency of the synchrotron component in the SED or can be related to the presence or absence of broad emission lines in the optical band. A commonly used diagnostic for splitting blazars in flat spectrum radio quasars (FSRQ) and BL Lac objects is the equivalent width of optical emission lines (EW) that is defined:

$$EW = \int \frac{F_\lambda - F_0}{F_0} d\lambda \tag{1.18}$$

where F_λ corresponds to the total flux (broad emission lines + continuum emission) while F_0 is the flux without the broad emission lines. Equivalent width is measured in Angstrom (Å).

Blazars are thus-divided into:

- **BL Lacertae (BL Lac objects):** with $EW < 5$ Å these objects have weak or no optical lines. These are further divided depending on the synchrotron peak frequency (ν_s): from the low energy peaked BL Lac (LBL) with the peak at lower energies with $\nu_s < 10^{14} Hz$, the intermediate energy peaked BL Lac with $10^{14} < \nu_s < 10^{15}$ until the high energy peaked BL Lac (HBL) with $\nu_s > 10^{15}$.

- **Flat Spectrum Radio Quasar (FSRQ):** with $EW > 5$ Å these objects have intense emission lines and the synchrotron peak frequencies are homogeneous to those of LBLs or even lower

An example of SED these objects taken from Spurio (2015) is shown in Fig: 1.12.

The blazars different sub-classes suggest a SED sequence: FSQR → LBL → IBL → HBL that sometimes is called "blazar sequence". The sequence shows an increase of synchrotron peak Energy, with the decrease of source luminosity (Ghisellini (2013)).

1.2.4 Gamma Ray Bursts

This kind of transient events is the most powerful phenomenon known nowadays but the GRB progenitors are also the least known. GRB are very powerful explosions that radiate through the universe for billions of light year. These events

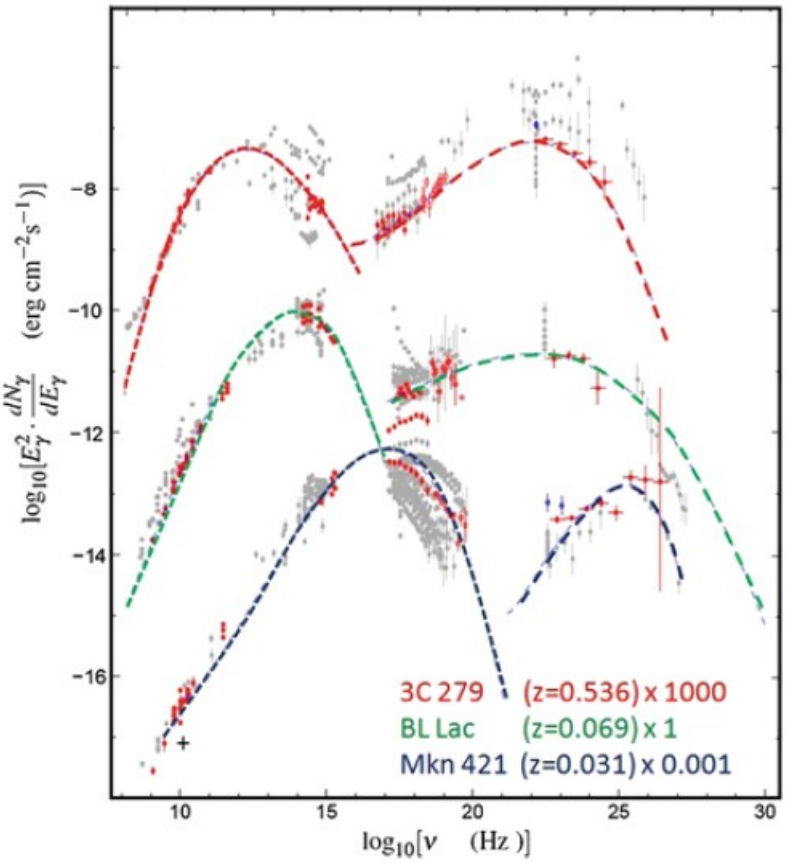


Figure 1.12: From Spurio (2015) The SED of three AGNs at different distances (z is the redshift) from different subclasses. The FSRQ 3C 279 (spectrum multiplied for 10^3, the BLLac a LBL and the Mrk421 HBL with spectrum multiplied by a factor of 10^{-3}. The dashed lines represent the best fit for the data assuming a leptonic model.

are rare, unpredictable and arrive at the Earth from outside the Milky Way. Furthermore these events show a wide variety of behavior. These are the reasons that make it difficult to identify the origin of these events.

The GRB most used model is called Fireball Model (Piran (1999)). Its assumptions are the following:

- the energy produced from one GRB is $10^{51} - 10^{52}$ erg
- the flux of the GRB is very variable as confirmed by the observed GRB and it should last for the time of the burst that can be of timescale of seconds-minutes-hours. Furthermore the burst afterglow can continue at lower energies for hours to days.

- The events are very rare (once per millions years for galaxy) and the overall GRB emission in γ-rays is about 10^{52}erg/10^{6}years /galaxy

The model that should be able to consider all these assumption must be very versatile. For these reasons the GRB classification is based on their time duration:

- **Short gamma-ray burst:** in this category there are $\sim$ 30% of the GRB detected, the GRB that during 2 or less seconds are considered in this group. This type of GRB are in general associated with regions with low or no star formation, central regions of galaxy clusters or elliptical galaxies.

- **Long gamma-ray burst:** The majority of detected GRB ($\sim$70%) are in this group. In this category there are GRBs with duration greater than 2 seconds. These GRB are easier to study with respect to the short ones because the afterglow is longer and brighter. For this reason the research works of the long GRB are more widespread. It was supposed that the observed long GRB arrive from star forming galaxies and are generated in core-collapse supernovae of massive star. The first long GRB associated with a supernova event (SN1998bw) is GRB980425 (Galama et al. (1998)). After that various other supernova events have been associated with GRB. For this reason it is supposed that the long GRB are connected with the death of massive stars.

- **Ultra-long gamma-ray burst:** These events are more rare than the others two groups, the duration for these events is more than 10^4 seconds (about 3 hours). This class of GRB was proposed after the detection of GRB11209A in 2011 that lasted for more than 7 hours. The sources proposed for these GRB are: collapse of blue supergiants, birth of magnetars (a pulsar with a strong magnetic field) or in a tidal distruption events (star destroyed by a SMBH).

1.3 Detectors for Gamma-ray Astronomy

In general the astrophysical experiments for detection of astroparticles are divided in two main categories, the space-based and the ground-based detectors. Both of these are very important for the γ rays detection but in general, due to the flux of High energies particles, that is strongly decreasing with energy, the space experiments (like the Agile or Fermi γ ray space telescopes) are able to detect less energetic gamma rays than the ground experiments, because of their limited collection area (order of 1m^2 at most). This, as mentioned in the previous sections, makes each experiment specialized for the detection of a specific region of the spectrum.

In general, ground-based experiments can be classified in two categories: the high altitude experiment array like HAWC, ARGO, LHAASO shower detector arrays (briefly described in Sec.1.3.4) and IACTs like MAGIC, HESS, VERITAS (described in Sec. 1.3.5). Combining the results of the space and ground telescopes one can have a deeper knowledge of the electromagnetic spectrum (Multiwavelength observation). In table: 1.2 is shown that the gamma-ray band comprehend many decades in energy, therefore no single detector can measure precisely all the gamma-ray band alone. A good summary of the sensitivity of gamma-ray experiments as a function of energy can be found in Fig. 1.13 (From Knödlseder (2016)). In the following are described in simple the space and ground based experiments

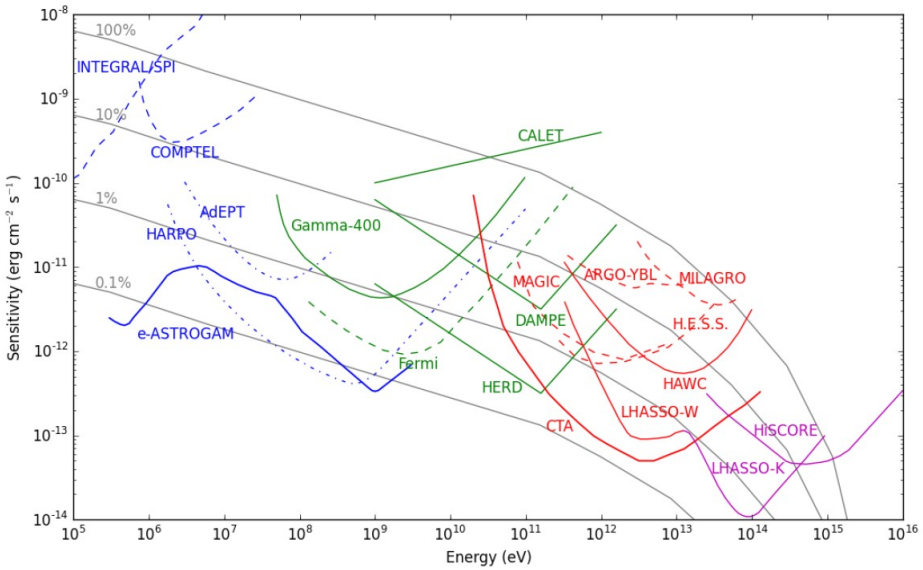


Figure 1.13: From Knödlseder (2016) the sensitivity as a function of Energy for different gamma ray detectors. The energies go from 10^5 eV to about 10^{16} eV.. The lower energies are covered by space telescopes experiments while going towards the high energies ground detectors dominate.

with a focus for the imaging atmospheric Cherenkov technique.

1.3.1 Direct Detection from Space Satellites

The satellite experiments cover energies from 100 keV to few hundreds of GeV. This is due to the flux of cosmic rays: more the energies increases more the energetic particles are rare.

The satellite telescope in general use calorimeters that must stop the arriving particle to detect its properties. The "stopping power" is proportional to the

weight of the telescope and the weight of a space telescope is strictly connected to the price of the telescope.

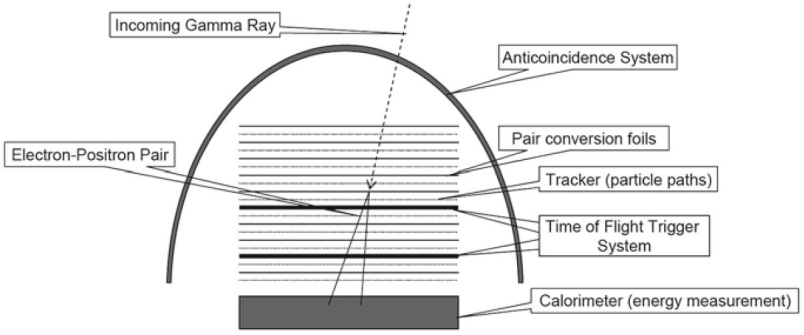


Figure 1.14: From DiSciascio (2019) Sketch of a telescope which converts γ to $e^{\pm}$ (pair-conversion telescope), reproducing the features of the EGRET experiment.

In the following the basic principles of operation of a space γ-ray telescope (EGRET Kanbach et al. (1989), CALET Torii, Marrocchesi, and Calet Collaboration (2019) and FERMI Atwood et al. (2009)) are summarized. The purpose of these telescopes are the recognition of the γ ray from the cosmic ray background and the evaluation of the γ ray arrival direction, time and energy. In Fig. 1.14 the main telescope components are represented. The first component that incoming particles cross is the anticoincidence system (AS); a sort of first discriminator between the background and γ ray based on electrical charge. If the signal meets the criteria of the AS the event is considered generated by a gamma ray.

After the AS, the γ ray interacts with one of the (pair) conversion foils and produces an electron/positron pair. Interleaved with the pair conversion foils there are tracker plates designed to detect the path of the electron and positrons, to reconstruct the interaction vertex and to detect the γ ray arrival direction.

Furthermore, in the middle there are two scintillator devices used to measure the time of flight of the particles in the detector and to confirm the directions. After this the particles enter in the calorimeter; this device is designed in such way that the hitting particles generate an electromagnetic shower, in this way the particles lose their energy with every interaction until they stop. Measuring the whole energy of the shower products allows the estimation of the energy of the parent particle.

1.3.2 Air Showers

Before describing the ground-based experiments it is necessary to explain the behaviour of the energetic particles that arrive at the Earth from the space. When a

particle hits the Earth atmosphere, the atmosphere acts as a calorimeter (1.3.1). The impact makes the particle generates a shower of less energetic particles going down to the ground in a cascade of further interactions. The cascade continues until the final particles no longer bring enough energy to generate further sub-showers. These air shower are called Extensive Air Showers (EAS). The maximum number of particles that can be generated from a primary particle is called "shower maximum (N_{max})"). The max height (X_{max}) position in function of the energy (E_0) of the primary particle is called "elongation rate (Δ)":

$$\Delta = \frac{dX_{max}}{d(Log_{10}(E_0))} \tag{1.19}$$

Δ and N_{max} are proportional to the energy of the primary particle, and are crucial to calculate the primary particle properties and to design the experiments as well. Ground-based experiments in fact are built at altitudes that are chosen based on the energy of the primary particles for which the detector is designed for.

Shower front detectors reveal the particles in the shower, and from their spatial and arrival time distributions infer the kind, the energy and the incoming direction of the primary particle that generated the shower. The major weakness of shower front detector is that these telescopes can detect only the last products of the shower. For this reason the recognition of the primary particle is not easy and the measure is indirect.

In general the estimation of the properties of the primary particles is performed with the help of Monte Carlo (MC) simulations of showers to be compared with the observed ones. There are various models used to create MC showers. They are based on two categories of showers: the Electromagnetic shower that is composed of photons, electrons and positrons, and hadronic showers composed mainly of hadrons but also of γs electrons and positrons.

In the following a basic[3] model of each type of shower is described in order to show why the separation of the gammas from the hadrons can be non trivial.

EM shower Model

The Heitler model (Heitler (1954)) is the most simple model for an EM shower, despite the simplicity this can be used to understand some of the properties of the EM showers. The EM showers are generated by a γ or a $e^{\mp}$ that interacts with the atmosphere. The involved decays are:

$$\gamma \rightarrow e^- + e^+ \quad \text{(Pair production)} \tag{1.20}$$

$$e^- \rightarrow \gamma + e^- \quad \text{(Brehmstrahlung)} \tag{1.21}$$

[3]the MC model are based on these but are much more complex

$$e^+ \rightarrow \gamma + e^+ \quad \text{(Brehmstrahlung)} \tag{1.22}$$

all of these decay are based on the simplified assumption that the energy of the primary particle E_0 is divided equally between the two products, furthermore the energy loss of a particle in atmosphere can be approximated with:

$$-\frac{dE}{d\xi} = \frac{E}{\lambda_T} \tag{1.23}$$

Integrating:

$$E(\xi) = E_0 e^{-\frac{\xi}{\lambda_T}} \tag{1.24}$$

where ξ is the particle travel distance in the mean (air) and $\lambda_T = 37\,\frac{g}{cm^2}$ is the radiation length of the air. The thickness of matter crossed by the particle before it loses an half of energy is:

$$\frac{E(\epsilon)}{E_0} = \frac{1}{2} = e^{-(\frac{\xi}{\lambda_T})} \tag{1.25}$$

$$\epsilon = \lambda_T ln2 \tag{1.26}$$

This means that an $e^{\mp}$ produces a γ and viceversa a γ produces an $e^{\mp}$ after travelling a distance in air equal to $d = \lambda_T ln2$. In this last passage it has been used the approximation that λ_T is the same for γ and $e^{\mp}$; with this the model overestimates the number of electrons/positrons, for this reason in general is introduced the attenuation constant $g = 10$ with $N_e = \frac{N_{max}}{g}$. In every decay the energy is divided in two equal parts between the two secondary particles. After this consideration is easy to draw a picture of the decay. In figure 1.15 a sketch of the Heitler model is shown.

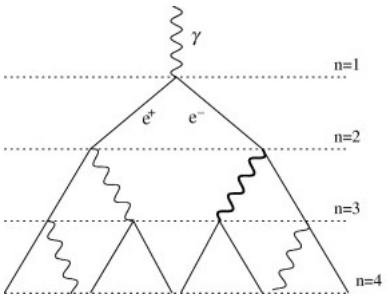


Figure 1.15: From Matthews (2005) Schematic views of an electromagnetic cascade.

Considering an air shower produced by a gamma ray of energy E_0 the number of particles after n half-life lengths d is 2^n with an energy for each particle of $\frac{E_0}{2^n}$.

The length X after n lengths is:

$$X = nd = n\lambda_T ln2 \tag{1.27}$$

Using the logarithm properties $ln_2 = log_2(2)/log_2(e)$ the number of particles can be rewritten as:

$$N = 2^n = 2^{X/\lambda_T ln2} = 2^{\frac{X}{\lambda_T} log_2(e)} = e^{\frac{X}{\lambda_T}} \tag{1.28}$$

The number of particles at the maximum of the shower (N_{max}), considering that the particle generation ends at E_{crit} defined as the energy level where the average collisional energy losses begin to exceed radiative losses (1.15), is:

$$\begin{cases} N_{max} = 2^n \\ E_0 = N_{max} E^{\gamma}_{crit} \end{cases} \tag{1.29}$$

from which one can obtain:

$$n = \frac{1}{ln2} ln \frac{E_0}{E^{\gamma}_{crit}} \tag{1.30}$$

From this one can say two things:

- the number of particles at the maximum of the shower is $N_{max} \propto E_0$
- the maximum shower depth is equal to $X_{max} = \lambda_T ln \frac{E_0}{E^{\gamma}_{crit}}$

For the last value one can obtain the elongation rate that is defined in Eq:1.19:

$$\Delta = \frac{dX_{max}}{dLogE_0} \tag{1.31}$$

and using the X_{max} value with logarithm substitution $Log_{10} \frac{E_0}{E_\gamma} = \frac{ln(E_0/E_\gamma)}{ln_{10}}$:

$$\Delta = \frac{d\lambda_T ln10 Log(E_0/E^{\gamma}_{crit})}{dLogE_0} = 2.3\lambda_T \approx 85 \frac{g}{cm^2} \tag{1.32}$$

Where it is estimated the elongation rate of a an EM shower in air.

Hadron shower model

There are many model for the hadron showers, these models are much more complex because a proton can decay in atmosphere in many different particles. A very simple model (but already more complex than the Heitler) considers that a p^+ can generate equally only π^+, π^-, π^0 particles (see Fig. 1.16). With this assumption are generated in the shower 2/3 of charged pions and 1/3 of neutral pions, with N_{ch} charged pions and $1/2N_{ch}$ neutral pions.

The π^0 has mean life time $\tau_{\pi^0} \approx 8 \cdot 10^{-17} s$ and decays 98% of the time in $\gamma + \gamma$. From both γ will be generated an EM shower described in 1.3.2 [4]

The $\pi^{\pm}$ mean life time is $\tau_{\pi^{\pm}} \approx 2.8 \cdot 10^{-8}$,much more than the π^0. For this reason these particles can interact before decay generating other π^0 and $\pi^{\pm}$ until the energy of the charged pions is too low. In figure 1.16 is shown a model of an hadronic shower. The major part of the shower sooner or later produces γ,

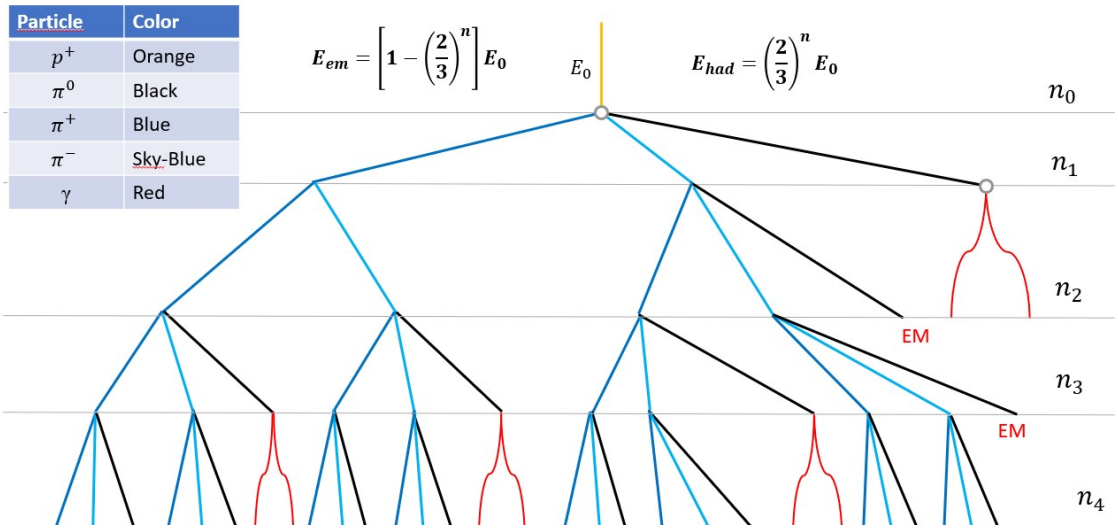


Figure 1.16: Sketch model of an hadronic shower with calculation of the Energy of the EM and the hadronic components

this means that the EM energy contribution is high. To estimate the energy of the primary particle one can fix the two values N_{ch} and E^{π}_{crit} (E_{had} in Fig: 1.16), considering that from the accelerator experiments it is known that the dependence of N_{ch} (the number of charged hadrons) from the energy is $\propto E^{\frac{1}{5}}$. Given that it is possible to approximate N_{ch} as a constant value. After n interactions (see Fig: 1.16) the numbers of charged pions are $N_{\pi^{\pm}} = N^n_{ch}$. After n_{crit} atmosphere layers the energy of the charged pion becomes less than $E^{\pi^{\pm}}_{crit}$:

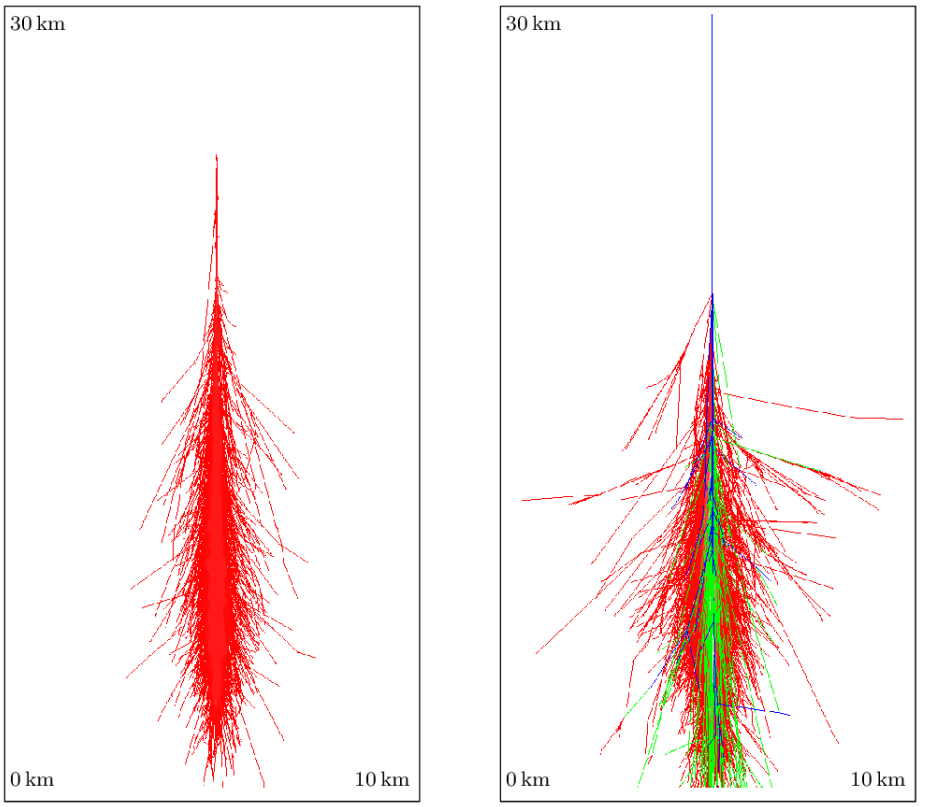
$$E^{\pi^{\pm}}_{crit} = \frac{E_0}{(3/2N_{ch})^{n_{crit}}} \tag{1.33}$$

Therefore:

$$n_{crit} = \frac{ln(E_0/E^{\pi^{\pm}}_{crit})}{ln(3/2N_{ch})} \tag{1.34}$$

[4]This is very important for the ground-based gamma ray experiments because this type of purely EM sub-showers can be easily misinterpreted as genuine gamma-like events.

These two simple sketch-models show how evaluating and simulating an hadronic shower can be complex. Furthermore, for this specific case muons were not considered. It is hence easy to understand why the computational resources required for hadronic simulations are much higher than for the EM ones. The models shown are the starting point to build the simulations model, to give an idea of the results after a gamma and hadron MC simulation is shown in Fig. 1.17 (from Häffner (2010)) the resulting shower profile of one EM shower (left) and one hadron shower (right).



(a) γ-ray induced air shower with 300 GeV and first interaction in 23 km height

(b) hadron induced air shower with 1 TeV and first interaction in 18 km height

Figure 1.17: From Häffner (2010): Simulations of γ-ray and hadron induced air showers. A γ-ray initiated air shower consists only of an electromagnetic component (illustrated in red), whereas the hadron-induced shower has apart from the hadronic component (blue) a distinct muon component (green) and electromagnetic sub showers (red).

Monte Carlo shower Simulations

In ground experiments the observation of events is indirect. For this reason important properties like energy and arrival direction of the original particle have to be reconstructed by properties of the secondary particles generated. To solve the problems of estimating the energy the position, and separation of gamma from hadron the ground experiments use a series of Monte Carlo algorithms to simulate the showers. A famous algorithm is called COsmic Ray Simulations for KAscade (CORSIKA). The detail about that algorithm can be found in Heck et al. (1998). With this algorithm is possible to simulate the EM and hadron-like extensive air showers shown in the previous chapters and in figure 1.17.

1.3.3 Detection of Air Showers from Ground

The Cherenkov radiation discovered by Pavel Cherenkov (Čerenkov (1937)) provides a very effective way for the ground experiments to detect and measure indirectly the Extensive Air Showers (EAS) originated from a primary particle.

The Cherenkov radiation is a glow of light generated by a charged particles moving in a dielectric medium faster than light. During the transit the charged particle polarizes and accelerates the dielectric particles, which therefore radiate. The Cherenkov effect is illustrated in Fig.1.18, when the speed of particle in the medium $v < c_{vac}/n_{mean}$ the wave fronts generated arrive to the observer separated and intensity of light is too low to be detected even from a very sophisticated electronic detector. Instead in the case where $v > c_{vac}/n_{mean}$ the wavefronts cross and the superimposition can be detected as a flash of light.

It can be noticed that the principle of the Cherenkov radiation is very similar to the sonic boom in air when the sonic wave is generated by an object faster than the sound speed. In this case, the wavefronts can arrive at the same time to the listener(observer). The Cherenkov light can be seen with naked eye in the water of a nuclear power plant. The Cherenkov light is generated also in atmosphere by the astroparticles that hit the Earth, however in this case is impossible to see the light with naked eyes but it can be detected with electronic detectors. Many experiments that exploit the Cherenkov effect exist for different types of particles and energies, in particular this work is focused on the gamma ray detection with the MAGIC IACT described from Sec. 2.

1.3.4 High Altitude Water Cherenkov Experiments

The High Altitude Water Cherenkov experiments (HAWC, MILAGRO) belongs to one of the two kinds of ground experiments, exploiting the Cherenkov effect, the

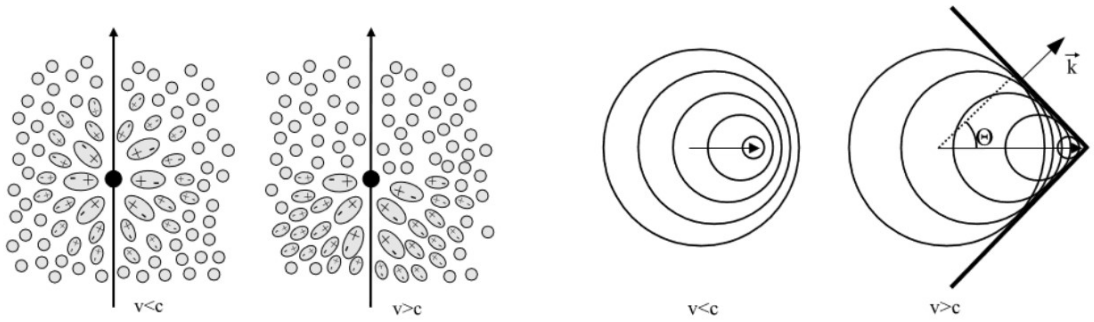


Figure 1.18: From de Naurois and Mazin (2015) Left: Illustration of the polarization of the medium induced by the crossing of a relativistic particle. Right: Construction of Cherenkov wave-front. v is the speed of the charged particle in the mean and $c = c_{vac}/n_{mean}$ is the speed of light in the mean

"Water Cherenkov" experiments. These "telescopes" can be used to detect CRs and γ ray.

For instance in HAWC, the technique consists of detecting the particles using many closed water tanks (Fig: 1.19) within which are placed some Photomultiplier Tube (PMT) used to detect the particles of the shower. These experiments have about 100% of duty cycle due to the closed tank that isolate the PMT from the natural light and a wide field of view, much wider than the IACTs (Sec. 1.3.5).. A weakness of these experiment is the low discrimination power for gamma/hadron particles with respect to the IACTs. In Fig. 1.13 it can be noticed that these (HAWC and LHAASO in particular) experiments are complementary to the IACTs because the sensitivity curves cover two different ranges of energies, with the IACTs extending down to few tens of GeVs.

In this thesis this technique is mentioned for completeness but will not be treated in depth.

1.3.5 Imaging Atmospheric Cherenkov Telescopes

The other kind of ground telescopes based on the Cherenkov effect are the Imaging Atmospheric Cherenkov Telescopes (IACT). Like HAWC (Smith (2015)) the technique is based on the detection of the Cherenkov radiation but instead of using the water these telescopes use the air as dielectric transparent mean. Currently the three most important telescopes using this technique are the Major Atmospheric Gamma Imaging Cherenkov (MAGIC), High Energy Stereoscopic System (HESS) and Very Energetic Radiation Imaging Telescope Array System (VERITAS). A newer observatory Cherenkov Telescopes Array (CTA) based on an evolution of the technologies of these three is under construction (Acharya et al. (2013)). These

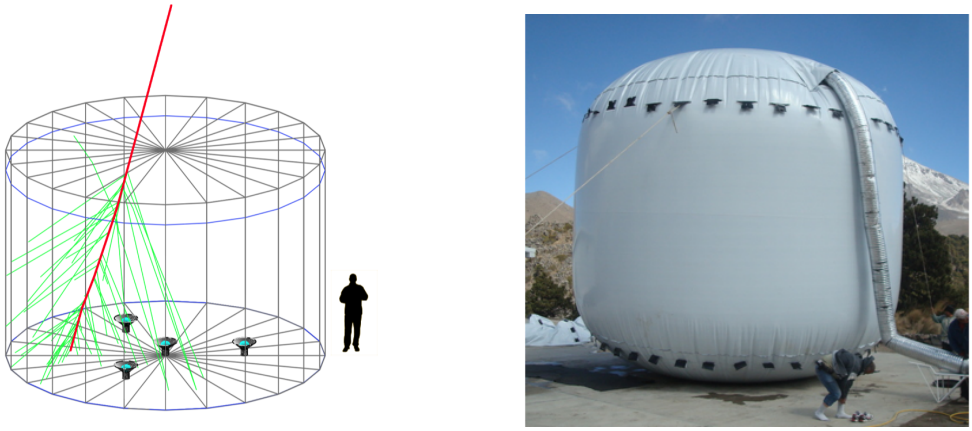


Figure 1.19: From Smith (2015) Left: internal sketch of HAWC water tank, Right: HAWC bladder

telescopes use a mirror to convey, with a reflection, the low intensity light of the Cherenkov radiation on a camera composed of photomultipliers tubes. The three main characteristics of the primary particle measured with an IACT are:

- The generic probability ("likeliness") of the particle to be a gamma is called "gammaness (g)". The opposite, the likeliness of the particle to be an hadron is called "hadronness (h)" where $g = 1 - h$.
- the arrival direction of the primary particle;
- the energy of the primary particle.

This works is focused on the MAGIC telescopes and for this reason it is described in detail and used as a general example of IACT.

The imaging technique

The primary particle impinging on the atmosphere produces an EAS (Sec.1.3.2) that can be detected by the Cherenkov telescope. The telescope detects the Cherenkov light as shown in Fig. 1.20 and records an image built with the charge generated in each PMT by the incoming light. After some calibration and cleaning steps, the shower image can be modelled as a sort of ellipse. In the standard analysis the information contained in the image is condensed in parameters called Hillas parameters (Hillas (1985)) shown in Fig.1.22. The Hillas parameters are used to estimate the "hadronness" and to reconstruct the arrival direction and the energy of the primary gamma rays. One of the main differences between the hadron-like and gamma-like showers is the shape of the image produced: a good

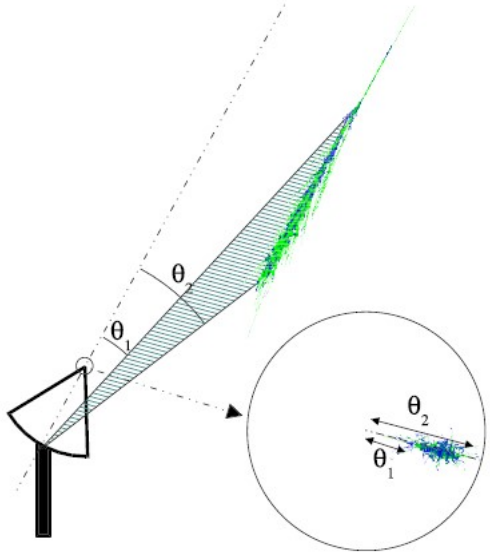


Figure 1.20: From de Naurois and Mazin (2015) The shower image is projected on the camera focal plane.

example can be viewed in Fig. 1.21, where on left images are shown respectively the hadron-like and gamma-like showers and on right images are shown the projection of these showers on the IACT camera plane. The gamma-like event on top right has a more tight-elliptical shape than the hadron-like on bottom right. The separation between gamma-like and hadron-like events is a crucial step, performed in the MAGIC standard analysis by means of a Machine Learning (ML) Random Forest (RF) algorithm that consider a set of the Hillas parameters plus some stereoscopic parameters introduced in the next Sec. 1.3.5.

Impact of the zenithal distance of the observations

The zenith distance (Zd) is the angular distance of a celestial body from the zenith. It influences a lot the detection of the air showers, in particular the Zd is related to the energy threshold and the effective collection area. If Zd increases, the EAS has to go through more atmosphere to get to the observer (Fig. 1.23). The path length for an high Zd shower can be approximated as:

$$L' \simeq \frac{L}{cos(Zd)} \tag{1.35}$$

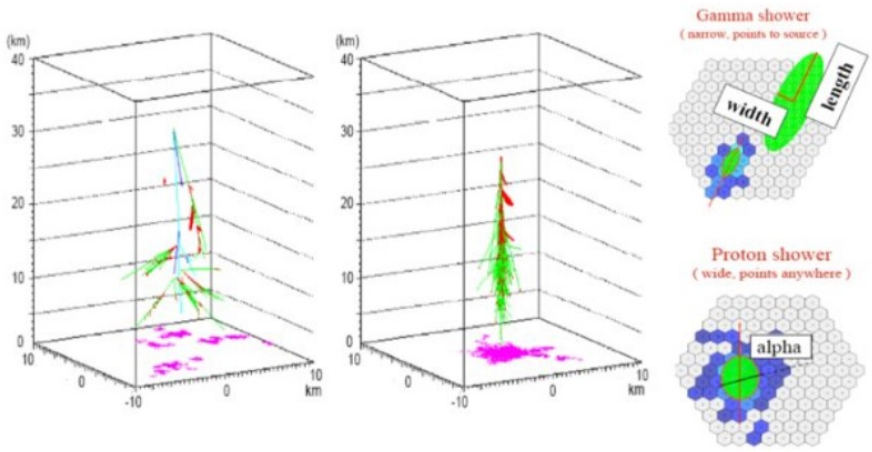


Figure 1.21: From:DiSciascio (2019) The two figures on the left are the scheme of the lateral development of air showers induced by protons (left side) and photons (right side). The hexagonal pictures on the right represent the projection of the two air showers on the IACT camera. The gamma-like event (top-right) is more compact and elliptical respect the hadron-like event (bottom-down)

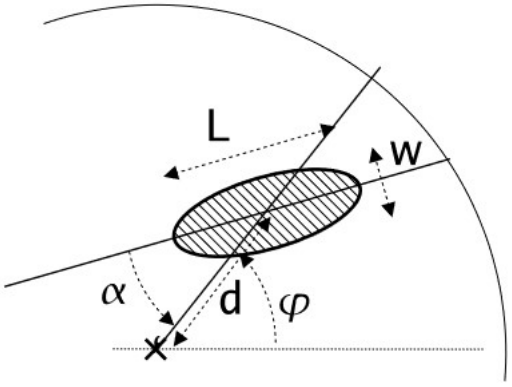


Figure 1.22: From Naurois (2006) The geometrical definition of the Hillas parameters: shower image length L and width W, *size* total charge in the ellipse, d nominal distance (angular distance between the centre of the camera and the gravity centre of image), ϕ azimuthal angle of the images main axis, α orientation angle

and the diameter of the light cone on the reflector, called "light pool radius", can be approximated in the same way:

$$l' \simeq \frac{l}{cos(Zd)} \tag{1.36}$$

Moreover considering that the amount of light arriving on the detector, that is the photon density ρ, is proportional to the area of base surface of the cone (l^2) one can rewrite the equation 1.36 to find the Cherenkov light density at high zenith angle:

$$\rho' \propto l'^2 \propto \frac{\rho}{cos^2(Zd)} \tag{1.37}$$

This implies the reduction of the Cherenkov photon density at high zenith distances and given that ρ is strictly connected to the energy causes an increment of the energy threshold (see the plot in sec 2.3.1). This influences also the effective collection area (see sec 2.12), that increases with the Zd of the observations.

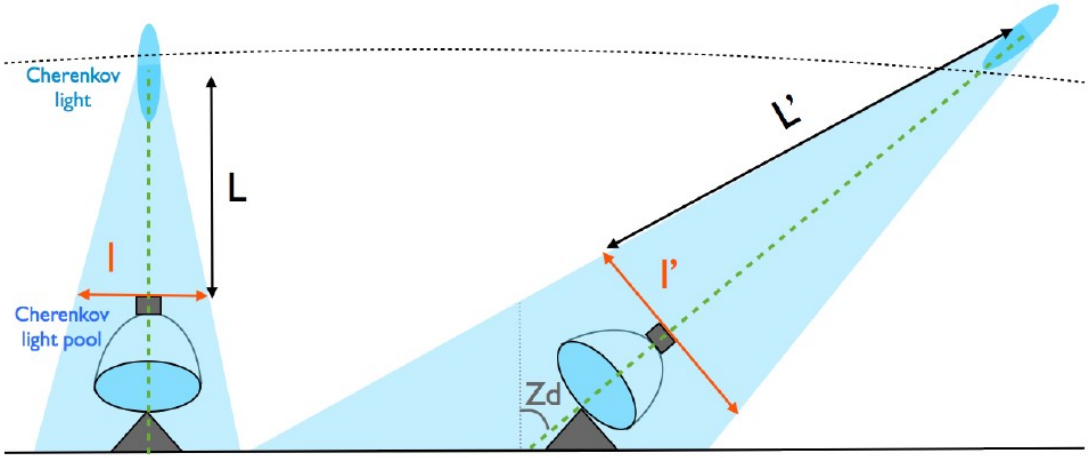


Figure 1.23: From Lopez-Coto (2015) Difference in the shower development between low Zd (left telescope) and high Zd (right telescope) observations. We can see that the distance from the camera to the point in the atmosphere where the showers start is smaller for low Zd observation (L) than for high Zd observations (L'). The diameter of the Cherenkov light pool also increases for high Zd observations ($l' \simeq l/cos(Zd)$).

Stereoscopic parameters

In addition to the Hillas (see Sec. 1.3.5) parameters, the stereoscopic parameters are used in the analysis when two or more telescopes belonging to an IACT acquire an image of the same event. These parameters can improve a lot the telescope performances, in particular they are used in γ-like/h-like separations algorithm (Sec. 2.2.1). In Fig. 1.24 is shown how stereoscopic observations can improve the performance. One of the most important observables is the primary particle direction that, in first approximation, can be estimated as shown in the middle

panel. In stereoscopic observations, all the telescopes point the same direction in the sky, but each telescope images the same shower event with a different parallax; the black lines represent the ellipse major axes. In each image, the source position in the sky lies somewhere along the major axis, but this position is undetermined and has to be estimated with a further analysis. With multiple images of the same region of the sky, these can be superimposed, and the intersection of the 4 lines gives a good first approximation about arrival direction. The arrival direction estimation can be improved with regression techniques (Sec. 2.2.1).

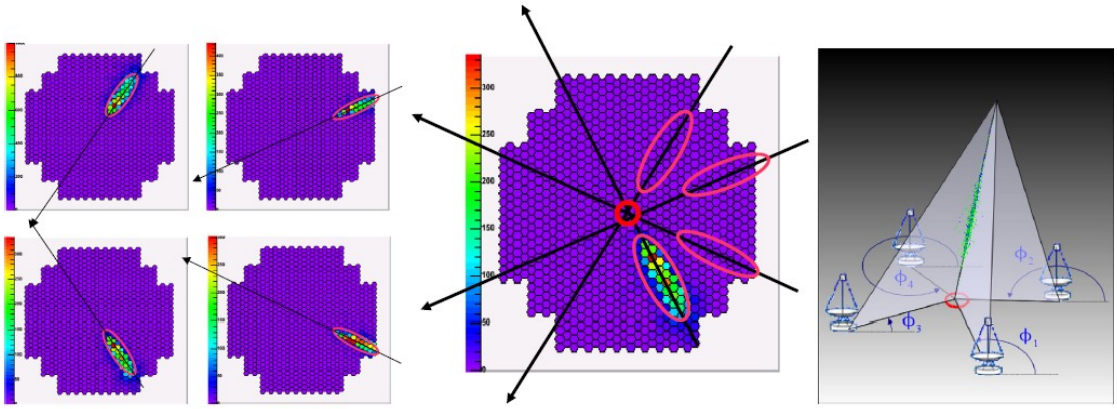


Figure 1.24: From de Naurois and Mazin (2015) Geometric reconstruction of source direction and impact parameter in stereoscopic mode. Left panel: in the camera frame, the main axis of the shower corresponds to a plane that contains the actual shower track and the telescope. The primary particle direction corresponds to a point on this main axis. Middle: The intersection of the main axis of the images recorded by the different telescopes immediately provides the primary particle direction. Right: Direct intersection of the planes containing the shower tracks and the telescopes provides the shower impact point on the ground.

Sources of background events

The shower images that constitute the background for VHE γ-ray astronomy can be classified depending on the primary particle originating the extended air shower (hadrons, leptons, diffuse photons) more in detail (see also Fig. 1.25):

- **Night sky background(NSB)**: this is due to the brightness of the night sky and is generated by the stars, Moon and light pollution, airglow and other effects. Given that the PMT receivers are very sensitive to the light they can be triggered by photons originated by NSB. This noise can be reduced using specific hardware trigger settings. An additional step to reduce this

noise is in software, by means of suitable cleaning settings. Gamma/Hadron separation algorithms do not consider this source of background, assuming that all the surviving events after cleaning are actual images of extendend air showers.

- **Hadrons**: this component is the most abundant given that hadrons are the largely dominant component of the CR. In general the algorithms performing the rejection of this background work considering the Hillas (Fig. 1.22) and Stereoscopic (Fig. 1.24) parameters. Because hadronic and gamma-ray showers are morphologically different, these parameters allow to discriminate the two. In the last years many algorithms that exploit all the images instead only the parameters have been proposed.

- **Electons and positrons**: These particle are generated both in hadron and in EM (see 1.15) showers. Showers initiated by $e^{\pm}$ in the CR are purely electromagnetic exactly as the gamma-ray initiated ones. Therefore there are intrinsically difficult to distinguish from gamma-ray initiated events and survive the classification based on Hillas parameters, as an isotropic background.

- **Diffuse gamma**: The sky is permeated with diffuse gamma rays mainly produced by non resolved sources (Fornasa and Sánchez-Conde (2015)). As in the case of electrons and positrons the showers are hardly indistinguishable from those produced by gamma emitted from the celestial sources of interest and again are istrotropic.

Morphology of the shower event

The typical morphology of showers with different progenitors are shown in Fig. 1.26.

- **Hadron shower:** (left panel) are typical characterized by multiple islands, the shower produce multiple sub-showers with respect γ-like. The main island presents an irregular shape with respect γ-like events with typical larger width. 1.22)

- **Muon event**(middle-panel) present the typical ring or arcs shape. These events are easy to separate because the shape is totally different than the gamma-like/hadron-like events.

- **Gamma-like shower** (right-panel) present a well defined elliptical shape, as shown in 1.25 in this category are included the events produced by the source of interest but also the diffuse gamma rays and the $e^{\pm}$.

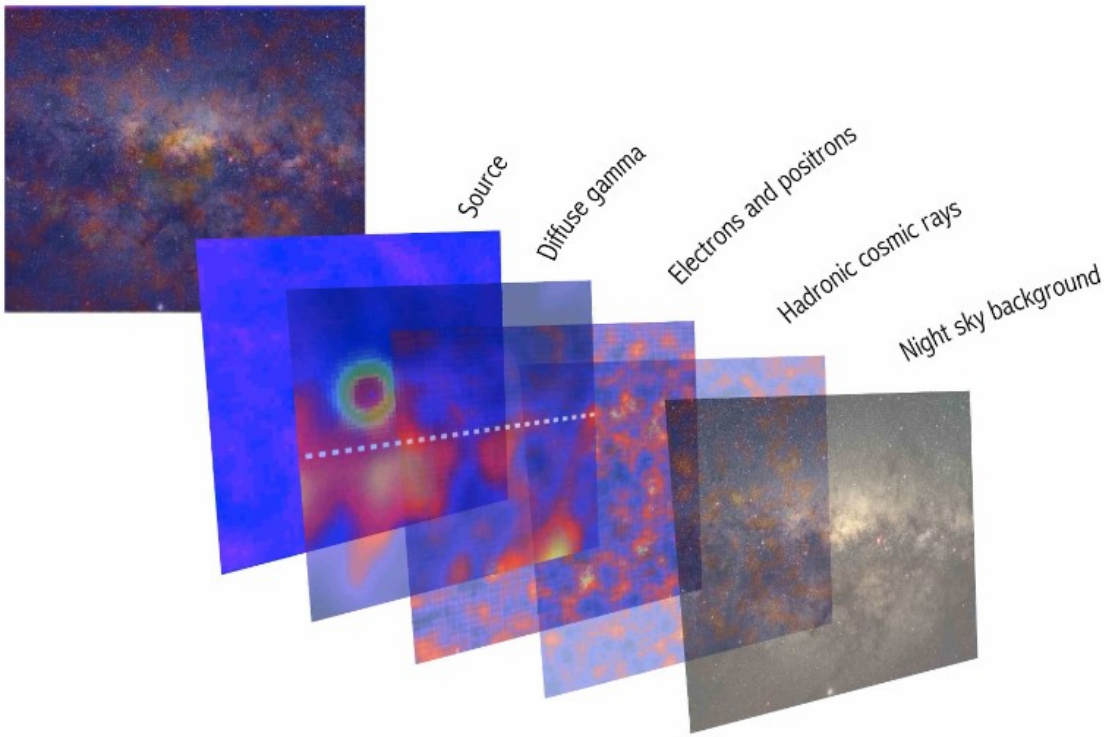


Figure 1.25: From Brun (2012) A schematic representation of the different layers of background for IACT telescope.

All of the previous showers are simulated, the images are easily recognizable. In real data samples the fraction of events for which the classification is non trivial, is large. In general, the background discrimination becomes easier with the energy of the events, given that events of higher energy produce brighter images with a better defined morphology.

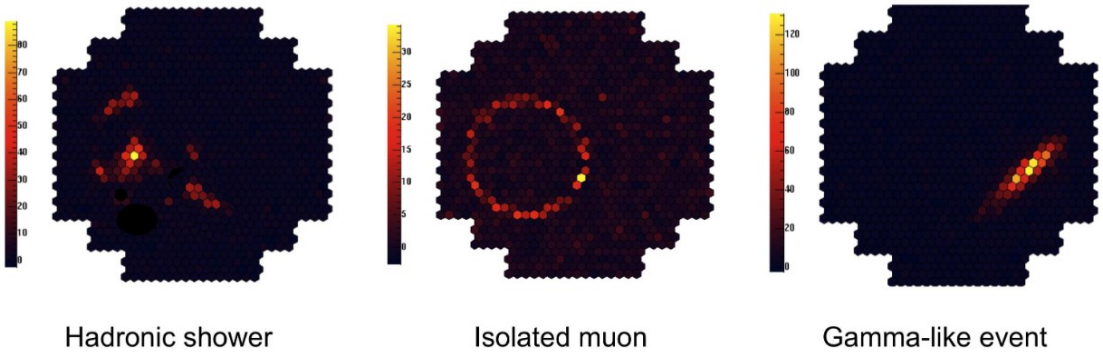


Figure 1.26: From Brun (2012) Simulated images of extended air shower events induced from cosmic particles, as observed in the focal plane of a Imaging Atmospheric Cherenkov telescope. From left to right: hadronic shower, muon, gamma-ray shower

Chapter 2

The MAGIC Telescopes

The Major Atmospheric Gamma Imaging Cherenkov telescope (MAGIC) is an Imaging Atmospheric Cherenkov Telescope (IACT) system composed by two telescopes operating in stereo modes (see fig: 2.1). The telescopes are situated at the Roque de los Muchachos observatory on the La Palma Canary island at 2200m of altitude and in first approximations they can be considered identical.

IACTs primary design goal is to reach the lowest possible energy threshold ($\sim$ 50 GeV), which means being able to detect the faintest showers light, this is achieved through fine pixelated cameras, fast sampling electronics and a large mirror area (Aleksić et al. (2016)).

In the case of MAGIC, the second design goal is a fast repositioning speed to capture transient events that occur quickly, including gamma-ray bursts. This is accomplished by using a lightweight ($\sim$ 70 tons) telescope construction built of reinforced carbon fiber tubes.

The main components of the MAGIC telescope are listed briefly in the following:

- The **Mirror surface**: it has a 17 m diameter and is made square, all-aluminum mirrors (964 panels 50 cm $\times$ 50 cm for MAGIC-I and 247 panels 1$\times$ 1 m for MAGIC-II), forming a parabola with a surface area of 236 m^2 (Doro et al. (2008)). A honeycomb structure that provides rigidity, high temperature conductivity, and low weight is sandwiched between two thin aluminum layers to create each mirror. Each raw blank has its surface diamond milled for excellent reflectivity and a slightly variable focal length to meet the reflector's overall parabolic form.

- The **Camera**: it is positioned in the reflector's focus, about 17 meters above the mirror dish. The MAGIC camera has a field of view (FoV) of 3.5° and is outfitted with 1039 evenly sized 0.1° pixels organized into 169 cluster modules. The trigger region includes 91 cluster modules covering a FoV of

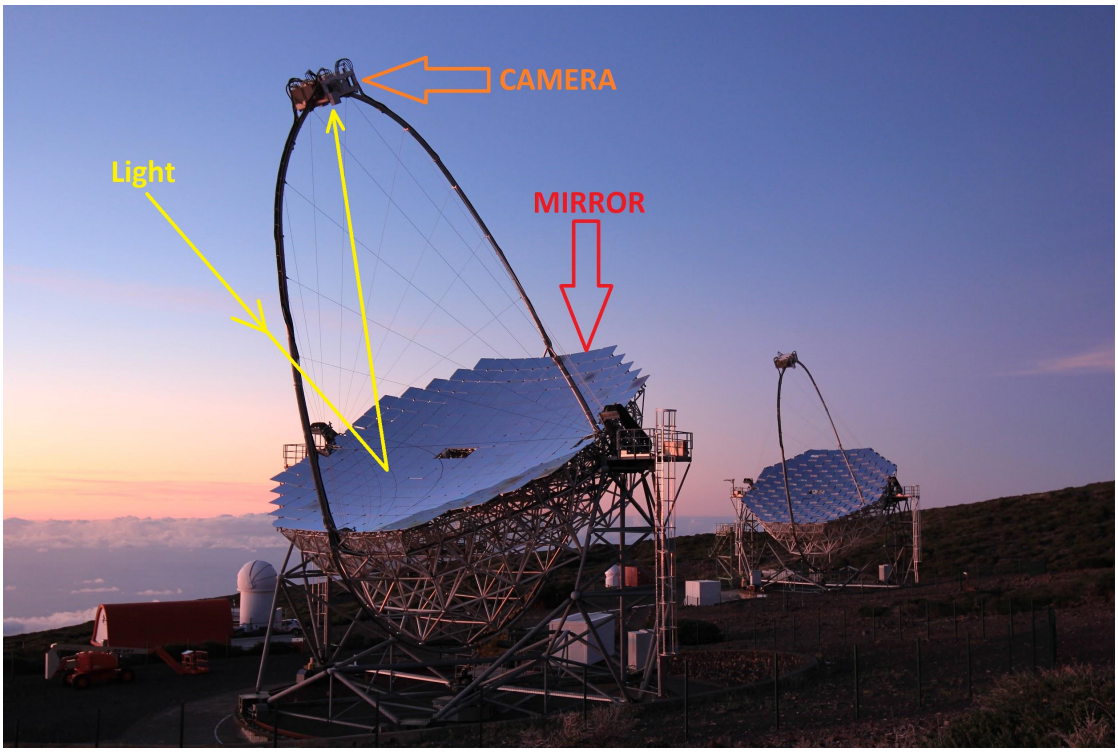


Figure 2.1: The MAGIC II (M2 left) and MAGIC I (M1 right) telescopes. Picture Credits: Derek Strom, Giovanni Ceribella and the MAGIC Collaboration. On MAGIC II the camera, the mirrors and the path of Cherenkov light are shown.

2.5° diameter. Each cluster consists of seven hexagonally arranged pixels. The camera can be controlled and maintained more easily because to the modular design. In order to reduce light losses caused by the dead space between the PMTs and to reject background light with a wide angle, the PMTs are coupled to Winston cone-type light guides, hexagonal in shape for a compact arrangement (Borla Tridon et al. (2009)).

- The **Readout**: it is situated in a designated counting house, where the camera's optical fiber-transmitted signals are received. The two primary components of the readout electronics are the receiver boards and the digitization electronics. Since Cherenkov flashes only a few ns, a fast sampling speed is required to improve the signal to noise ratio and fully utilize the arrival time information. The MAGIC readout has been originally designed for sampling the signals at 2 GSamples/s. Aleksić et al. (2016)), in the last years the sampling rate has been reduced to 1.64 GSamples/s.

- The **Trigger**: it is also located in the counting house, next to the readout electronics, to which it provides the necessary signals to perform the readout of the original signals from the PMTs and convert them to digital data.

  The trigger system is divided in three levels whose main functions are:

  - **Level 0 (L0)**: it is a amplitude discriminator operating on each pixel individually. If the analog signals exceeds a programmable threshold, a digital signal of fixed length is generated.
  - **Level 1 (L1)**: it is arranged in 19 macrocells of 36 pixels each (see Fig:2.3), with a partial overlap. A next-neighbor logic of 2, 3, 4 and 5 contiguous pixels is implemented in each macrocell (referred as 2NN, 3NN, 4NN and 5NN pattern logic). Only one logic pattern can be selected at observation start. If any of the 19 microcells triggers the programmed logic, a L1 trigger signal is generated and transmitted to the next trigger level.
  - **Stereo trigger (L3)**: the two L1 trigger signals are stretched to 100 ns and delayed according to the pointing direction of each telescope to take into account the differences in the arrival times of the Cherenkov light at the corresponding focal planes. A logical "AND" operation is made between the two signals, and the resulting signal (L3 output) is sent back to the individual telescope readout.

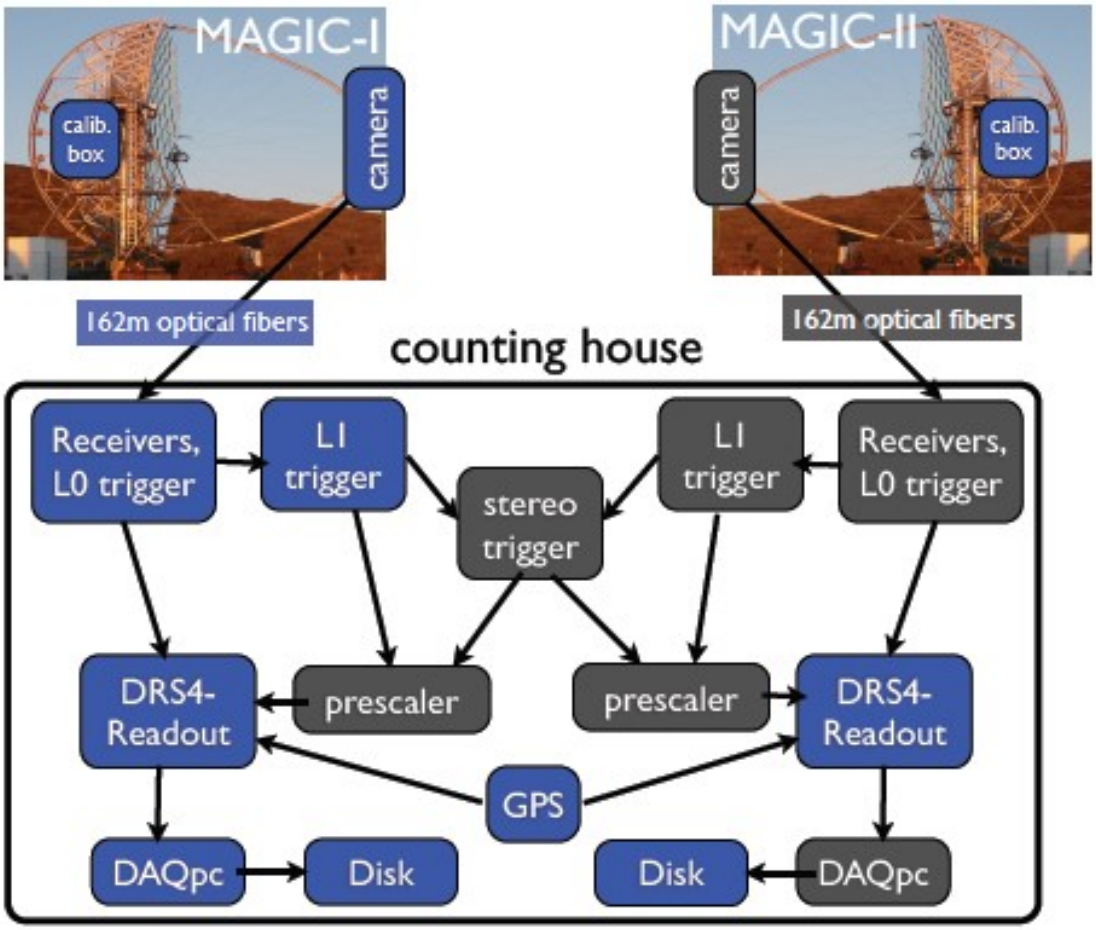


Figure 2.2: From Aleksić et al. (2016) Schematic view of the readout and trigger chain of the MAGIC telescopes.

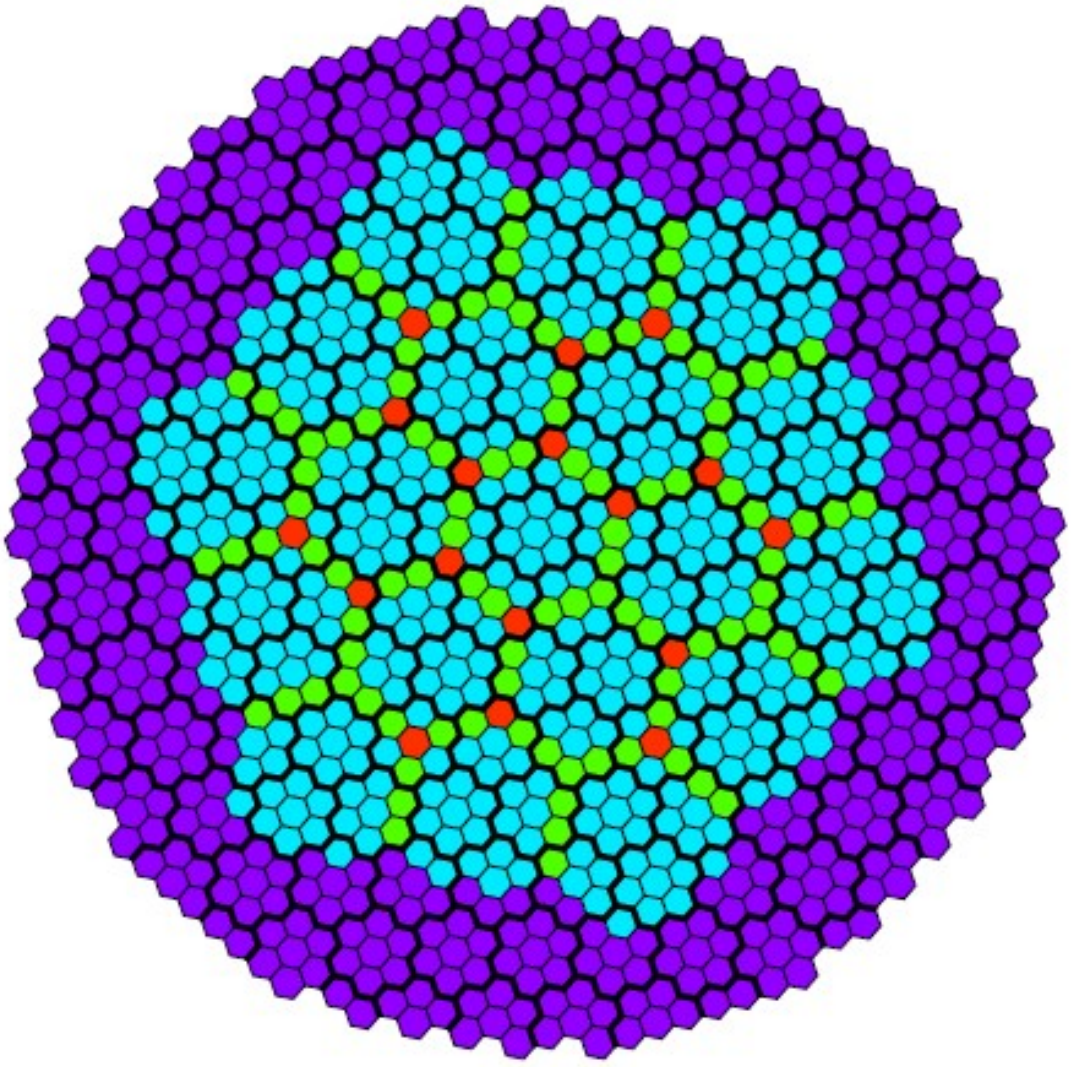


Figure 2.3: From Aleksić et al. (2016), Geometry of the MAGIC camera. Shown are the 1039 pixels arranged in clusters of 7 pixels each. The L1 macrocells (36 pixels each) cover the inner part of the camera. Pixels can be shared by two macrocells (green pixels) or three macrocells (red pixels).

2.1 MAGIC Data processing

The MAGIC "standard" data analysis (Zanin et al. (2013)) is done with the MARS software, a ROOT-based framework written to analyze the MAGIC data from low to high level where the events with all their estimated properties are reconstructed. A scheme of the analysis pipeline is shown in fig: 2.4.

In the following sections the most important blocks of the analysis pipeline are described.

2.1.1 Raw data

The data directly acquired by the telescope are referred as "raw data", that is the un-treated information from the PMTs as they are recorded by the readout system without any significant data manipulation besides some baseline subtraction and correction procedures (cell-wise pedestals and time-lapse correction). (figure: 2.2). For more information see Aleksić et al. (2016).

2.1.2 Calibrated data

The main quantities that must be extracted from the raw data are the charge and arrival time of the signal. The first step is the estimation of the baseline from the PMT waveform, in figure: 2.5 it is shown an example of the light pulse with the associated baseline, taken from dedicated pedestal data runs.

The area below the pulse signal and above the baseline represent the PMT charge. The main method to extract the baseline in MAGIC is by taking the mean of the waveform interval that does not contain a pulse or spurious signal (e.g. the ringing shape in figure).

The pulse area is obtained with technique called "sliding window". This technique takes a window of k slices (out of n slices) and the window area is calculated as the k slices are moving along the full waveform. The largest calculated value is taken as the light pulse area.

Before being processed by the analysis chain, the raw data need to be calibrated. This process is performed by the `sorcerer` executable in fig:2.2. The data from the PMTs are converted from digital counts to the equivalent number of photoelectrons (N_{phe}) representing the real charge collected by the PMT. The technique used to uniform the camera output is called "flat-fielding", see Gaug (2006) for more information.

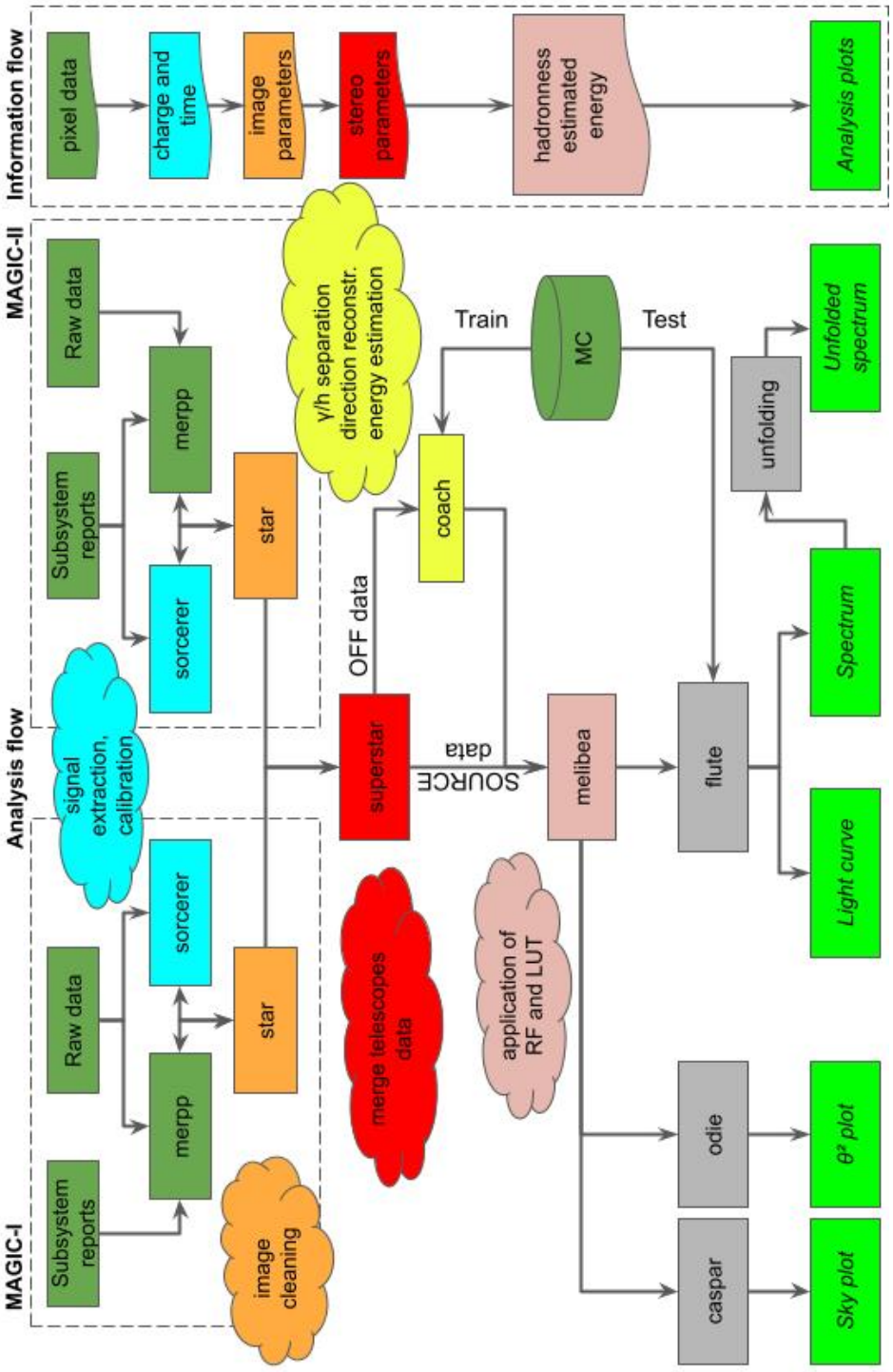


Figure 2.4: Scheme of the MAGIC analysis pipeline. On the left side, the analysis flow is shown, where data from the single telescopes are processed and later merged for the final analysis. On the right side, it is shown the corresponding data transformation, from raw data to high level information used in the analysis plots. Details on each analysis step are provided in the text.

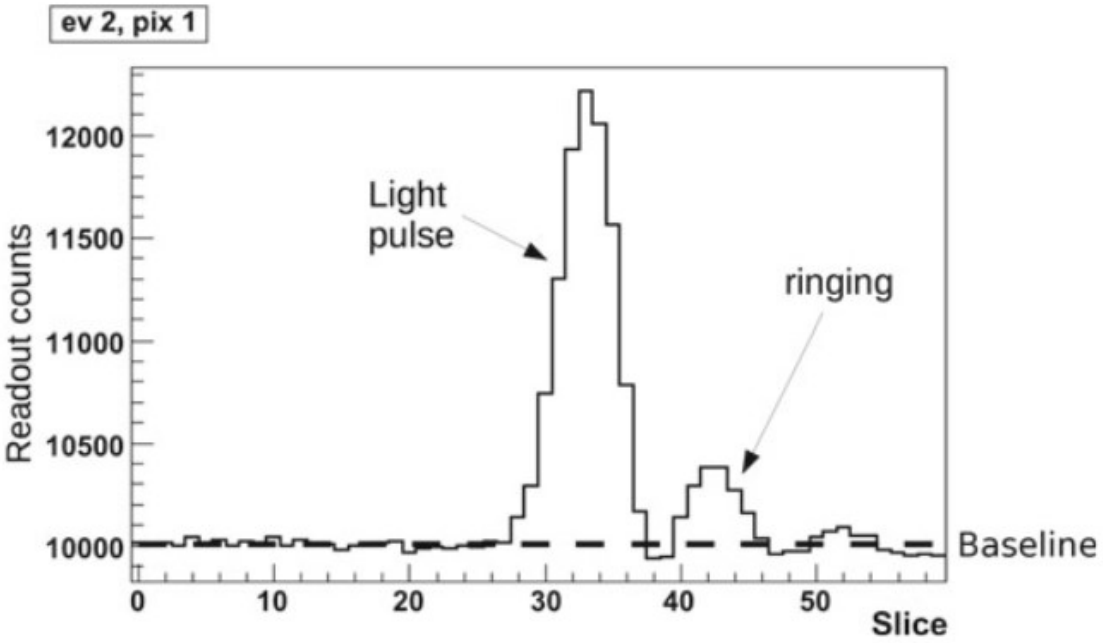


Figure 2.5: From Carreto Fidalgo (2019) A typical PMT signal digitized by the readout. The readout counts are plotted versus the readout slices, which correspond to the switching capacitors. The so-called ringing after the light pulse is an artifact of the readout. Besides light pulses, the readout can also be triggered by the so-called afterpulses from the PMTs, which are large amplitude signals caused by an ion accelerated back to the photocathode of the PMT.

2.1.3 Image cleaning

After the calibration procedure, the pixels data have been converted to meaningful information, charge and arrival time. In the following, we will refer to their collection as charge and arrival time "maps".

To extract information about the shower topology, the Hillas parameters (Hillas (1985)) are usually calculated. However, at this stage of the analysis, still some spurious signal is still present, due to the night sky or noisy pixels. It is therefore mandatory to perform a "cleaning" of the camera information in order to remove the pixels that are not associated with the shower light and may cause an incorrect estimation of the image parameters.

The description of the cleaning algorithm is important for this thesis work because the CNN classification is strongly influenced by the type of applied cleaning. A good cleaning algorithm should be a compromise between the simplicity and the effectiveness, in the following the two main cleaning methods used for the MAGIC data (Aliu et al. (2009)) are listed:

- **Absolute image cleaning:** This simple and robust algorithm has been used in MAGIC until 2007. It exploits only the charge value and it is based on two thresholds: Q_{core} and $Q_{boundary}$ (with $Q_{core} > Q_{boundary}$). All 2NN pixels with charge $N_{phe} \geq Q_{core}$ are marked as "core" pixels. The remaining pixels contiguous to core pixels with $N_{phe} \geq Q_{boundary}$ are also included as "boundary" pixels. The rest of the pixels are discarded. Typical threshold values are $Q_{core} = 10$ and $Q_{boundary} = 5$.

- **Time Absolute cleaning** This cleaning considers both the charge maps and the arrival time maps. It starts calculating the Q_{core} pixels from the charge maps like the previous absolute cleaning. These pixels must respect the Eq. 2.1 condition

$$t_{avg} - t_{core} \leq t_{arr} \leq t_{avg} + t_{core} \tag{2.1}$$

  where t_{avg} is the average arrival time, t_{core} is a time value arbitrary defined and t_{arr} is the arrival time of the signal.

  The next step is to select the pixels that are contiguous to core pixels and satisfy the Eq.: 2.2.

$$t_{arr-boundary} - t_{boundary} \leq t_{arr-core} \leq t_{arr-boundary} + t_{boundary} \tag{2.2}$$

  where $t_{arr-core}$ is the arrival time of the core pixel, $t_{arr-boundary}$ is the arrival time of the boundary pixel and $t_{boundary}$ is an arbitrary value with $t_{boundary} \leq$

t_{core}. The use of the arrival time maps allows to relax the charge threshold parameters $Q_{core/boundary}$ as well as to exploit the arrival time information.

Typical charge thresholds used in the standard analysis are $Q_{core} = 6$ and $Q_{boundaries} = 3.5$, while the time parameters values are $t_{core} = 4.5$ ns and $t_{boundary} = 1.5$ ns.

- **Sum image cleaning:** A new cleaning algorithm is being used, based on the sum of the pixels charge in different neighbor configurations (2NN, 3NN and 4NN). If the total charge exceeds a definite threshold and all the NN arrival time differences are within a given interval (depending on the NN combination) , the pixels are considered part of the shower image. This technique, together with signal clipping on the summed signal, is effective to reduce the effect of the PMT afterpulses and the night sky background effect on the camera. This cleaning is currently used for the MAGIC L3 trigger (stereo) observations.

  At the end of the cleaning procedure, the Hillas parameters are calculated to extract information about the image topology.

  The cleaning and Hillas extraction operations are performed by the `star` script, see fig:2.2.

2.2 Stereoscopic data merging

After the cleaning process, the data of the two telescopes are ready to be merged in one joint dataset with the additions of the main stereoscopic parameters. We review here the main variables that play a crucial role in the analysis, refer to fig: 2.6:

- Arrival direction: The arrival direction of the primary particle must be obtained from the images collected by the two MAGIC telescope. If the two images are plotted on the same plane (top portion of fig: 2.6), the arrival direction can be found by projecting the major axes and looking for an intersection, which is the shower incident angle or arrival direction.

- Impact point: next variable to calculate is the impact point, that is the position at ground where the shower axis intercepts the telescopes ground reference. By looking at Fig. 2.6, we see that, for each telescope, the image major axis on the camera and the arrival direction define a plane. The intersection of these two planes is the shower impact point at the ground reference.

- Shower maximum height:

  Last variable of interest is the shower maximum height ($Hmax$) that is the position where the shower development is maximum. In this case the procedure is more complex since we are going to look for an altitude figure based on partial and irregular sampling of the shower development. Since the image centroid defines the average angle of incident photons, by extrapolating back along the arrival direction line, the intersection should define the point of maximum height. An exact application of this method seldom results in a satisfying result, therefore the altitude of interest is defined as the point that minimizes the distance between the extrapolated lines from MAGIC-I, MAGIC-II and the reconstructed impact position. lines.

All of these parameters estimation can be improved if one knows the source position. In this case, instead of taking the image major axis, the line connecting the source position and the image centroid can be used.

Other variables of interest are the Cherenkov radius and the Cherenkov density. These are estimated assuming a Cherenkov emission from a single electron with energy equal to the critical energy (86 MeV) at the shower maximum height. The reconstruction of all these parameters in MARS is done with the script called `superstar` in Fig. 2.4.

2.2.1 Monte Carlo

The γ-ray shower events are typically overwhelmed by the hadron-like background, therefore discriminating the γ-ray events from a real data sample is challenging. Furthermore, even if one would successfully tag the γ ray event, some properties would remain unknown (e.g. the energy of the primary particle) because of the indirect nature of the Cherenkov technique that sample only the final development of the shower.

In general the γ-ray properties are estimated through simulation by a Monte Carlo (MC). Considering that all the MC properties are known, by associating showers with the best matching MC events , the unknown properties of the real events can be inferred from the similar MC.

For what concerns the hadron-like events, a good dataset can be obtained from the real data since it is possible for IACTs to monitor a sky region where no known γ-ray sources are present (so-called "dark patch"). The collected dataset in this condition is usually called "OFF data" and it is not expected to show an excess of gamma-like events. This solution is very effective since the simulation of hadron showers is a very time-consuming procedure, both in terms of computer elaboration time and resources.

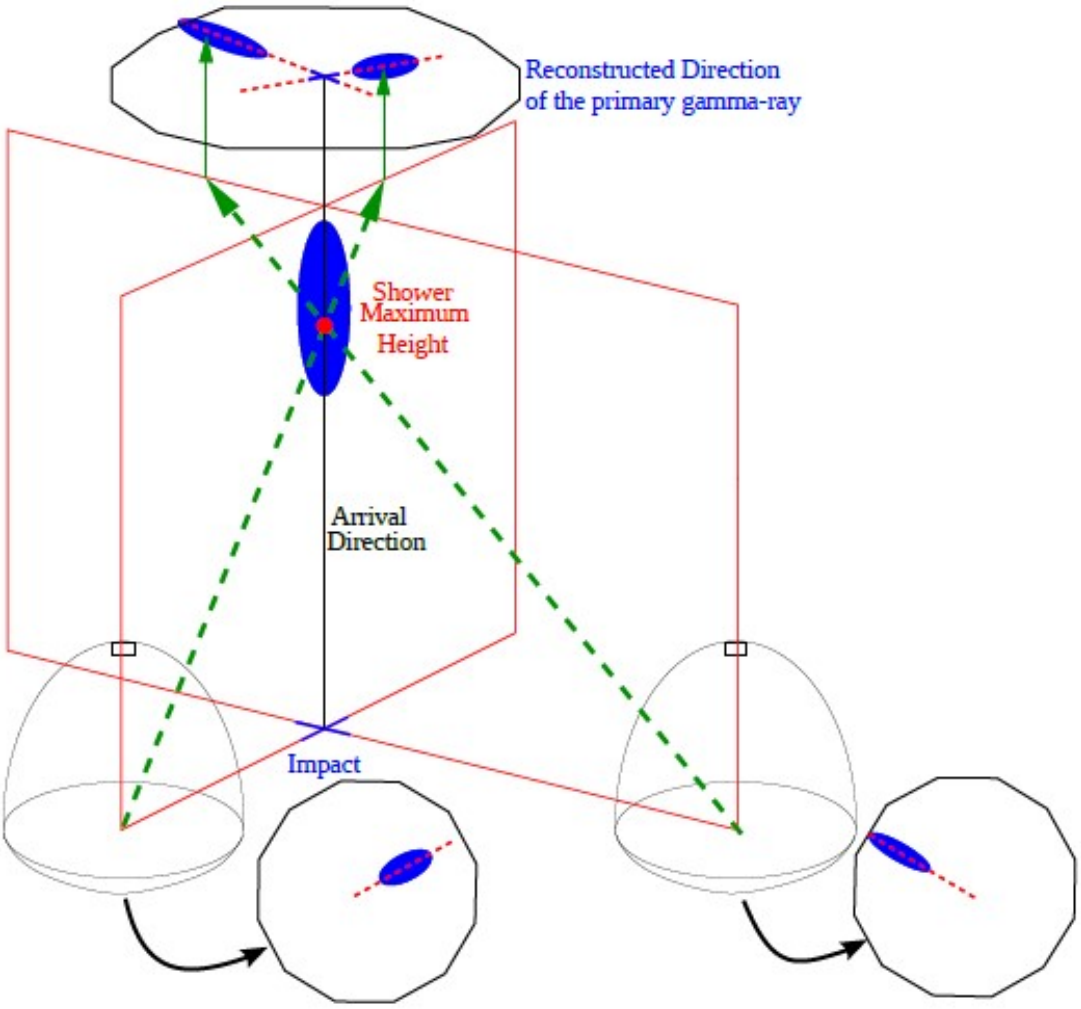


Figure 2.6: From Berti (2018) Example of the reconstruction of some stereo parameters.

This can be understood comparing the two pictures of the shower models: the EM shower (Fig:1.15) and the hadron shower model (Fig:1.16). The EM model is less complex than the hadron model but for the EM model are already needed good computing resources. MAGIC simulation uses the CORSIKA software (1.3.2) to simulate showers using input parameters like the primary particle type, energy and incident angle and several others parameters.

This part of analysis is performed by the `coach` script (Fig.2.4) and it runs parallel to the main analysis branch. Coach uses two sets of data: the simulated γ-ray shower events and, the dark patch shower events.

Gamma-hadron separation

Extensive air showers generated by hadrons are the majority of events and generate images that are typically more disperse than the ones generated by gamma rays. Especially at low energies, electrons produce showers that are very similar in shape to gamma-like images, as well as fluctuations of the signal due to background light, such as the night sky background, Moon light, or flashes from cars driving nearby the telescopes. Moreover, muons with ultra-relativistic velocities hitting near the telescope, generate rings or portion of rings that can mimic a low-energy shower.

Once the events have been calibrated and merged, they must be associated to a probability of coming from a hadron via a variable called "hadronness".

MAGIC uses a classification algorithm called Random Forest (RF) to separate hadron-like from gamma-like event images. As it will be exaplained in Sec. 2.2.1, the RF is a "supervised machine learning algorithm" that uses a model from a sample with known properties ("training sample") and applies it to a sample with unknown properties ("test sample") to predict the prabability that is a hadron-like event.

To do so, the Hillas and the Stereo parameters (section 2.2) of the MC and the dark patch described in Sec. 2.2.1 are used.

The Hillas and stereo parameters selected from the known samples are input in the RF algorithm that creates many conditional parameters trees. Each event is then input to a tree to obtain the classification result 0 if the event is gamma-like and 1 if it's hadron-like. This procedure is repeated for 100 trees to have a set of 100 events that result in a probability:

$$p_{h/\gamma} = \frac{\sum_{i=1}^{100} p_i}{100} \tag{2.3}$$

The sum of the results divided by the total number of trees gives the hadronness probability for the unknown event.

The RF training sample is selected by the analyzer, by choosing the correct MC [1] and the correct dark γ-ray sources belonging to the same time period

and similar Zenith degree angle. After the data samples have been selected the RF can be built. The calculation of the probability for the analysis sample is performed in the next analysis step by the Melibea script (Sec:2.2.1).

A detail description of the RF algorithm can be found in Albert et al. (2008a).

Energy reconstruction

The energy of the primary particle is proportional to the number of Cherenkov photons produced by the air shower, also proportional to the photoelectrons N_{phe} charge measured by the telescope (the Hillas parameter $size$). This is the main parameter to make a first estimation of the energy of the event. Others energy-related parameters are, for example, the shower maximum height, the impact, the source zenith and azimuth angles and several other variables.

Like for the gamma/hadron separation, the energy of the real events is unknown because only the final part of the shower is detected. It is possible to use the MC data and to associate an energy to each event by using a Look Up Table (LUT).

A multi dimensional table containing the mean energy of each MC events belonging to each parameter bin is built ($< E_{MC} > (Par_{bin})$). The energy of the real events is obtained by looking at the corresponding parameters MC bin from which the $< E_{MC} >$ that will be the estimated energy of the real event is obtained.

Position reconstruction

The calculation of the arrival direction of the shower (Sec.2.2) can be improved by using a method known as Distance between the Image centroid and the Source Position (DISP). This method was first used for MAGIC single telescope observations, (Domingo-Santamaria et al. (2005)), but can be adapted for stereo data. Actually MAGIC calculates the stereo DISP using an RF algorithm (Aleksić et al. (2010)).

Hadronness

The analysis process shown in 2.4 converges in the melibea script that uses the results of the MC coach stage described in the Sec. 2.2.1 together with the superstar data of the analyzed source to calculates the hadronness, the energy and the arrival position of the origin particle of the real events.

[1] the MC are rebuilt each times the state of the telescope changes, in particular are related with the states of the cameras and the mirrors

This part is strictly connected to the original work (Ch. 4) because the MARS hadronness is used to slightly bias the training set.

After this block, the data events can be considered high-level data because these events will contain the physical parameters from which their physical properties can be directly calculated.

2.2.2 Significance

The next step in the analysis chain is to quantify the gamma-ray signal coming from the source.

Before introducing the significance calculation, the observation techniques of MAGIC must be described (Bretz et al. (2005)):

- ON-OFF mode: the telescope is pointed so that the source is located in the camera center (ON position). After the data have been acquired, the telescope is moved to a new position (OFF position) towards a dark patch region in the sky and new data are recorded. Since both measurements are not done with the same instrument at the same time and the data taking conditions are different, scaling of the background measurement is necessary to achieve an agreement of the background levels of the ON and OFF-source observations. A drawback of this method is a low duty cycle because ON and OFF data must be recorded separately.

- Wobble mode: the camera is moved off-axis and the source is located on the side of the camera FoV (see fig: 2.7).

  In this case, the OFF data are coming from a camera region far from the source position and several OFF positions can be defined. Typically, in MAGIC the ON and OFF positions are positioned at a distance of $w = 0.4°$.

The wobble method is a clever technique because it maximizes the observation time and limits the systematic due to cameras dis-uniformity as well as increasing the duty cycle.

With these definitions, the statistical significance used in MAGIC is based on the LiMa formula (Li and Ma (1983)) can be written as:

$$\begin{aligned}\sigma_{Li\&Ma} =&\sqrt{2}\Bigg\{ N_{on} \cdot ln\left[\left(\frac{1+\tau}{\tau}\right)\left(\frac{N_{on}}{N_{on}+N_{off}}\right)\right]+ \\ &\frac{N_{off}}{\tau} \cdot ln\left[(1+\tau)\left(\frac{N_{on}}{\tau \cdot N_{on}+N_{off}}\right)\right]\Bigg\}^{\frac{1}{2}}\end{aligned} \tag{2.4}$$

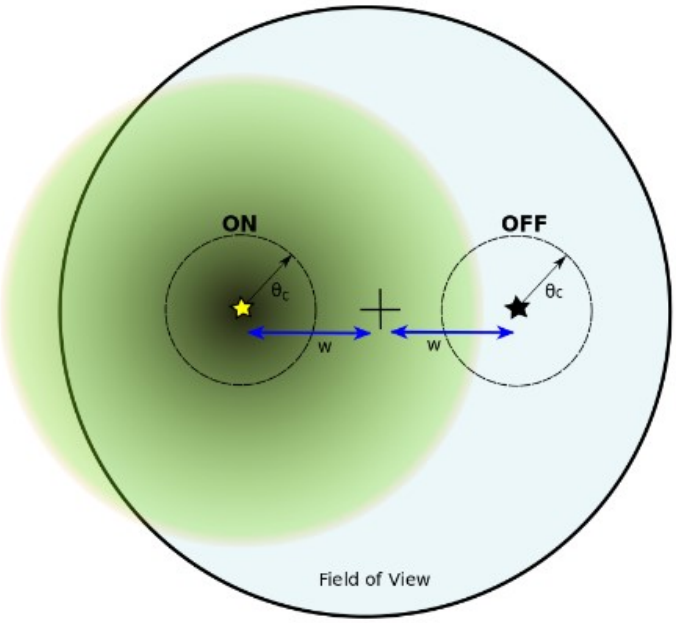


Figure 2.7: From Palacio, Navarro-Girones, and Rico (2019) Schematic configuration of the FoV during wobble mode observations. The telescope axis (black cross) has an offset distance w with regards to the center of the source under study (yellow star). Signal (ON) region is defined as a circle around the center of the source, with angular size θ_c. One background control region (circular region around OFF, black star) is defined with same angular size, symmetrically with regard to the signal region. The leakage effect is schematically shown where, for moderately extended source (green area), signal events are also expected to be reconstructed inside OFF.

where N_{on} is the number of gamma-like events towards the source direction, N_{off} is the number of gamma-like events towards the OFF direction and τ is a time normalization value to balance the contribution of the ON and OFF events.

The excess of events can be plotted in the θ^2 space, defined as the radius square of a circle on the camera FoV. If we take a circle of radius ($\theta = \sqrt{\theta^2}$) around the ON position and a similar circle around the OFF direction(s) we can plot the DISP (Sec. 2.2.1) in bins of θ^2. An example of θ^2 plot of a source is shown in Fig.2.8, the top plot uses a cut $\theta^2 \leq 0.15$, $size1/2 > 60$ and no hadronness cut, while the bottom plot uses the same cuts for θ^2, $size1/2$ and hadronness ≤ 0.28. In this example, the ON part shows an excess of events with respect to the three other background areas and, above $\theta^2 \sim 0.08$ value the ON and OFF backgrounds are similar.

Name	Had $< cut$	$\theta^2 < cut$	$cut <$size1&2	$cut < E_{est}$
cut-Low Energy (LE)	0.28	0.02	60	—
cut-Full Range (FR)	0.16	0.009	300	—
cut-High Energy (HE)	0.1	0.007	400	1000

Table 2.1: MAGIC standard cuts used for odie source evaluation

In the MAGIC analysis there are three standard settings that are used: the Low energy (LE), High energy (HE) and Full Range energy (FR), listed in table: 2.1. These cuts are used both for ON and OFF data, each time changing the center of the circle as shown in Fig.2.7. The ON data are centered on the source and the OFF data are centered on the OFF circle(s).

The events in the ON circle (N_{on}) and the mean of the events in the OFF circles ($< N_{off} >$) are computed and the number of excess events (N_{exc}):

$$N_{exc} = N_{on} - < N_{off} > \tag{2.5}$$

The MAGIC script used to calculates the θ^2 plot, the standard significance and Li&Ma significance is called `Odie` (see fig. 2.4).

2.2.3 Spectral Energy Distribution

The last step of the standard analysis consists in calculating the spectral energy distribution 2.2.3 and the light curve.

In the analysis block scheme (fig. 2.4) this operation is performed by the `flute` script. This script will not be described in detail because is not the main topic of this work.

The energy spectrum generated of the EM emission produced by an astrophysical source and observed by different telescopes can cover various energy decades for example from radio to gamma rays. An usefull and significant physical representation of this emission as a function of the energy is called Spectral Energy Distribution (SED). SED plot (an example is showed in fig:2.9). The SED function can be expressed as:

$$J(E_\gamma) = \int_{E_1}^{E_2} E_\gamma \cdot \frac{dN_\gamma}{dE_\gamma} dE_\gamma \quad \text{units: erg cm}^{-2}\,\text{s}^{-1} \tag{2.6}$$

Where E_γ is the energy of the arriving photon and the $N_\gamma = n_\gamma/(T \cdot A)$ is the number of photon on unit of time and area.

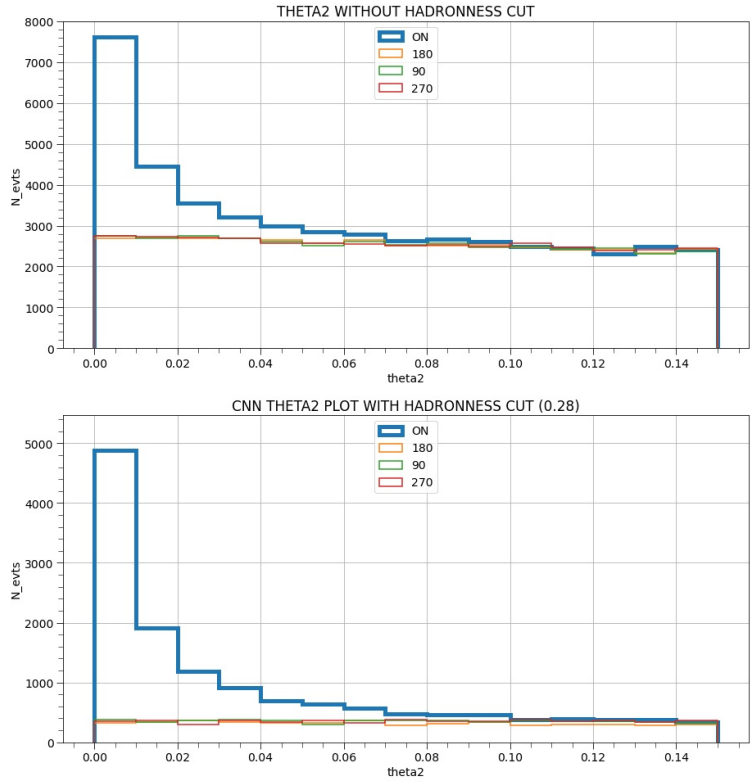


Figure 2.8: Example of θ^2 plot with Mrk421 cross-check sample (Sec.4.5). In the plots are represented the ON (Blue) and 3 OFF (180 90 and 270 others three color) positions using the¿ cut $\theta^2 \leq 0.15$, $n_{bins} = 15$, $size1/2 < 60$. Top no hadronness cut, down hadronnes cut $h \leq 0.28$.

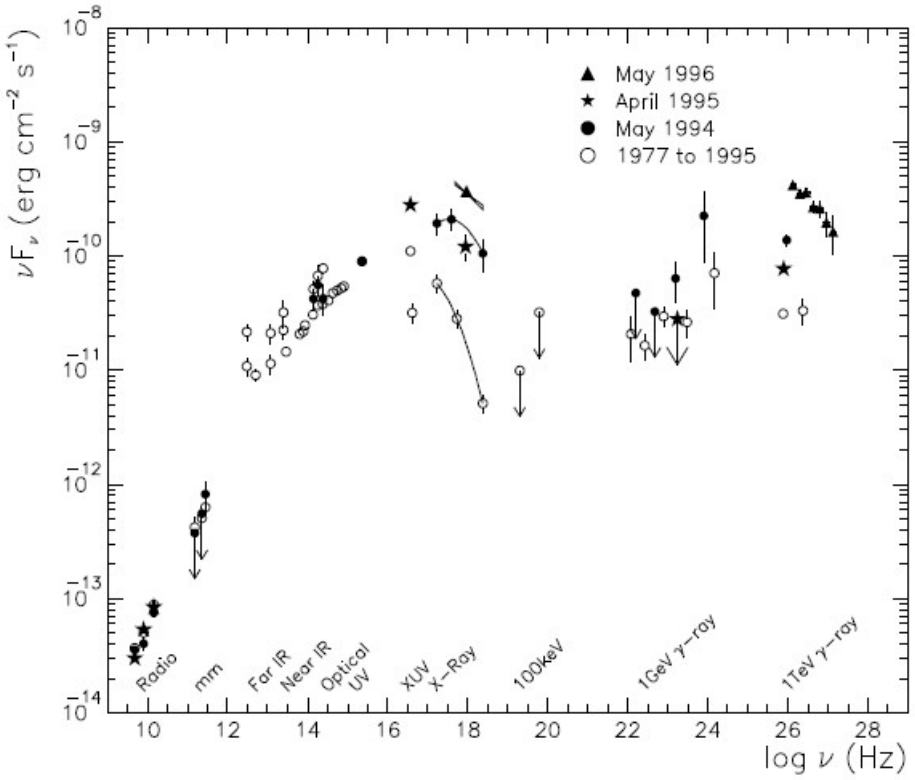


Figure 2.9: From Buckley et al. (1997) the experimental multiwavelength SED points for Mrk421 blazar. On the y-axis is reported the energy on area and time units, on the x-axis the specific flux density multiplied by the energy of the photon emitted

Light Curve

The Light Curve (LC) is a plot of the light intensity as a function of time. The LC is used to study the changes in astronomical sources over time. Together with the SED, the LC is very useful to infer some physical properties of the sources: for example in pulsars due to the rotation the LC is periodical (E.g. Abdo et al. (2010)). An example of periodical LC of the CrabNebula pulsar is shown in Fig. 2.10.

2.3 MAGIC performances

In this section the current performances of MAGIC stereo system will be presented (Aleksić et al. (2016)). The performances have been evaluated by the MAGIC

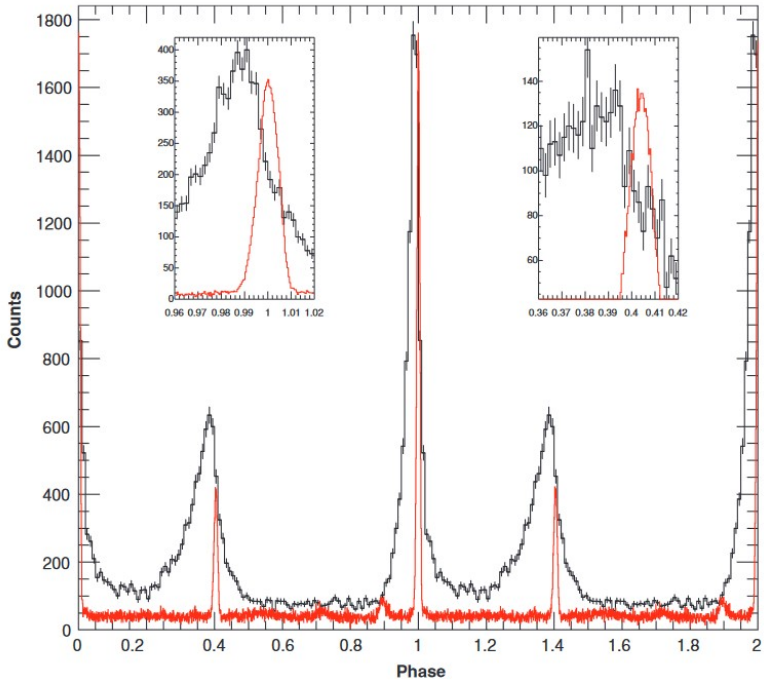


Figure 2.10: From Abdo et al. (2010). Light curve obtained with photons above 100 MeV. The light curve profile is binned to 0.01 of pulsar phase. Insets show the pulse shapes near the peaks, binned to 0.002 in phase. The radio light curve (red line) is overlaid (arbitrary units). The main peak of the radio pulse seen at 1.4 GHz is at phase 0. Two cycles are shown.

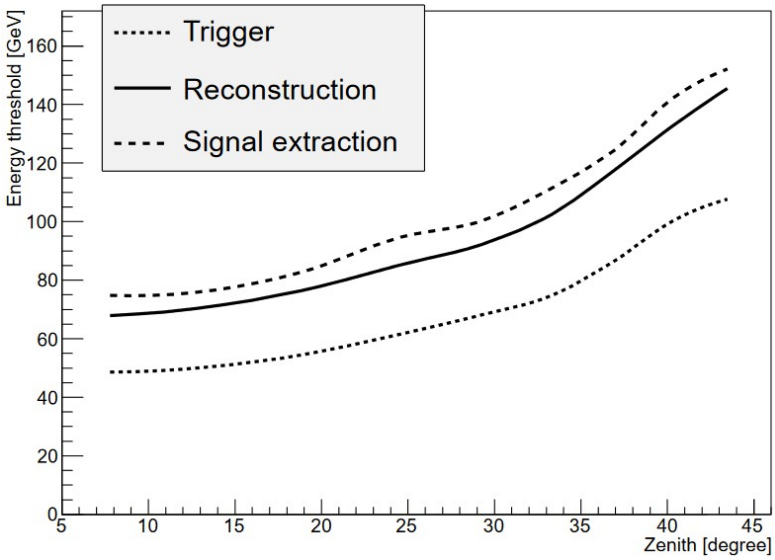


Figure 2.11: Energy threshold of the MAGIC telescopes as a function of the zenith angle of the observations Aleksić et al. (2016) . Long dashed curve: threshold at the trigger level. Solid line: only events with images that survived image cleaning in each telescope with at least 50 phe. Dashed line: with additional cuts of $Hadronness < 0.5$ and $\theta^2 < 0.03^{\circ 2}$ applied

collaboration using Crab Nebula data samples taken at low and medium zenith angles and MC simulations.

2.3.1 Energy Threshold

The energy threshold of an observation is defined in Aleksić et al. (2016) as the peak of the energy distribution of simulated MC data for a source with a power law spectrum with spectral index of 2.6. This parameter is very important because allows to understand which will be the minimum energy that can be measured with MAGIC. The energy threshold is different for each observation because depends on the trigger and is influenced by the zenith angles of the obsrvation and by the spectrum of the source. The energy threshold of the MAGIC telescope as a function of the zenith observation is shown in figure 2.11. The Zd influence on the energy threshold is due by the fact that the shower has to go through a thicker layer of atmosphere than a shower that comes from a low zenith. This raises the energy threshold of the high Zd observations.

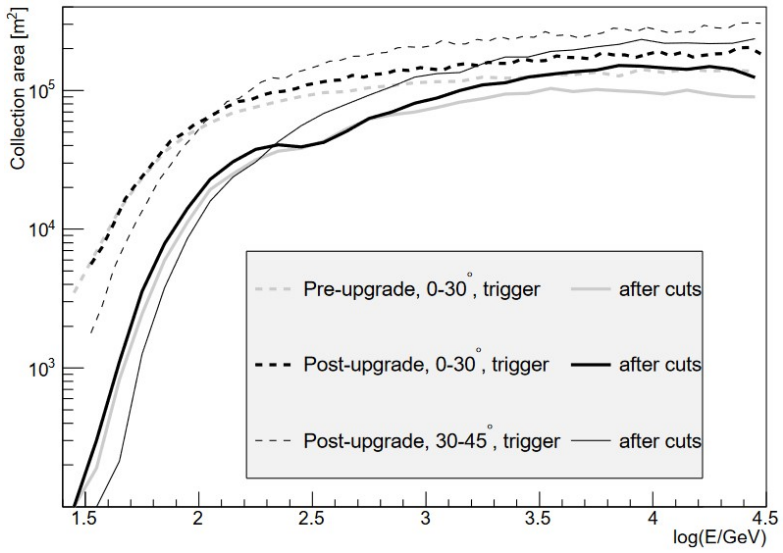


Figure 2.12: From Aleksić et al. (2016): Collection area of the MAGIC telescopes after the upgrade at the trigger level (dashed lines) and after all cuts (solid lines). Thick lines show the collection area for low zenith angle observations, while thin lines correspond to medium zenith angle. For comparison, the corresponding pre-upgrade collection areas are shown with gray lines

2.3.2 Effective collection area

The effective area is literally the instrument geometrical area where the gamma rays can be detected by the instrument, for small array of telescopes, like MAGIC, the collection area is mainly determined by the size of the Cherenkov light pool, this depends on the Zd (Sec.1.3.5), at low Zd The typical Cherenkov light pool size is $\sim 120m$ (Aleksić et al. (2016)). To calculate the effective area in MAGIC one can apply the definition of the effective collection area:

$$A_{eff} = \frac{N(E)}{N_0(E)} \pi r_{max}^2 \tag{2.7}$$

where r_{max} is a simulated value called maximum simulated shower impact. $N_0(E)$ is the number of simulated events and $N(E)$ is the number of events that surviving the trigger condition or a given set of cuts. In Fig. 2.12 are shown the collection area functions based on periods, Zd, trigger and before and after the cuts.

2.3.3 Sensitivity

Another important value used to evaluate the telescope performance is the sensitivity, this parameter represent the minimum signal that can be detected in 50

hours with 5σ. The significance σ using this time the Gaussian significance:

$$\sigma = \frac{N_{ex}}{\sqrt{(N_{off})}} \tag{2.8}$$

where N_{ex} is the number of excess $N_{ex} = N_{on} - N_{off}$. If one consider a Crab Nebula observed for a time t the significance for a time $t_0 = 50h$ can be write:

$$\sigma(t_0) = \sqrt{\frac{t_0}{t}} \frac{N_{ex}}{\sqrt{N_{off}}} \tag{2.9}$$

And then the sensitivity with a minimum signal of 5σ in $50h$ is defined as:

$$Sensitivity = \frac{5\sigma}{\sigma(50h)} = 5\sigma \frac{\sqrt{N_{off}}}{N_{ex}} \sqrt{\frac{t}{50h}} \tag{2.10}$$

Beside the performance calculations that is useful to compare the MAGIC telescope with other instruments the sensitivity can be used to estimate the flux that one can detect for a given observation time.

The MAGIC sensitivity curves (Aleksić et al. (2016)) are shown in Fig. 2.13: on left is shown the integral sensitivity that is calculated searching the highest sensitivity value that can be obtained applying a set of cuts (hadronnes,size, θ^2, etc...) for events that are above the energy threshold (2.3.1) of the telescope. On right the is shown differential sensitivity that is the highest sensitivity in a energy bin (or in a set of energy bins) applying a set of cuts.

2.4 Standard Analysis of MAGIC Sources

In the following sections are present the MARS Standard Analysis (SA) results for some reference sources.

2.4.1 Monte Carlo and OFF data selection

The MC and OFF data were selected (González Muñoz (2015)) with the parameters shown in tables: 2.2 and 2.3.

The source 1ES0927+500 in the table represents a little issue because the standard analysis of this source has been performed with RF matrices obtained using, as off data sample, a small part of the source data sample itself. In general this is not a good practice because the analysis might be biased. However the 1ES0927+500 data sample used to train the RF is very small in comparison to the overall off data sample.

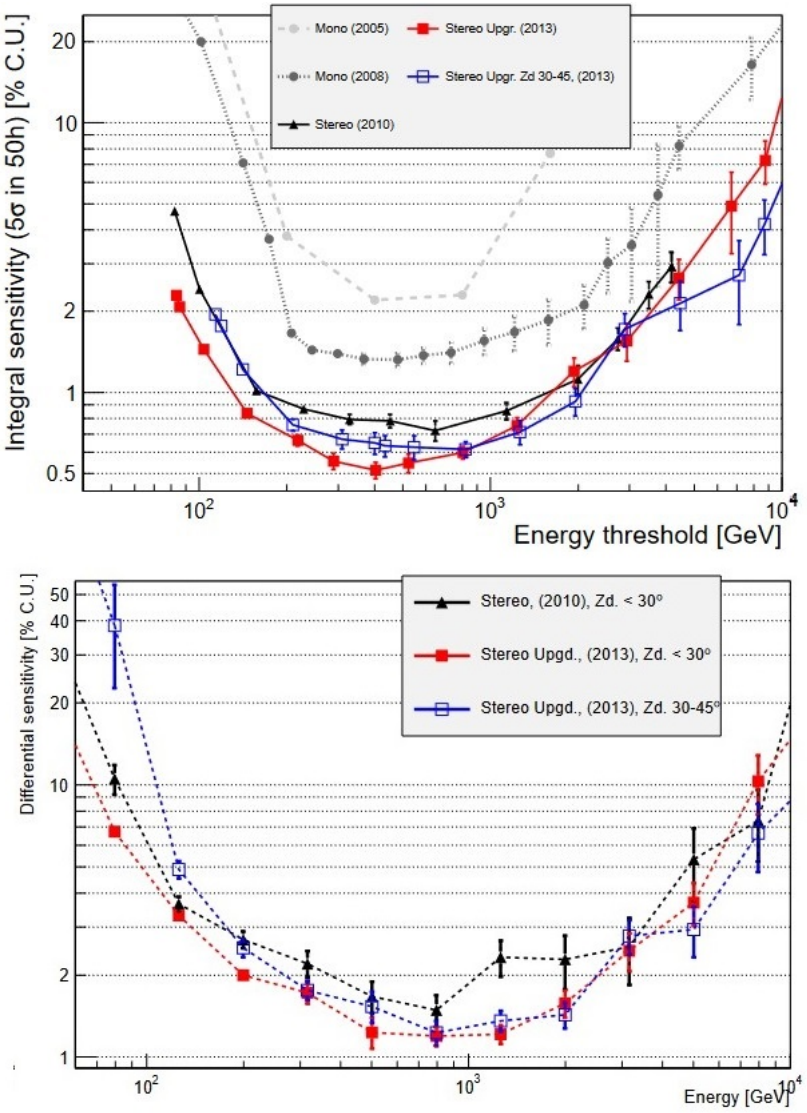


Figure 2.13: Top panel: integral sensitivity of the MAGIC telescopes as a function of the energy and for different periods (colors). Bottom panel: differential sensitivity of the MAGIC telescopes as a function of energy and for different periods (colors). Grey and dark grey dots: Mono (single telescope) sensitivity. Black triangles: stereo (both telescopes) with old MAGIC-I camera. Red and blue squares: stereo sensitivity after upgrade of 2012-13 (Aleksić et al. (2012)) at zenith lower than 30° and between 30° and 40° respectively.

Name	dates covered	zenith	type
ST.03.02	2013-01-18 to 2013-07-26	05-35	ringwobble
ST.03.02	2013-01-18 to 2013-07-26	35-50	ringwobble
ST.03.02	2013-01-18 to 2013-07-26	50-62	ringwobble

Table 2.2: MC data used for the standard data analysis

Source name	Date
SegueA	2013-01-15
	2013-01-16
	2013-01-17
	2013-01-18
	2013-01-19
	2013-01-20
	2013-01-21
DarkPatch11	2013-02-08
DarkPatch28	2013-01-21
DarkPatch32	2013-02-11
DarkPatch33	2013-02-10
	2013-02-15
Geminga	2013-02-12
GRB130504	2013-05-04
1ES0927+500	2013-04-02

Table 2.3: List of the off sources used as hadron sample for RF construction (González Muñoz (2015))

2.4.2 Crab Nebula

The Crab Nebula (Hester (2008)) is a very famous supernova remnant with a pulsar in the center. The Crab Nebula is the reference source for the γ ray astronomy and is important because is very easily detectable at VHE. Its luminosity and SED are of high intensity and nearly constant. Furthermore, observing the Crab Nebula one can study an astrophysical system composed by four observable components: the Pulsar, the Pulsar's wind, the thermal filaments, the freely expanding ejecta.

The Crab Nebula is therefore used to test the goodness of other analyses provided that the examined periods are close and similar observing conditions. The Crab SED is also used to verify for consistency.

Crab Nebula cross-check analysis

For the Mrk421 cross-check analysis, the following days of Crab nebula observation were selected: 2013-02-10, 2013-03-12, 2013-03-15.

The obtained SED is shown in Fig. 2.14 and the SED is in a good agreement with the ones already published (Aleksić et al. (2015), Aleksić et al. (2016), Albert et al. (2008b) and Aharonian et al. (2006)).

The other part of check regards the sensitivity obtained after the Full-Range cut (Tab. 2.1), this should be around 0.7% Crab Units (Aleksić et al. (2016)), in Fig. 2.15 for low zenith (Zd= 0-50) the value is: $0.72 \pm 0.03\%\ CU$.

2.4.3 Markarian 421

Markarian 421 (Mrk421) (Sec.1.2.3) is a High Frequency BL Lac (HBL) object. The SED is shown in Fig. 1.12. Its SED has the typical blazar shape, composed by the Synchrotron and the Inverse Compton bumps.

The Mrk421 is one of the closest Blazars to Earth (redshift=0.031) and one of the brightest. All these reasons makes it one of the most observed sources after the Crab Nebula because can be detected easily even during the non-flaring state (Aleksić et al. (2015)).

In many papers the multiwavelength SED is studied as well as its variability (see e.g. The MAGIC Collaboration et al. (2022)). This source is a very good prototype to study the emission mechanisms of the Blazars class.

In the next sections, the standard analysis of two data samples of the Mrk421 will be shown. The first sample belongs to the extraordinary flare of April $10^{th} - 16^{th}$, 2013 (González Muñoz (2015), Acciari et al. (2020)) while the second is another flare of Mrk421 happened ten days before the first sample. They will be later used in chapter 4 as main and cross-check samples.

Standard Analysis of Markarian 421 flare

Since this sample will be used later as main dataset, it is necessary to check that the sample consists of "good" shower images. The quality of the dataset is done by using the standard analysis, starting from superstar (fig. 2.4).

The April $10^{th} - 16^{th}$, 2013 period has been selected, discarding the moon and the twilight data ($\sim$30%) and keeping only the dark sample. All of the available Zd angles ($5 \div 62$) were kept to have high statistics and to increase the variability range.

The standard analysis results can be seen in Fig. 2.16 where the comparison using the Full Range cut (tab. 2.1) with González Muñoz (2015) is shown. In Fig.

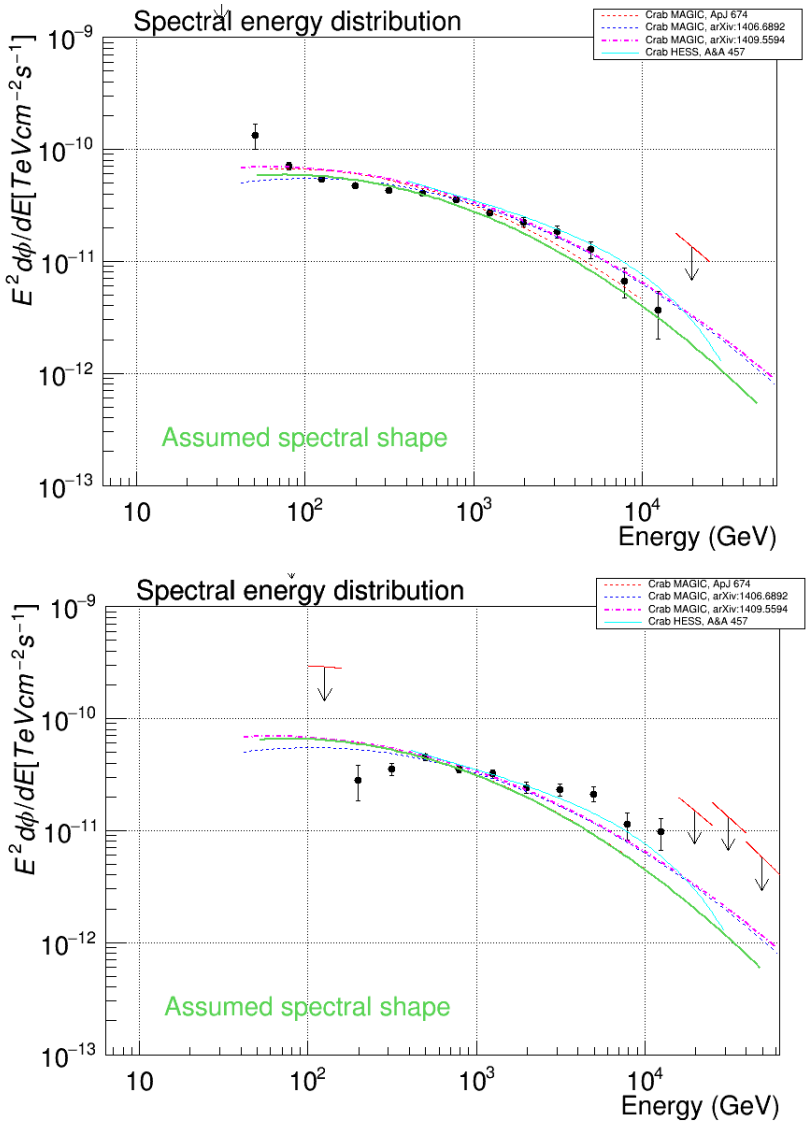


Figure 2.14: CrabNebula SEDs. Top panel: SED using Low Zenith data ($Zd = 0 \div 50$); Bottom panel: High Zenith data SED ($Zd = 50 \div 62$).

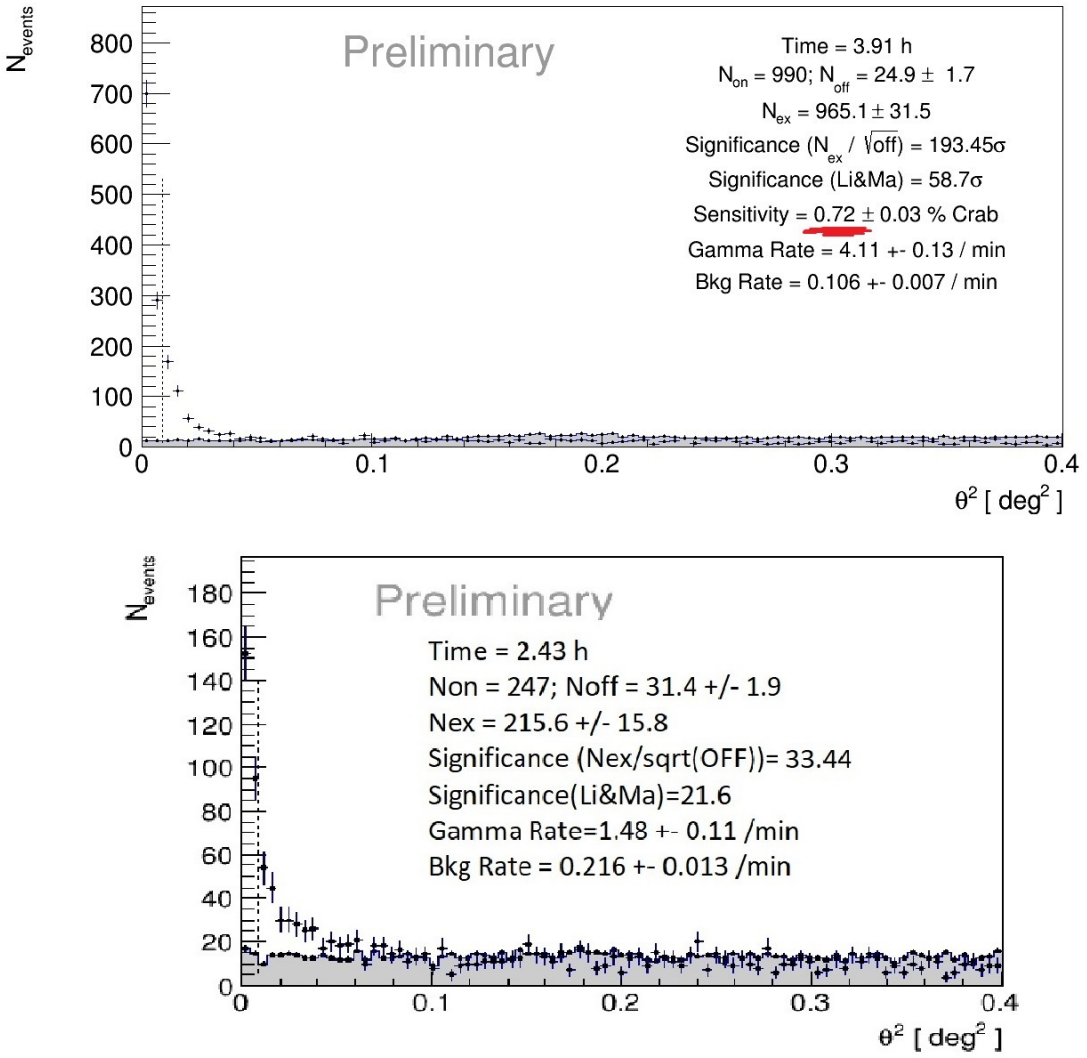


Figure 2.15: θ^2 plot for Crab nebula Full Range Energy (Table 2.1). The dotted line near the y axis represents the θ^2 cut. (Top) Low Zenith data ($Zd = 0 \div 50$) with the sensitivity near 0.7% like in standard analysis (Aleksić et al. (2016)). (Bottom) High Zenith data ($Zd = 50 \div 62$).

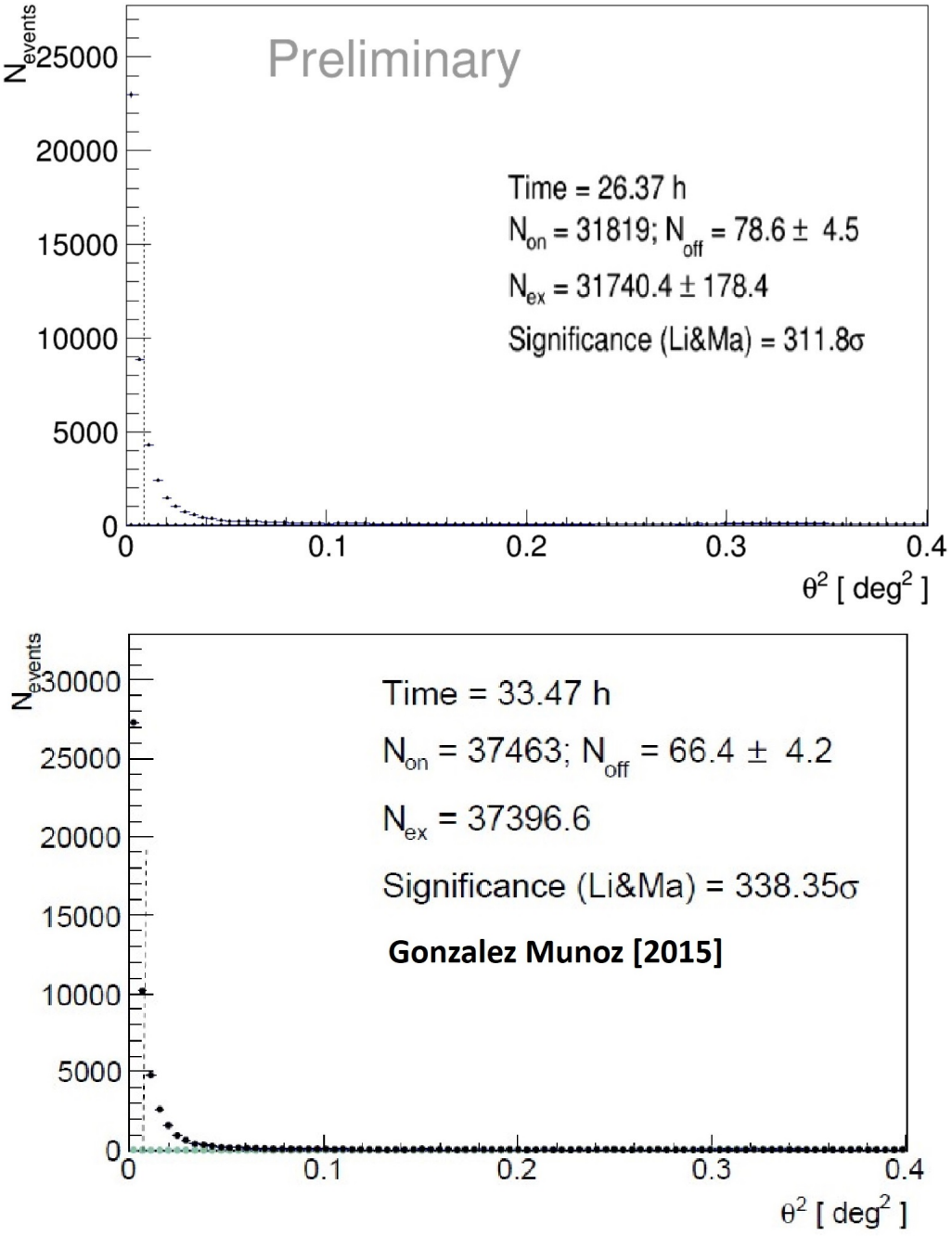


Figure 2.16: Mrk421 April $10^{th} - 16^{th}$, 2013 sample comparison. Top: Full Range θ^2 Standard Analysis (this work). Bottom: González Muñoz (2015) analysis.

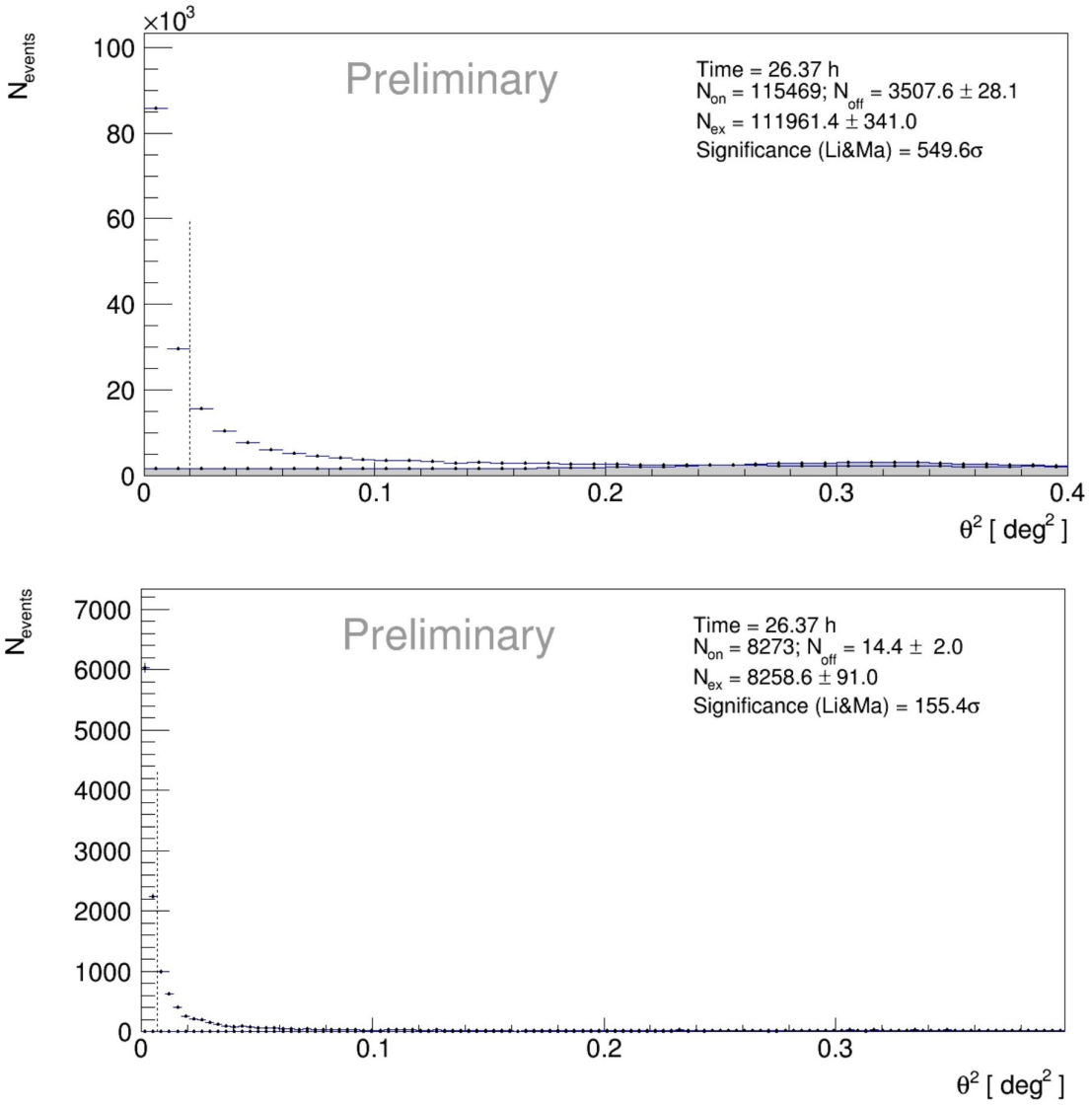


Figure 2.17: Mrk421 April sample (10th to 16th) θ^2 plots. Top my Low Energy cut θ^2 plot (cut in Tab. 2.1). Bottom my High Energy cut θ^2 plot (cut in Tab. 2.1)

2.17 are shown the θ^2 plots after LE an HE cuts (tab. 2.1) and in Fig. 2.18 is shown the comparison of the LCs.

Markarian 421 Cross-Check analysis

Another useful Mrk421 dataset for independent cross-check can be obtained by collecting the data of three days: March 30^{th}, April 1^{st} and April ^{2nd}, 2013. The same MC and dark patch sample used for the Mrk421 main dataset analysis and the Crab cross-check (Sec. 2.4.2).

Fig. 2.19 and 2.20 show the Standard Analysis results for the LE,FR,HE cuts (Table 2.1), that confirm a flux higher than the Crab one.

2.4.4 1ES0927+500

1ES0927+500 is a Bl Lac object confirmed by Fermi-LAT and Swift with redshift = 0.187. Considering synchrotron emission that peaks at exceptionally high energies ($\nu_s \sim 10^{17}$ Hz) in the medium and hard X-ray band, this BL Lac is classified in the Extreme HBL Blazar class (Foffano et al. (2019)). This source was observed a couple of times by MAGIC telescope but every time without a hint of a signal, probably due to its distance.

This source is another interesting sample because it falls in the same period of the Mrk421 extraordinary flare, and because it is an undetected VHE source.

As explained in Sec. 2.2.1, there is a little issue since one set of 1ES0927+500 data (2013-04-02), see Table 2.3) has been used to train the RF for the analysis of the same source.

In Fig. 2.21 and Fig. 2.22 the 1ES0927+500 standard analysis results of the day 2013-04-09 are shown.

2.4.5 BL Lacertae

During my period as flare advocate for the MAGIC collaboration, an exceptional flare of BL Lacertae occurred, the preliminary results for LE cuts showed $\sigma_{Li\&Ma} > 100\,\sigma$ of significance.

This has been an exceptional occasion to join the flare analysis group in view of an article publication. In the following section the results of the cross-check analysis by myself will be presented. This source will also be used as a cross-check sample in chapter 4.

The BL Lac (Acciari et al. (2019)) is a prototype of the BL Lac object, classified as Low Energy Blazar (LBL, sometimes IBL) due to the synchrotron peak frequency and with a redshift $z = 0.069$. The BL Lac is a very variable object in a wide range of energy, usually not detect above 5σ by MAGIC, but it is sometimes

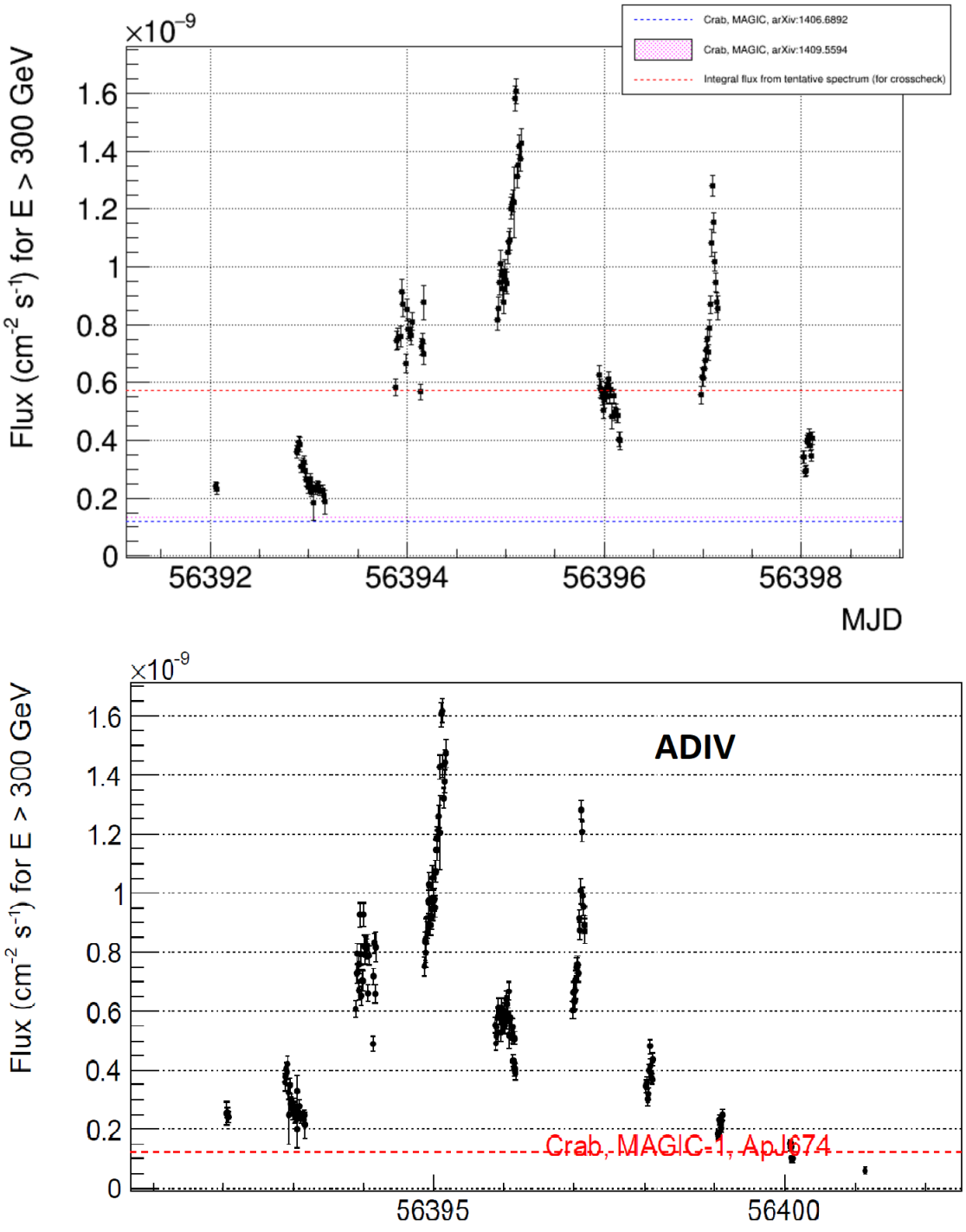


Figure 2.18: Mrk421 April $10^{th} - 16^{th}$, 2013 flare light curve comparison, this work (top) and González Muñoz (2015) (bottom). The curve are very similar, the few differences are maybe due to the conservative choice of discarding the moon days in the data selection.

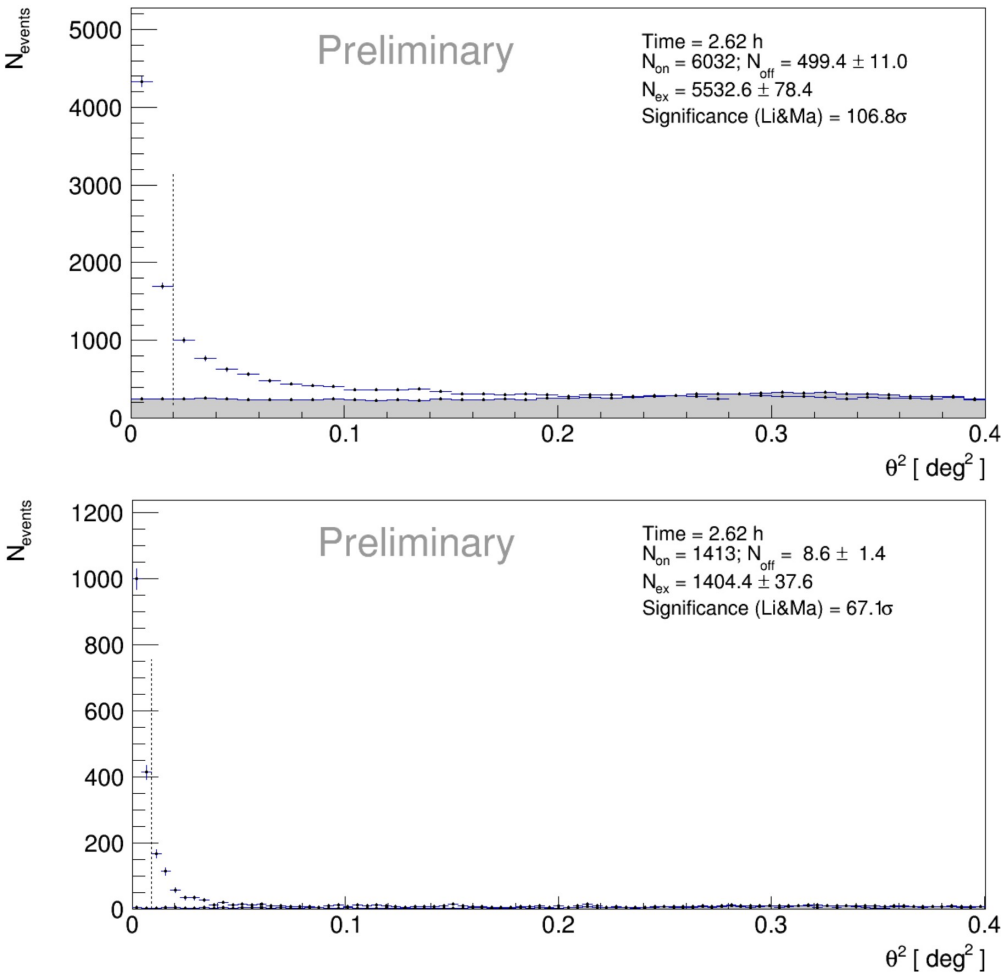


Figure 2.19: Mrk421 cross-check sample θ^2 plots. Low Energy cut (top) and Full Range Energy cut (bottom). The cuts are described in Tab. 2.1

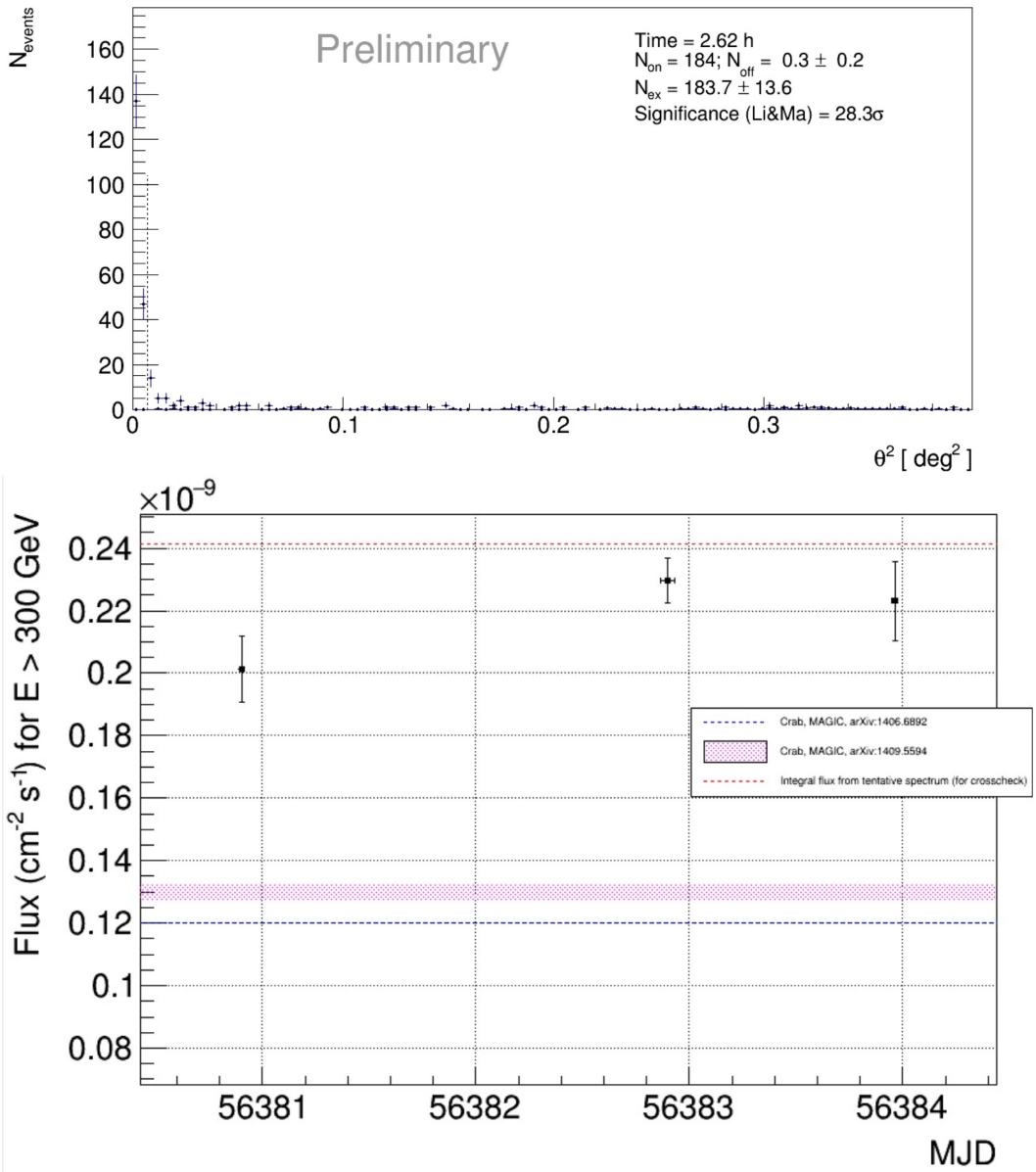


Figure 2.20: Mrk421 cross-check sample sample θ^2 plot and LC. Top:High Energy cut θ^2 plot (Tab. 2.1). Bottom: LC plot.

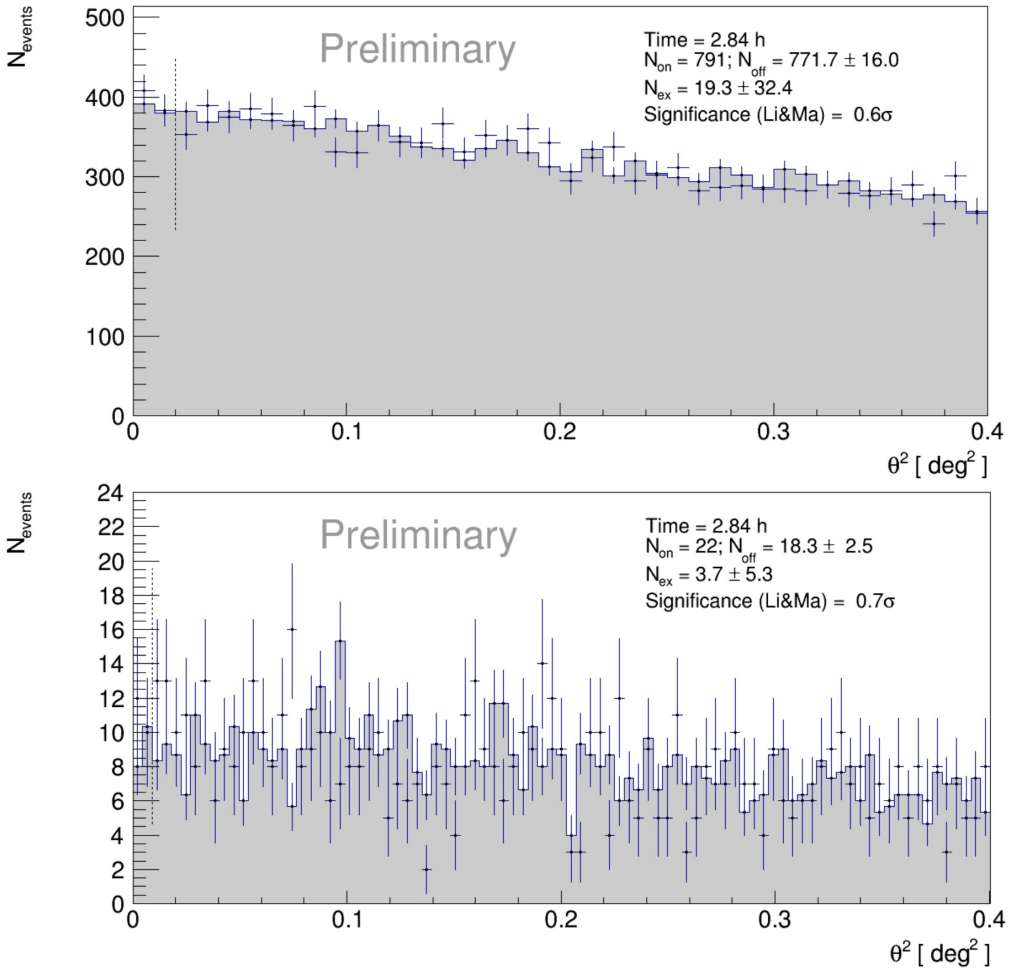


Figure 2.21: 1ES0927+500 2013-04-09 sample θ^2 plots. Top Low Energy cut θ^2 plot. Bottom Full Range cut θ^2 plot. The cuts are described in Tab. 2.1

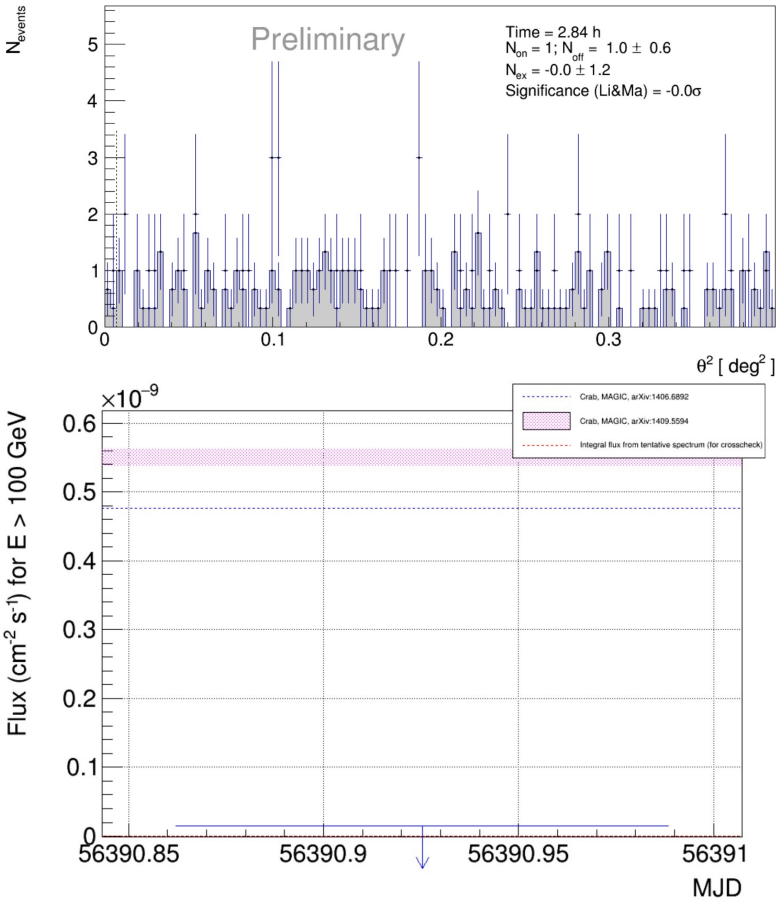


Figure 2.22: 1ES0927+500 2013-04-09 sample θ^2 plot and Light Curve. Top High Energy cut θ^2 plot (Tab. (2.1). Bottom LC plot

observed in flares. One of these flares led to the discovery of BL Lac as gamma-ray emitters in the VHE band by MAGIC (Albert et al. (2007)).

The flare of 20-09-2020 was very peculiar. A specific set of MC data was necessary to allow the analysis because of ashes released by alarm forest fire in August 2020 had altered the mirror performance. The sample of MC and of OFF data used for the RF training are showed respectively in tables 2.4 and 2.5. The OFF sample cover a wide range in Zd because of observations was extended as much as possible in consequence of the exceptional state that was identified almost real time by the observing crew thank to the MOLA (Tescaro et al. (2013)) analysis.

After the selection, three separate zenith angle sets were used to calculate the hadronness probability, the DISP position of the shower and the energy for each event with `coach` and `melibea` (see 2.2.1 and 2.2.1).

Name	dates covered	zenith	type
ST.03.15	2020-09-15 to 2020-10-18	05-35	ringwobble
ST.03.15	2020-09-15 to 2020-10-18	35-50	ringwobble
ST.03.15	2020-09-15 to 2020-10-18	50-62	ringwobble

Table 2.4: Sample of MC used to analyze the BL Lac 20-09-2020 data

Source name	Zenith	Date
87GB225250.5	05-35	20200917
NGC1275	05-35	20200917
IC-200926A	05-35	20200927
TXS0210+515	05-35	20201012
GRB201014	05-35	20201015
GRB201015	05-35	20201016
IC-200926A	35-50	20200927 20200929
Arp187	35-50	20200928 20201016
1ES1741+196	35-50	20201016
GRB201015	35-50	20201016
GRB201017A	35-50	20201018
1ES1959+650	50-62	20200922
B2-1811+31	50-62	20201009 20201010 20201015
1ES0229+200	50-62	20201014
1ES1741+196	50-62	20201016

Table 2.5: The sample used as hadron-like divided by Zenith angle used for the analysis of BL Lac 20-09-2020

Crab cross-check for BL Lac

Analogously to the Mrk421 (see Ch.2.4.3), a Crab Nebula cross-check analysis is needed to confirm that the analysis steps are correctly performed.

For the Crab Nebula check the dataset from the nights `2020-09-18,19,20,27` `2020-10-11,12,13,14` was selected.

The Crab cross-check summary is shown in figure 2.23, including all the three zenith ranges.

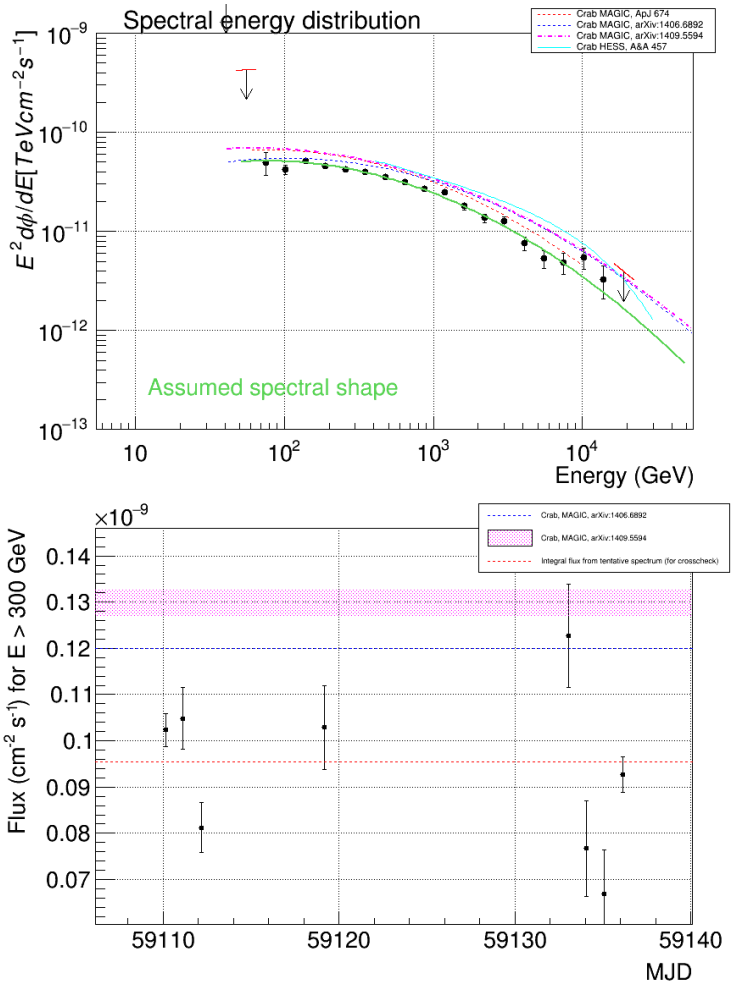


Figure 2.23: CrabNebula SED and Light curve plots. (Top) The SED for the data with zenith angles 0÷62. (Bottom) The light curve divided by nights. The SED is slightly lower than the reference value, this is not related to an analysis problem but is due to the quality of the Crab Nebula data. This is confirmed by the LC where the data have very lower flux than their reference value. Given that the MC time windows are very tight, it is impossible to select better Crab Nebula data

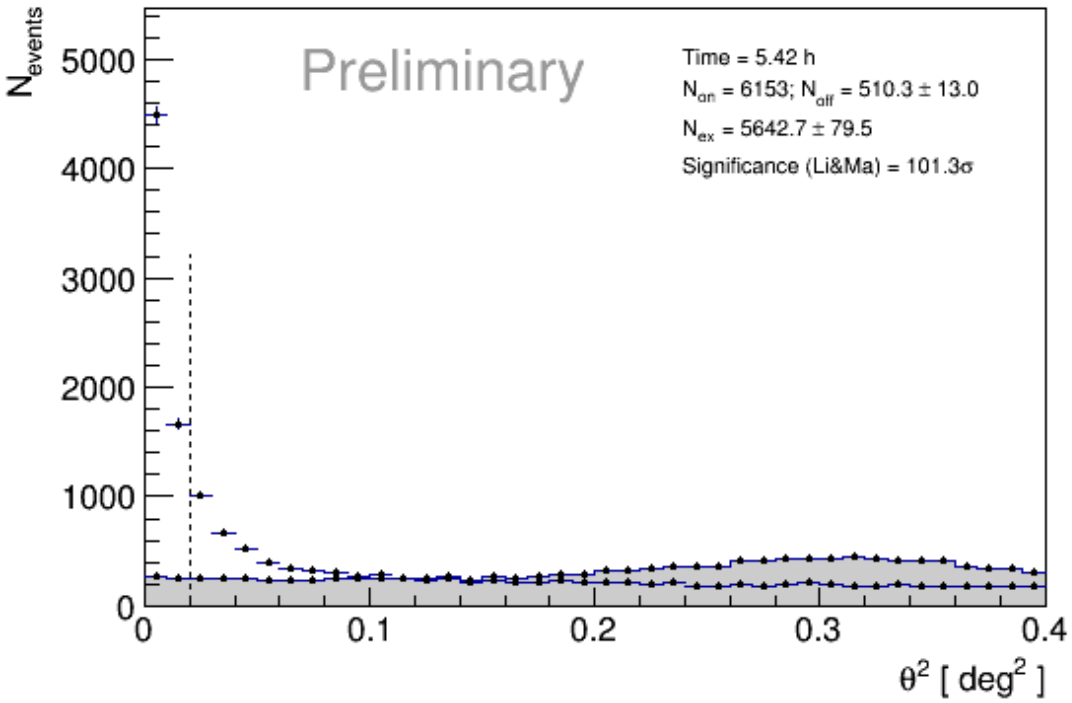


Figure 2.24: BL Lac of day 2020-09-20 preliminary analysis results. θ^2 plot with Low Energy cuts (Tab 2.1)

BL Lacertae cross-check results

The preliminary results for the BL Lac are summarized in Fig. 2.24 and Fig. 2.25 where the θ^2 plot, the SED with a spectral index of 3.6 and the Light curve with a run-wise time binning are shown. From the SED plot, it can be seen that the source has a steep spectrum and it is in agreement with the LBL nature of the source while, from the light curve, the high variability of the source flux can be seen.

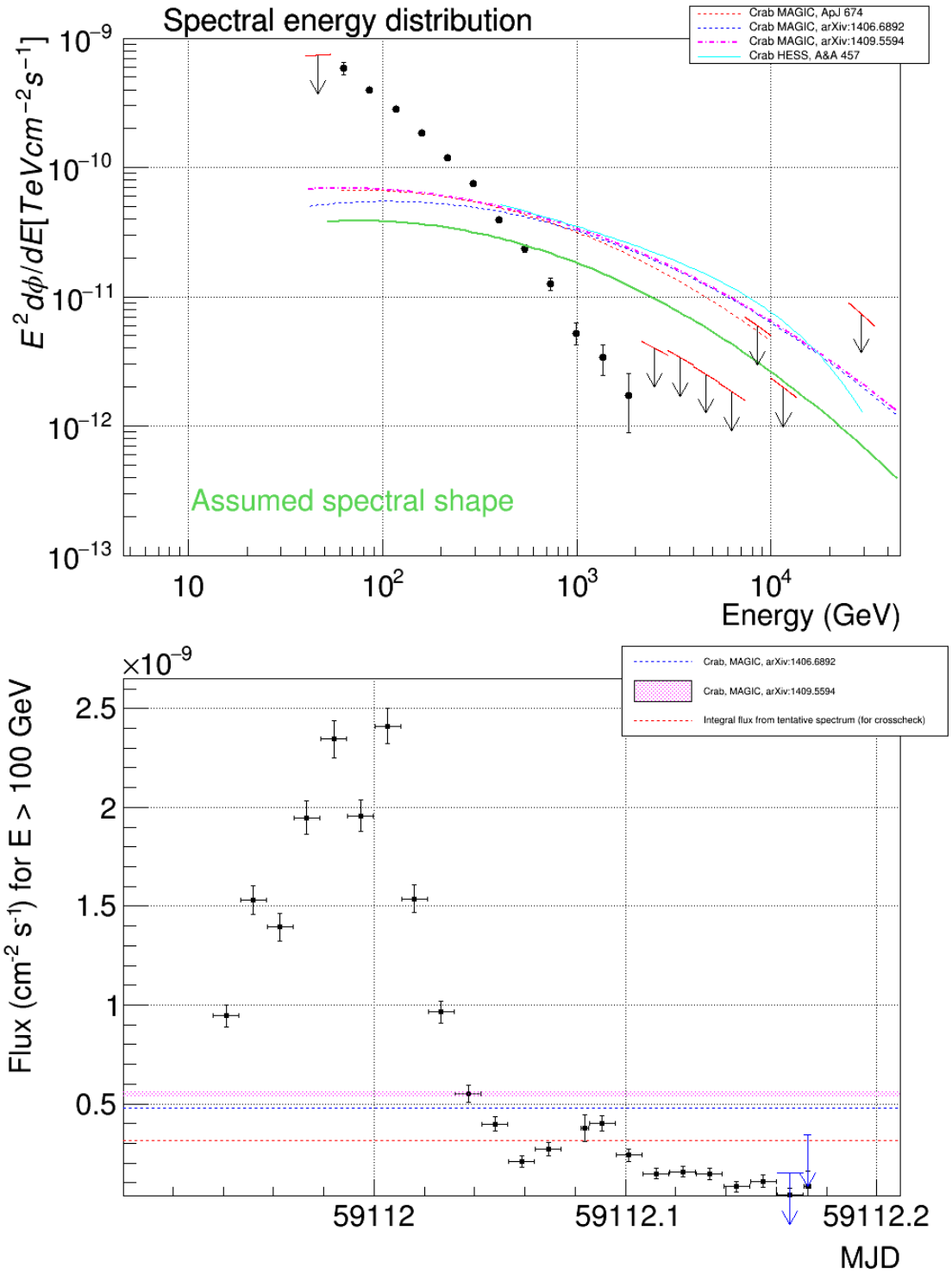


Figure 2.25: Summary of the BL Lac of day 2020-09-20 preliminary analysis results. Top SED plot, bottom LC plot

Chapter 3

Machine Learning techniques

In recent years, Machine Learning (ML) and Deep Learning (DL) algorithms have been increasingly used in data science, from one side due to the massive amount of data that scientists want to process and analyze, and from the other one, thanks to the excellent computing power available at reasonable costs. The purpose of Machine Learning (ML) and Deep Learning (DL) "supervised" algorithms is that of making predictions based on the analysis of a vast amount of well-known cases. These algorithms generally require much-extended calculation time, but the continuous improvement in computer technologies and, particularly in parallel computing, has significantly reduced overall execution times.

In this chapter, I will present a short summary of how the ML/DL algorithms are implemented and used, with a specific focus on the Convolutional Neural Network (CNN).

3.1 Machine Learning

Machine Learning algorithms are software programs that are capable of learning from data and making predictions in new situations based on what they have learned. The learning process is also called training, and the algorithms we are interested in are labeled as supervised ones in the sense that they "learn" from well-known cases. Non-supervised algorithms are not currently in scope; anyway, they are used to partition data in classes not known initially in a totally autonomous way. Such programs are generally defined as "autoencoders".

3.1.1 Artificial Neural Network and Deep Learning

Artificial Neural Network (ANN) are programs that try to imitate the structure of the human brain; they are based on so-called artificial neurons interconnected

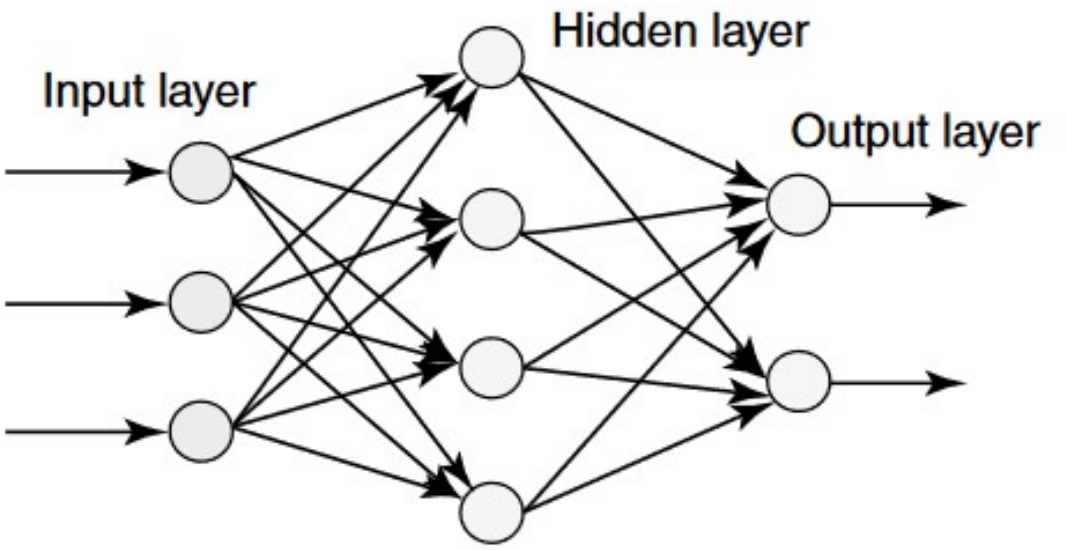


Figure 3.1: Example of multi-layers Artificial Neural Network (Abraham, 2005)

between themselves through `nodes` as natural neurons do and stacked in `layers`. The answer of an artificial neuron to an external stimulus is ruled by a bias and a weight that, together with the signal, are the input parameters of the `activation function`.

The figure 3.1 shows an example of an ANN where the circles are the nodes, the arrows are the weights, and the layers are represented by different shades of greys. Three different layers are represented, from left to right: the input layer, the hidden layer, and the output layer. An ANN with more than three layers is called Deep Neural Network (DNN) precisely because of the presence of the deep layers. The field that studies the DNN is called Deep Learning (DL). Summing up AI $>$ ANN $\geq$ DL.

From now on I will focus on DL that can be considered a specialized subset of ANN and then on CNN that are an even more elaborated version of DL, namely the Convolutional Neural Network (CNN). These supervised feed-forward networks are becoming very popular in the last few years because they perform optimally in analyzing a set of images. The CNN are inspired by biological processes such as those of the brain's visual cortex, whose neuron connection patterns are imitated by CNN. These networks are usually composed of two parts: the convolutional part (`Conv`) that extracts a feature map from the image through convolutional and pooling layers, and the fully connected layers Fully Connected (FC). The convolutional layer executes the convolutional operations on each image's pixels with a kernel matrix, while the pooling one applies filters that reduce the images' size.

DL	non-DL ML
Automatic features extraction	Human-driven features extraction
Greater quantity of data needed	Smaller quantity of data needed
High-end hardware required	Medium-level hardware required
Can manage a problem as a whole	Must subdivide the problem into more minor tasks
Performance optimization is tough	ML programming is near straightforward
Longer train-time required	Shorter train-time required
High accuracy obtained	Low accuracy obtained

Table 3.1: some important differences between DL and ML summarized from Dargan et al., 2020

Lastly, the convolutional block is connected to the Fully Connected (FC) section through a layer that flattens the output of the first block in a single-size array. The FC block is where the network "learns" the features extracted from the last part. The CNN are typically used in image classification, segmentation, and processing. I want to use this network to classify events from our telescopes, using their images, as obtained by MARS pre-processing. In the first stage of our research, I used a handmade CNN (4.8) and, subsequently, I did some experiments with the well-known deep network model InceptionResNetV2 **famous˙network˙mos**. In the next future, I have in plan to change it with the latest EfficientNetB7.

3.1.2 Differences between Machine Learning and Deep Learning

The differences between DL and ML are also the reason why the DL is becoming very famous in the last few years. I listed the main differences in table 3.1 following. Dargan et al., 2020

Automatic data processing and high accuracy results are the main reasons why DL have become increasingly more prevalent. At the same time, the significant advances in the available computing power, including the enormous development

in GPUs, have mitigated the issues about the resource requirements for DL algorithms.

3.2 Linear regression

The linear regression is the well-known technique used to express the linear relationship between a scalar and one or more variables of the problem. The ordinary least squares method is the linear-regression technique used to approximate a solution of a linearly-distributed data problem.

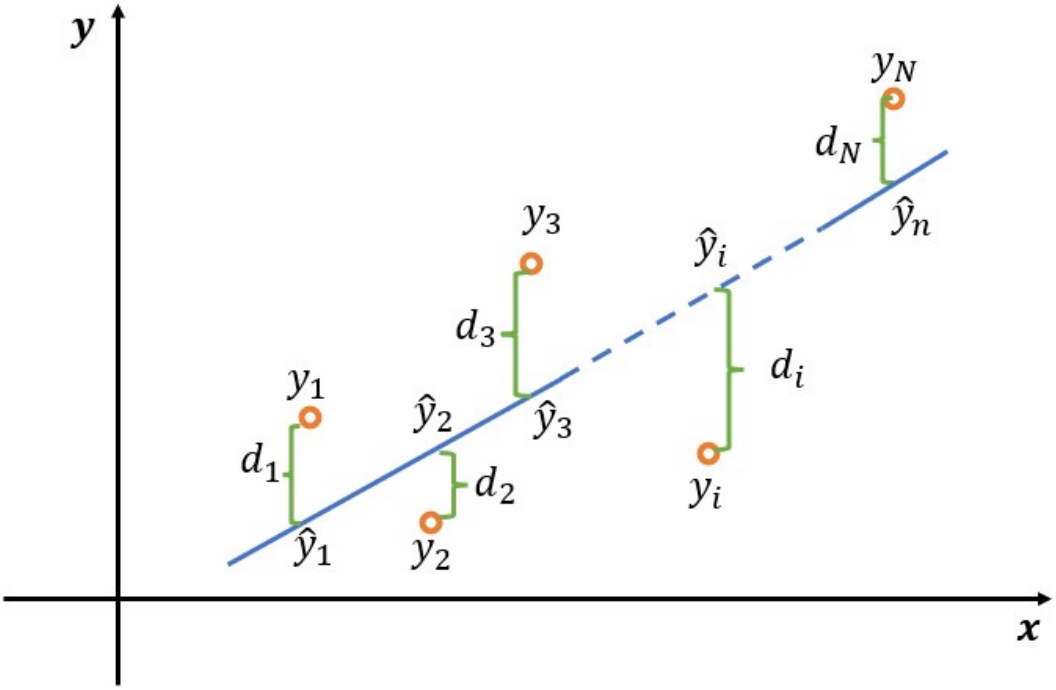


Figure 3.2: Ordinary Least Squares regression plot, where y_i are the empirical data, $\hat{y}_i$ are the estimated data, d_i are the distance/differences between the empirical and estimated data $(\hat{y}_i - y_i)$. The blue line is the estimated linear fit calculated with the distances

Following the Figure: 3.2 I want to calculate the linear regression for a generic linearly-distributed set of data. As well-known, we must find the minimum of the sum of d_i squares, which can be calculated by setting the first derivative equal to zero.

$$\sum_{i=1}^{N}(d_i)^2 = \sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{3.1}$$

I write $\hat{y}_i$ in the form of a linear function:

$$\hat{y}_i = ax_i + b \tag{3.2}$$

This function is called "linear regression function" and, in this case, it is represented by a straight line with a angular coefficient and b for the intercept.

$$\sum_{i=1}^{N}(ax_i + b - y_i)^2 = f(a, b) \tag{3.3}$$

Finally, I obtained:

$$\begin{cases} a = \dfrac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - (\sum x_i)^2} \\ \\ b = \dfrac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{N\sum x_i^2 - (\sum x_i)^2} \end{cases} \tag{3.4}$$

This is the simplest case of the regression algorithm, but I have mentioned it since it represents the starting point of the more elaborated ANN algorithms.

3.2.1 Artificial Neural Network

As I have already pointed out, the Artificial Neural Network (ANN) algorithms have been conceived getting inspiration from the biological brain processes. neurons (nodes) and synapses (connections) are represented by functions as those shown in figure 3.3 and 3.4.

The general purpose of the ANN is to solve complex regression or classification problems. Basing on what is represented in the fig 3.3 and fig 3.4 I are going to describe in practice how a simple ANN algorithm works.

I start from linear regression to solve a problem described by points:

$$x_i, y_i \in \mathbb{R}, i = 1, ..., N$$

where x_i are the N independent input variables and y_i are the N dependent output variables. I want to find the "best" function for estimating the y_i values:

$$\hat{y}_i(x_i) = ax_i + b$$

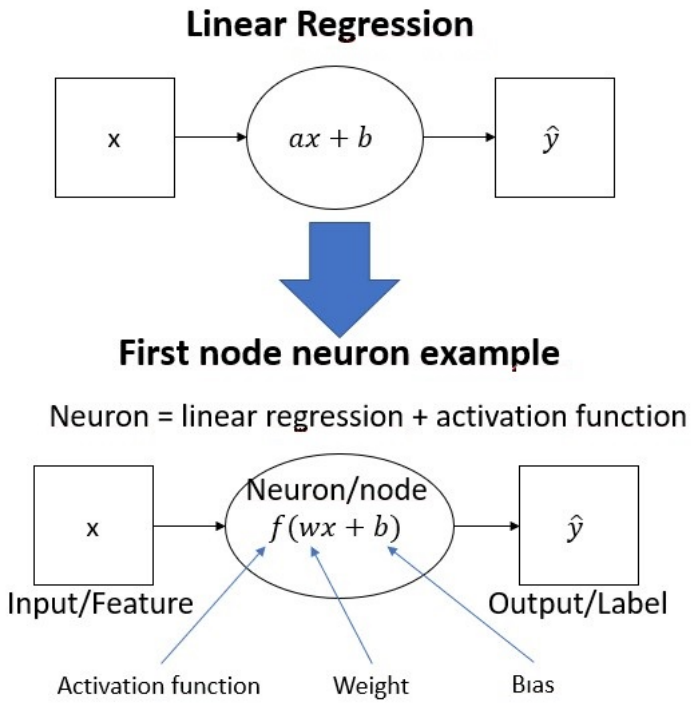


Figure 3.3: example of simple neuron/node and synapses/connection.

to do this one should use a function like that in eq: 3.1:

$$\sum_{i=0}^{N}(\hat{y}(x_i) - y_i)^2$$

these types of functions in ANN are called "Loss (or Cost) function"[1]. They represent the distance between the empirical/target output value and the estimated/calculated output value for y. The goal is to minimize the loss function so the actual values of $\hat{y}(x_i)$ will be very close to that of the points y_i. I listed some type of loss function with theirs common use in tab: 3.2.

The other pillar of ANN algorithms is the "activation function" (f in 3.3). This function acts as a switch that turns ON or OFF the node depending on its input. I will discuss in detail these type of functions in their dedicated section.

The input quantities are called "features" and consist of the characteristic properties of the data[2]. The output quantities are called "labels": making a comparison with the linear regression function eq 3.2, the angular coefficients become the so-called "weights", and the intercept terms are "biases". Nodes can be combined in complex structures and, consequentially, inputs data, weights, and biases will be represented by tensors.

In general, I can say that the algorithm searches for the best set of weights that minimize the loss function. The biases are not relevant to the performance of the network, but they form a set of constant values used for increasing the network generalization. When some inputs are sent to the network, the engaged nodes react depending on their weights and the relative activation function so that the last node array will present an output.

The output array must be compared with the array of expectation values, and the Loss function is computed. The process of training consists of iterative searching for the best set of weights that minimize the Loss function. Each iteration is called `epoch`; for each epoch, a new set of weights is generated through a very peculiar algorithm that starts from the output and goes backward, neuron by neuron, adjusting the weights to obtain the "right answer". Such an algorithm is better known as the "back-propagation", and it is the core of the training process.

In figure: 3.4 I show how a simple network schema can appear. The behavior and the performance of the network strongly depends on a set of parameters that are traditionaly called "hyperparameters": tuning a ANN is an empirical procedure by which the operator vary the hyperparameters searching for the best answer of the network.

[1] In literature sometimes Loss and Cost functions are different: I will be using the two terms as synonyms

[2] the name features derives from Machine Learning: here the features are human-chosen; in ANN instead, they are extracted automatically by the AI network

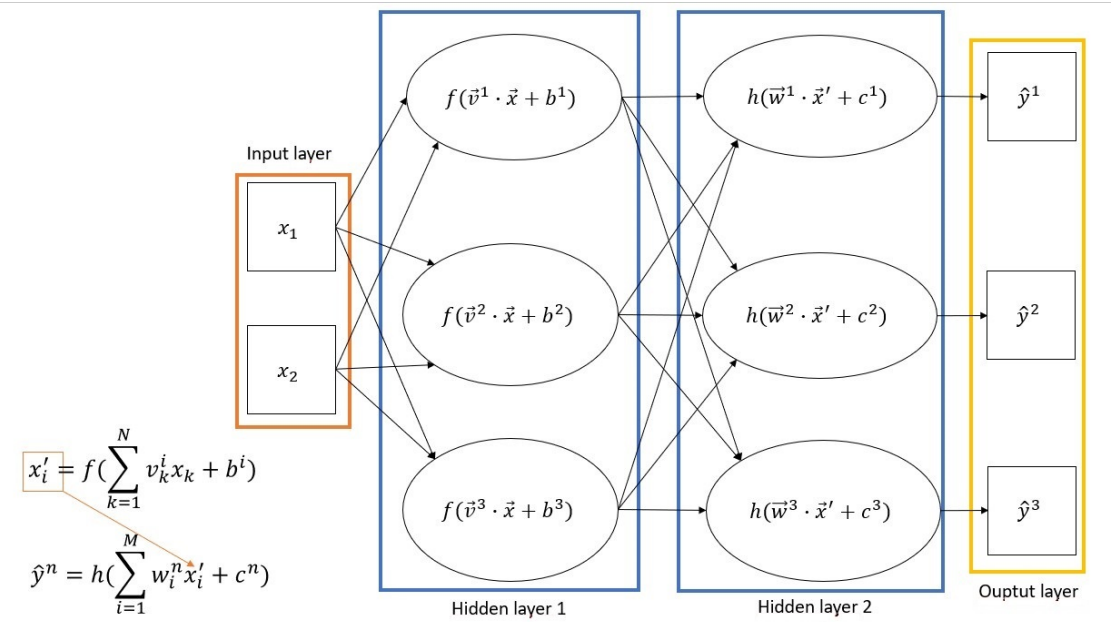


Figure 3.4: An example of DNN with a 2-features input layer, 2 hidden layers with 3 nodes, and 1 output layer with 3 labels

LOSS function

There are many types of loss functions, and each is used in different kinds of problems; I listed some of the most used loss functions with their typical use in the table: 3.2.

Activation functions

The activation functions are used to activate/deactivate the nodes and to introduce non-linearity. The most straightforward function is the binary step function (fig: 3.5 left), but it is not the most used since its response is linear; the sigmoid function is preferred.

Another activation function largely used by data scientists (fig: 3.5 right) is the Rectified Linear Unit (ReLU), whose output, in the active part, is directly proportional to the input.

There are plenty of activation functions, each with different benefits and drawbacks for every situation and necessity. A good practice is to choose the function type depending on the task and the type of network, and for implementation reasons, it is generally the same for all the nodes of the same layer. For the output layer, the function is chosen according to the typology of the problem (e.g., regression, binary or multiclass classification).

loss function	$L(y_i, \hat{y}_i)$ equation	typical use
Mean squared error	$\frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2$	Regression
Mean squared logarithmic error	$\frac{1}{N}(\log(y_i + 1) - \log(\hat{y}_i + 1))^2$	Regression
Mean absolute error	$\frac{1}{N}\sum_{i=0}^{N}\lvert y_i - \hat{y}_i\rvert$	Regression
Binary cross-entropy	$-\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)$	Binary classification
Squared Hinge	$\sum_{i=0}^{N}(max(0, 1 - y_i\hat{y}_i)^2)$	Binary classification
MultiClass Cross Entropy	$-\sum_{i=1}^{N} y_i \cdot \log(\hat{y}_i)$	Multiclass classification

Table 3.2: Most common Loss functions and their general uses

As I have already said, for all the nodes in the hidden layers, the same activation function should be used, and the most adopted is the ReLU.

I have listed the most used activation functions and their common use in table 3.3.

3.3 Convolutional Neural Networks

I want now to drill down on the structure of a Convolutional Neural Network (CNN) that is the network topology I have selected to classify the event images coming from our telescopes. The discussion will include sections that are common to other types of ANN and parts that are unique to CNN. As I have already said, the CNNs are supervised, feed-forward networks that perform very efficiently in the image analysis. In general, the CNN scheme can be decomposed in two sub-blocks (O'Shea and Nash, 2015): the convolutional layers that extract a features map from the images, and the Fully Connected (FC) part, where the network learns the features extracted in the `Conv` part so that it will be able to recognize unknown images. In Fig. 3.6 is shown a basic model of CNN.

The `Conv` part comprises layers that perform a filtering operation on the images. The two main filtering layers are Convolutional and Pooling.

Activation function	Equation	Derivative	Common use
Linear	x	1	Regression Output
Binary step	$\begin{cases} 0 \; for \; x < 0 \\ 1 \; for \; x \geq 0 \end{cases}$	$\begin{cases} 0 \; for \; x \neq 0 \\ undef. \; for \; x = 0 \end{cases}$	Binary classification
Softmax	$s_i = \frac{e^{z_i}}{\sum_{l=1}^{K} e^{z_l}} \; for \; \vec{z}$	$\frac{\partial z_i}{\partial z_j} = \begin{cases} 1 \; for \; i = j \\ 0 \; other \end{cases}$	Multiclass classification output
Sigmoid	$\sigma(x) = \frac{1}{1 + e^{-x}}$	$\sigma(x)(1 - \sigma(x))$	Binary/Multilabel output RNN hidden
Hyperbolic Tangent	$tanh(x)$	$1 - tanh^2(x)$	Recurrent NN Hidden
ReLU	$y = \begin{cases} 0 \; for \; x \leq 0 \\ x \; for \; x > 0 \end{cases}$	$\begin{cases} 0 \; for \; x < 0 \\ 1 \; for \; x > 0 \\ undef. \; for \; x = 0 \end{cases}$	CNN Multilayer Perceptron Hidden

Table 3.3: List of the main Activation functions and their common use

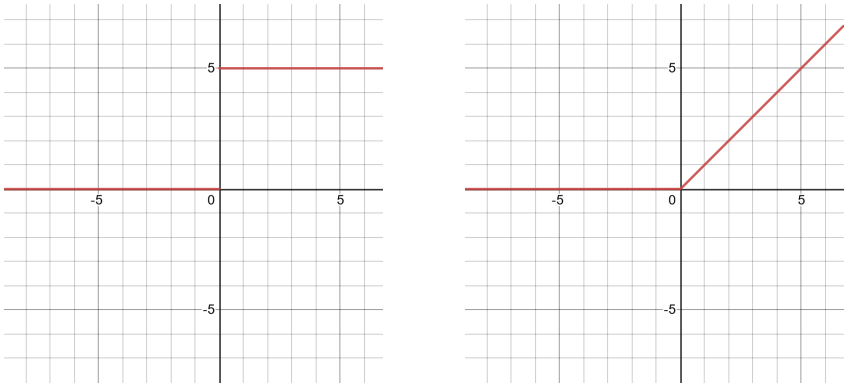


Figure 3.5: Left: Binary step function. Right: ReLU function

Convolutional

Convolutional layers execute a convolution operation between the data matrix and an auxiliary matrix called the kernel. The kernel matrix slides over the data matrix until it is entirely covered (Fig. 3.7). The kernel matrix element values are randomly chosen between 0 and 1. The size and the number of convolutional filters are set up in the network model. The set of matrixes obtained by applying the convolutional filters is called a "feature map".

Pooling

Pooling layers apply a pooling filter on the data matrix. As in the convolutional operation, the pooling is performed with a sliding window technique. This operation is critical since it reduces the size of the input tensor, depending on the pooling settings. The main types of pooling are:

- Max pooling: takes the maximum value from the sliding matrix and puts it in the output matrix.
- Average pooling: takes the average value from the sliding matrix and puts it in the output matrix.

A typical a CNN model is composed of many Convolutional layers alternating with pooling layers (3.6).
At the end of the Conv part, a flattening layer transforms the processed image tensors into a 1-size vector. This vector of features is sent as input for the FC part. A FC part presents a DNN (see Fig.3.4) structure. The name of this block indicates that all the nodes of a layer are fully connected to all the nodes of the following one.

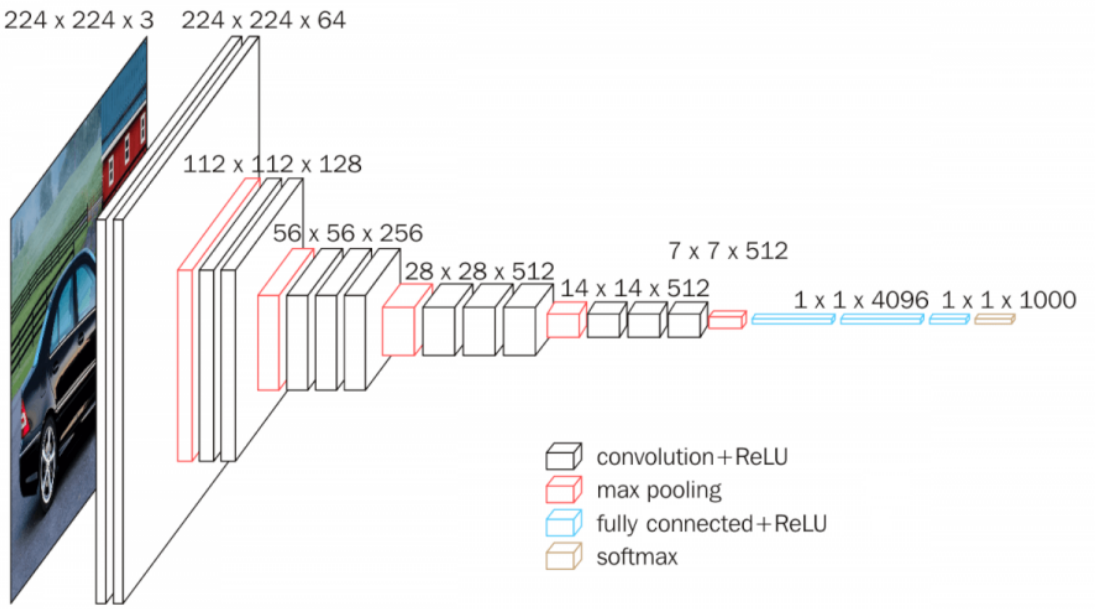


Figure 3.6: graphical representation of the CNN VGG16 scheme (Simonyan and Zisserman, 2014)

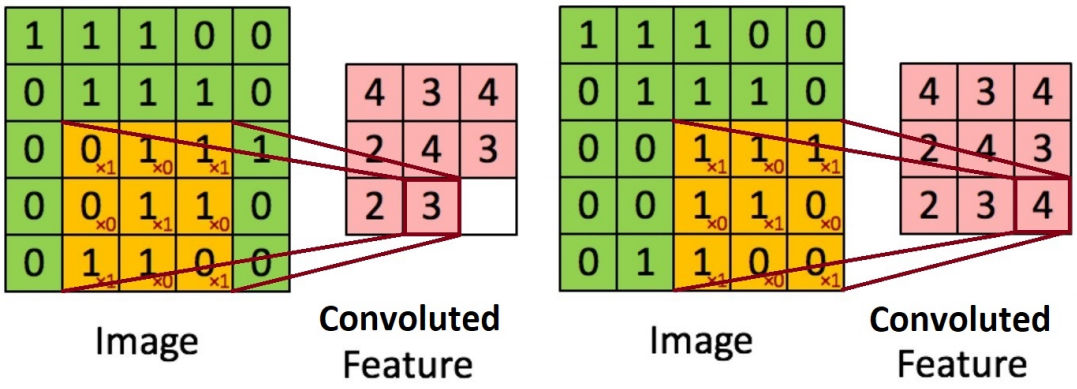


Figure 3.7: Example of the last two convolutional multiplication of data matrix (in green color) with a 3x3 kernel (in the yellow windows). In pink, we can see the resulting matrix where each number corresponds to a single convolutional operation.

At the end of the FC, there is a classification layer (in Fig. 3.6 the softmax layer) that assigns the input image to a specific class with its relative probability.

3.4 Dataset Pre-processing, Training, Test and Validation

The preparation of the dataset is the starting point in the process of building an ANN project. I here remind that a dataset in ML is just a collection of data represented in a tabular form of rows and columns, where the columns are the properties of the instances and the rows represent a unique data instance.

Data scientists usually perform some standard operations on datasets, such as anonymizing the data or making them consistently uniform. In the case that the sample is not numerous as requested by the network topology, a data-augmentation procedure can be applied to the dataset. Data augmentation techniques must be carefully chosen, but they effectively allow for extending the data sample widely.

When the dataset is ready to be processed, it is randomly divided into two subsets that will feed the training and validation phases. It is a good rule to have a separate dataset to be used for the final test, and it is usually done by extracting a portion of the original data sample before the splitting. The most common percentages of splitting vary from 80% to 50% for the training subset and, consequently, from 20% to 50% for the validation part. The test dataset will be utterly unknown to the network, and it will be used to measure the actual capability of the network to make correct predictions from unknown data. When I say "correct" predictions, I obviously mean accordingly to the accuracy of the network.

3.4.1 Training

The training is an iterative process during which the network learns from data. After each epoch training cycle, the Loss Function is calculated by applying the network with its current weights to the validation dataset. The value of the Loss Function is the metric to evaluate the "goodness" of the model. To limit the use of computational memory, within a single epoch, the training dataset can be sliced into smaller chunks called "batches". For each epoch, the number of batches is the same, but from one epoch to another, the batches can be different since the dataset can be shuffled. The final goal of the training cycle is to find the weight set that minimizes the value of the Loss Function. During each cycle, the weights are adjusted by the backpropagation algorithm. It is good practice to obtain plots of the Loss Function and the Accuracy vs. the training epoch: checking the training curves is very useful to understand if the network is working correctly.

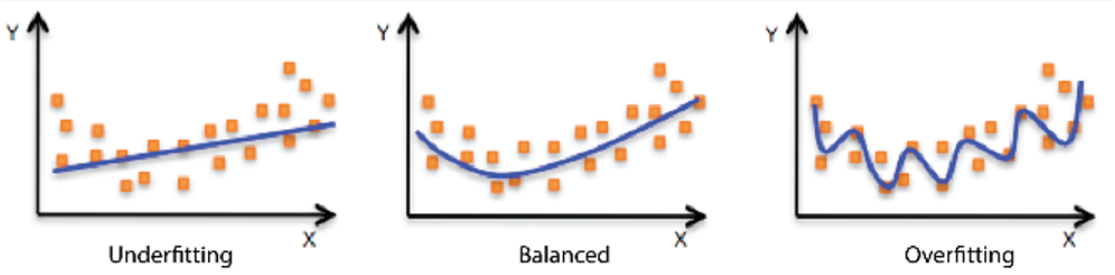


Figure 3.8: Example of Underfitting (left), balanced fit (center) and overfitting (right) from Amazon Machine Learning developer guide.

In particular, we can detect abnormal network behaviors such as the so-called "underfitting" or "overfitting".

3.4.2 Overfitting and Underfitting

Overfitting and underfitting phenomena are two of the most common issues in ML.

- **Overfitting** is when the estimated data reproduce too "perfectly" the training data sample, with a loss of generality: the network has not been able to extract and learn the right characterizing features but has learned from specific details. So, the performances on unknown test data are typically abysmal.
- **Underfitting** is the opposite of overfitting: the training process can be improved, and the loss function can be minimized further. Even in this case, the network will perform poorly on unknown data. I have shown an example of that in fig: 3.8.

3.4.3 Validation

Validation is the way to evaluate the current weight set with the goal of finding the optimal configuration for the network: the configuration is the combination of the network model with a proper set of hyper-parameters and the weight collection. The validation is the calculation of the Loss Function through the application of the current network configuration in making a prediction on the validation subset. It is performed for each epoch of the training. In DL most adopted libraries (such as Tensorflow or Pytorch), special functions are available that record the best results obtained or break the training cycle as soon as the Loss Function goes below a predetermined threshold or the Accuracy goes beyond an expected value.

	Origin positive	Origin negative
predicted positive	TP	FP
predicted negative	FN	TN

Table 3.4: Example of Confusion Matrix for binary classification

3.4.4 Test

When an optimal configuration has been found during the training phase, the network must be tested on unknown data. As I said before, the test dataset can be extracted from the original datasets. Sometimes public datasets can be used to benchmark the trained network's performance. As an example, if I would design a network to classify the numbers from 0 to 9, I could download images from MNIST famous dataset.

3.4.5 Back Propagation

Back Propagation (BP) is a core technique of an ANN since it is used to adjust all the weights to obtain the correct response of the network. It is called Back Propagation (BP) just because weights, starting from the final output, are consequently varied from the final layers back to the input ones.

As already said, the final goal is to obtain the best weight set possible and minimize the Loss Function with respect to those weights themselves for every single input-output sample.

3.4.6 Metrics

As I already pointed out, the general goal of training is to reduce the Loss Function on each epoch, trying to minimize it overall. Other functions have been introduced to help data scientists evaluate the results and give a network's performance more significance. I will now introduce a handy tool, the so-called "confusion matrix", starting from a binary case. In a supervised binary classification problem, network performance is tested with an input dataset for which classification labels are well-known. The network will produce a set of predicted labels that can be compared with the so-called ground-true: so it is possible to create the confusion matrix (Tab.3.4).

The acronyms of the confusion matrix are explained in the following:

- **True Positive(TP)** are the elements belonging to class "positive" that the model predict as belonging to class "positive".

Metrics name	Metrics definition
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision / Positive Predicted Value (PPV)	$\frac{TP}{TP+FP}$
Recall / Sensitivity / True Positive Rate (TPR)	$\frac{TP}{TP+FN}$
Specifity or True Negative Rate (TNR)	$\frac{TN}{TN+FP}$

Table 3.5: Most used Metrics definitions (Davis and Goadrich, 2006 and Zhu, Zeng, Wang, et al., 2010)

- **False Positive (FP)** are the elements belonging to class "negative" that the model predict as belonging to class "positive".
- **False Negative (FN)** are the elements belonging to class "positive" that the model predict as belonging to class "negative".
- **True Negative(TN)** are the elements belonging to class "negative" that the model predict as belonging to class "negative".

From these four elements, it is possible to define some metrics. Some of the most common metrics are listed in Tab.3.5.

In addition to Loss Function, current DL libraries allow the operator to specify one o more of the metrics above, making the final results much more understandable to researchers.

3.4.7 Regularization

I want to conclude our panoramic view of CNN with some hints on regulararization techniques. The regularizations are techniques used to improve the performance of an ann. They are typically used to stabilize the training process or to lower overfitting. Among the most commonly used regularization techniques, we can find:

- **L1 and L2 regularization**: These techniques consist in adding a terms to the Loss function (described in sec: 3.2.1)for example:

$$Loss_{L1} = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} |w_i| \tag{3.5}$$

$$Loss_{L2} = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} |w_i|^2 \tag{3.6}$$

With the addition of the term (λ) the Loss derivative changes and intuitively is shifted to prevent the "perfect" fitting of the real problem and avoid overfitting. A good explanation of this technique can be found in Ng, 2004.

- **Dropout** The dropout definition can be found in Srivastava et al., 2014. This technique consists in randomly dropping some nodes and their related connections from the neural network during the training. This operation is re-iterated for each epoch. This technique is used most in the FC block.

- **Batch normalization layers** They are described in Ioffe and Szegedy, 2015 and consider that the distribution of each layer's input changes during the training process since the previous layers change too. For this reason, careful parameter initialization and lower learning rates are required, slowing down the training time. In the article, the author defines a value called "covariant shift" used for normalizing the layer inputs. The `Batch` normalization permits to increase in the learning rates and allows being less careful in the parameter initialization. In general, this technique is used in the `Conv` part of the CNN.

3.4.8 Fine tuning and Transfer Learning

These are two techniques used to re-train each model partially with a new set of training data in order to reduce the time calculation and to reduce the quantity of data requested for this operation. The two techniques are described in detail in Tajbakhsh et al., 2016 are:

- **Transfer Learning:** consists in locking the weights of all the layers of the neural network obtained with the dataset and adding a top classification layer that is trained to classify a similar dataset.

- **Fine Tuning:** Fine Tuning is a technique by which a network is first trained massively with public datasets or datasets different from the problem of interest. In this case, weights and layers are not locked, so it is possible to complete the training with the dataset in scope, on top of what has been previously done.

3.4.9 Data Augmentation

As I have already anticipated, this technique is used to improve data generalization by increasing the quantity of the data. It consists in applying some transformations to the original sample images to obtain much more samples. Among the most common operations for data augmentation, we find:

- Rotation and flipping
- Translation
- Re-scaling
- Brightness modification
- Others

These operations can be applied in the pre-processing stage, before or during the training stage. The advantage of the first technique is that the new images are saved on the storage, and they are added permanently to the data sample. The second technique permits saving storage space, but the generated samples are not persistent. Nevertheless, a new set is generated on each epoch so that the obtained generalization is more effective than in the first case. For a more in-depth explanation of these techniques, see:Shorten and Khoshgoftaar, 2019.

Chapter 4

Convolutional Neural Networks on MAGIC data

The information that an IACT system, like MAGIC, exploits for each event includes the charge and the arrival time distributions acquired by the cameras. In the standard MAGIC analysis, the shower information is described in a set geometrical parameters known as "Hillas parameters" (Hillas (1985), see Fig. 1.22). This set of variables was popular at the beginning of the IACT era when the hardware resources were limited. Current computer technology, and especially the parallel computing capability of GPUs, allows the processing of a large quantity of data. The original idea of this work is to exploit modern techniques, such as CNN algorithms, to improve the performance of analysis with respect to the MAGIC standard analysis.

This chapter focuses on the development of an original CNN algorithm that is capable of performing an efficient gamma/hadron separation based on "real" data sample only, in particular, the dataset for the Mrk421 flare observed by MAGIC in April 2013 will be used.

Initially, I trained the CNN using the charge distributions only, in a second phase, I added the time information, too. Finally, I will describe the steps to create a general model, and I will make a comparison of the results obtained by this model with some control data samples.

4.1 Training with Monte Carlo Data

CNNs (Sec. 3.3) are specifically designed for image analysis, they belong to the class of supervised neural networks (Sec. 3.1) and need as input a set of well-known data. Simulated (MC) data (Sec. 2.2.1) have well-known and defined properties. With simulations software like CORSIKA (Sec. 1.3.2), sets of gamma-like and

hadron-like events can be easily produced and used for training a supervised neural network. The generation of a hadron-like sample is more demanding in terms of computational resources than for a set of gamma-like events of comparable statistics; for this reason, obtaining a well-diversified hadron-like ensemble of events is not a light task.

In the majority of the literature regarding IACT and CNN, it is common to use MC data to train a CNN and then apply the obtained model to a MC and/or to a real test dataset. A common CNN scheme used for gamma/hadron separation can be found in Shilon et al. (2019). In Fig. 4.1 one can recognize the input layer whose size is the number of image pixels (64) $\times$ the number of images (64) for a single event $\times$ number of telescopes (4), this CNN is used to predict the arrival direction of the shower.

Furthermore, as described in Sec. 3.3, the CNN topology can be divided in two sub-schemes: the `Conv` layers and the `FC` ones. The first set extracts the features from the images, and the second set classifies them. Moreover, in the network are present few `Dropout` layers that are used to regularize the CNN (Sec. 3.4.7). Another example of CNN trained with simulated MC dataset is proposed in Juryšek, Lyard, and Walter (2021).

In general, it is known that these methods give good results when they are applied to a test MC dataset, but they can still fail when tested on an observed dataset. This is due to the discrepancies between the simulated and the observed events. For instance, the observed events can be affected by the level of NSB, weather conditions, and the hardware status, and not all these effects can be taken into account in a simulation.

To overcome these problems, some solutions have been proposed. For example, Parsons and Ohm (2020) connects an extra ANN scheme that includes Hillas and stereo parameters on the CNN. This sub-scheme is merged with the CNN scheme using a flatten layer before the fully connected layers scheme. Another approach is to simulate the effect of NSB by adding Poissonian noise in the images (Jacquemont et al. (2021)). These solutions are very different: the former acts directly on the network topology while the latter is performed later, in the image preparation phase.

In synthesis, the research for a solution to overcome the CNN classification issue caused by simulated/observation differences leads to adding some features in the network or the images. Training a CNN using observation data is a genuinely unexplored path.

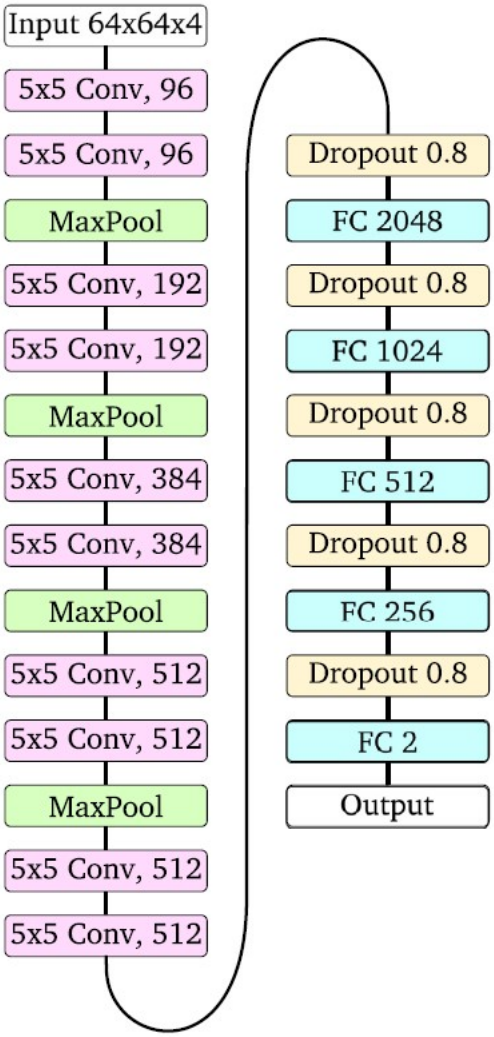


Figure 4.1: CNN architecture used for regression (Shilon et al. (2019)). This network is used to predict the shower arrival direction.

4.2 Application to MAGIC Data

The purpose of the new approach is to avoid the systematics potentially introduced by discrepancies between simulated and telescope data in the standard CNN classification algorithms.

In order to achieve this goal, we will make use of datasets from telescope observation to train the CNN. This solution has not been fully explored yet, probably because the supervised CNN requires a clearly classified dataset and using a "real" dataset: it is challenging to find a well-formed gamma-like sample while, in the MC data, each event type can be undoubtedly identified as a gamma-like or hadron-like

event since the particle type is specified in the MC configuration. On the other hand, the primary advantage of using real data is the reduction of the differences between results from the training and the test datasets.

A similar task was treated in Escolano (2019), and the results obtained with the observation data were similar to the results obtained with MC RF. Our working hypothesis is that, since both CNN and RF are supervised types of ANN, and provided that the CNN can exploit more information than the RF, using a CNN could lead to an improvement in the gamma-like/hadron-like separation.

4.3 The Markarian 421 Flare Dataset

Flares are phenomena consisting of a sudden increase in the luminosity flux of astrophysical sources, usually associated with a physical event that happens in the source itself.

Given that the flare produces a very large quantity of gamma rays that overwhelms the hadron-like background along the source direction, these events can be very useful to properly train a CNN . In addition to the flare data, it is necessary that during its observation, good atmospheric conditions have occurred, and also, no bright star is in the camera FOV to prevent the CNN from learning any unwanted details.

A promising source candidate is the Mrk421 (Sec. 2.4.3) and its extraordinary April $10^{th} - 16^{th}$, 2013 flare. The Mrk421 flare has been extensively studied, for example in Acciari et al. (2020) and González Muñoz (2015). We will use the latter as a reference for the Standard Analysis. The shower-arrival direction and the "hadronness" obtained from the Standard Analysis (Sec. 2.4.3) can be a good training set.

4.3.1 Dataset creation

To prepare the dataset for training, we first have to apply the cleaning, which will avoid poor learning by the neural network.

Images Transformation

The dataset from which I started is shown in Table 4.1 where I describe all the dataset properties. This dataset was obtained by merging two Standard Analysis stages (Sec. 2.4), calibration (Low-level stage), and Melibea (High-level stage). I did not discard data in this process, with a few exceptions when going from calibrated (sec:2.1.2) to star level (2.1.3 data). During this phase, the raw charge

Description	**MAGIC-1**	**MAGIC-2**
Charges	$[q_1, q_2, q_3, \dots]$	$[q_1, q_2, q_3, \dots]$
Arrival Times	$[t_1, t_2, t_3, \dots]$	$[t_1, t_2, t_3, \dots]$
Total charge	$size1$	$size2$

Description	**Variable(s)**	
Event Hadronness	$hadronness$	
Source Position	x_pos	y_pos
Shower DISP	$disp\_stereo\_x$	$disp\_stereo\_y$
LUT Energy	$energy\_stereo$	
Run Identifier	run_number	
Event Stereo Identifier	$stereo\_id$	

Table 4.1: Example of a simplified dataset used for CNN γ/h separations.

and arrival time information is discarded to reduce the data volume and to lower the processing time [1].

We keep the charges and the arrival time maps from calibrated/star (low-level data), and the hadronness and the arrival direction from Melibea (high-level data). In this way, we could keep the raw charge and time maps associated with the high-level parameters.

The merging and storing operation can be summarized with the following steps:

- from calibrated files, the event type is checked for both the telescopes, and if it is not a data event (e.g. it could be a so-called pedestal event, used to evaluate the cleaning efficiency) is discarded
- all the remaining calibrated data of one run, from both telescopes, are loaded in memory; this operation is due to the calibrated stored logic (2.1.2)
- the calibrated events of one run for both telescopes are then merged
- from Melibea files, the event type is checked for both the telescopes, and if it is not a stereo event, it is discarded
- the selected Calibrated and Melibea events are finally merged using a common-event index (*stereo_id* in 4.1).

[1] Star has an option to maintain the charges and arrival time maps so that one can use the data after this option or use directly the calibrated, both ways are correct

- before dumping on disk the final dataset, all the events with $size1$ or $size2$ less than 50 are discarded like in standard analysis.

The shown method is definitely the most memory-expensive, and it could be non-workable depending on the RAM quantity available on the hardware used for the analysis. A good alternative could be to load in memory only events with the same *stereo_id* from the M1 and the M2 subruns. In this way, we could have only two subruns, one for each telescope, loaded in memory simultaneously.

For our analysis, we used a high-performance computer with a large RAM capacity[2] so we did not have to manage memory issues and we could load all the dataset events at once.

Charge and Arrival Time Distributions

At this stage, we can start working with the real data, more precisely with the two kinds of "images" provided by each telescope, namely the charges and arrival times.

Fig. 4.2 shows an example of these images. The top pictures show the pixels charge in both telescopes, while the bottom images show the corresponding arrival times map. In the same area where the charge maps show a shower image, a localized group of pixels with consistent arrival times is evident. In the rest of both cameras, the arrival times are scattered incoherently since they are not associated with light coming from the same shower. The time maps, therefore, can be helpful to discern the arrival direction of the shower since they strengthen the charge information.

Image cleaning

In the MAGIC standard analysis, the charge and time pixels maps are cleaned with different techniques described (Sec. 2.1.3). The maps are summarized with the Hillas (Hillas (1985)) parameters of the images. Then all the maps are discarded.

Several tests of cleaning were done by changing configurations on charge maps and time maps. Some examples of cleaning both for charges and arrival times are shown in Fig. 4.3. In order to be more efficient, once the charge maps were cleaned, the same selection mask was applied to select the pixels for the time cleaning.

The right choice of cleaning level is very important because it has significant effects on the CNN performance, a non-optimal choice could take to a wrong selection of ON and OFF events, thereby tainting the signal.

An example of raw versus cleaned image is shown in Fig.4.4. In this case, we used the 6-3.5 absolute cleaning, which turned out to be an effective choice. The

[2]IBM Power System AC922 (512 GB RAM)

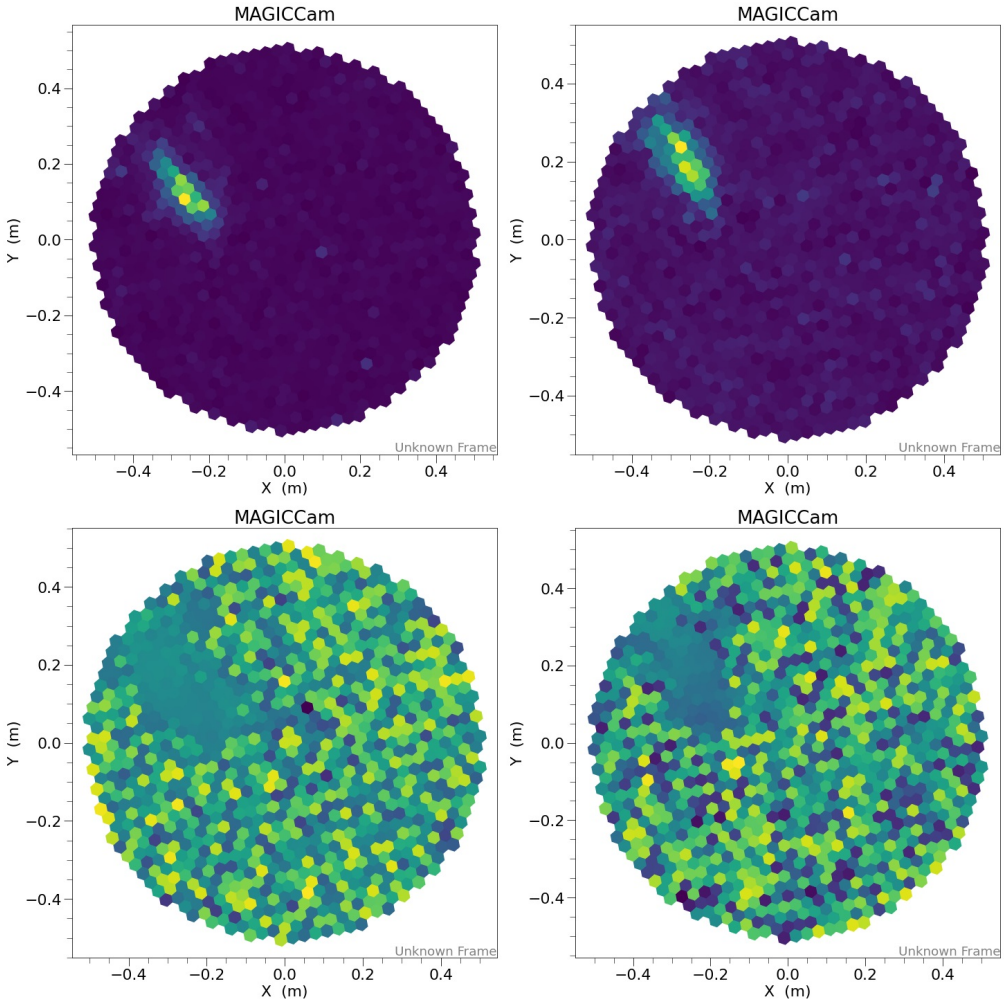


Figure 4.2: Example of raw charges distribution and arrival times distributions. On the left (right) column the MAGIC 1 (MAGIC 2) cameras are shown. The top (bottom) images show the charge (arrival time) distributions on the cameras.

images cleaning is performed with the help of the `mars_cleaning_1st_pass` script of the `ctapipe` package (Kosack et al. (2022)).

The cleaning must be performed before the next step of hexagonal to square interpolation, otherwise the charges and arrival time parameters after the square interpolation should be recalculated.

Hexagonal-to-Square transformation

In general, an image is represented by a three-index tensor which can be figured out as a superposition of three matrices, each corresponding to one of the three colors

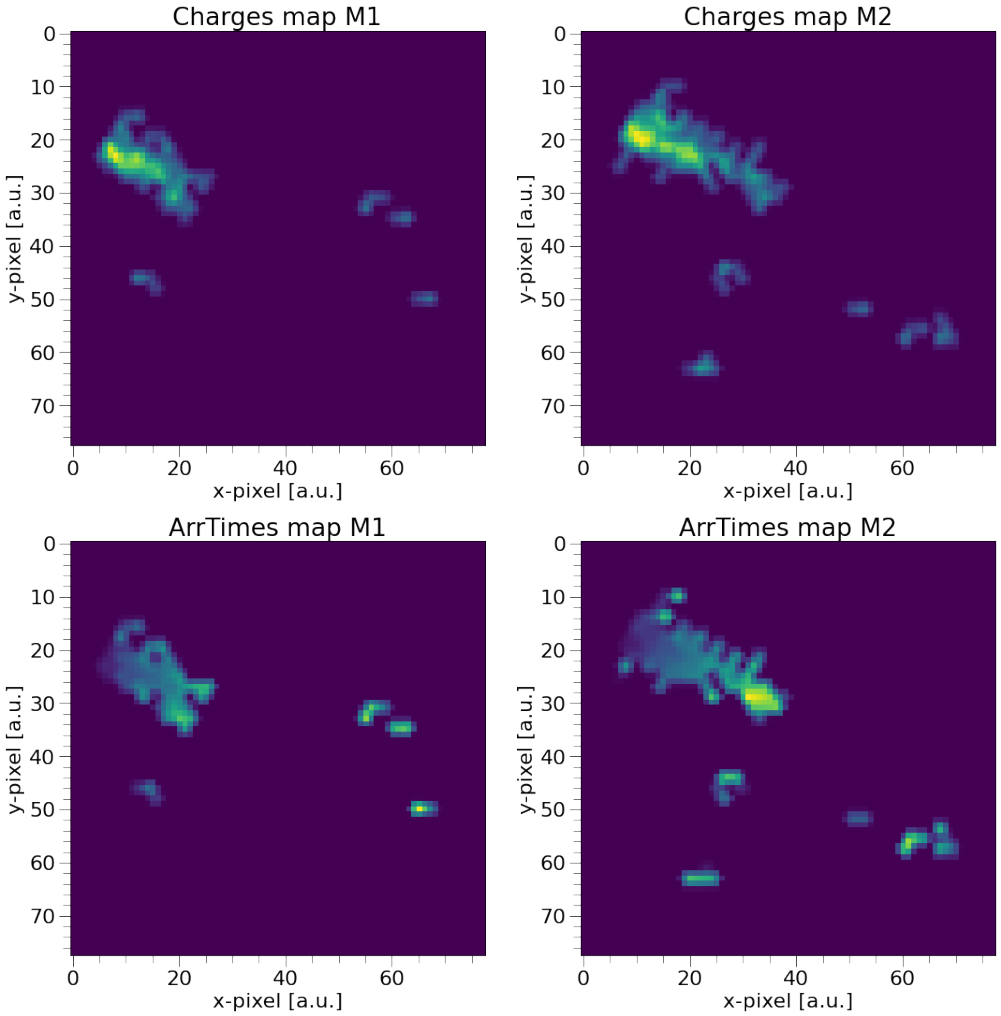


Figure 4.3: Square grid maps after 6-3.5 cleaning applied on charge and arrival time distributions. On the left(right), the MAGIC-1(MAGIC-2) camera is shown; on the top(bottom), the charge(arrival time) distributions on the camera.

(Red, Green, Blue). Usually, the first two indices of each matrix are respectively the Cartesian axes x and y, and the third index, called *channel*, specifies which color. Our samples are composed of black and white images, so we will have one channel only.

To build our CNN we use Tensorflow, a public-domain A.I. package by Google (Abadi et al. (2016)). Tensorflow is designed to process square images. In the case of MAGIC, the camera images are displayed in hexagonal fashion and they must be converted with an interpolation technique.

To convert the images we use the custom python package `DL1-DATA-HANDLER`

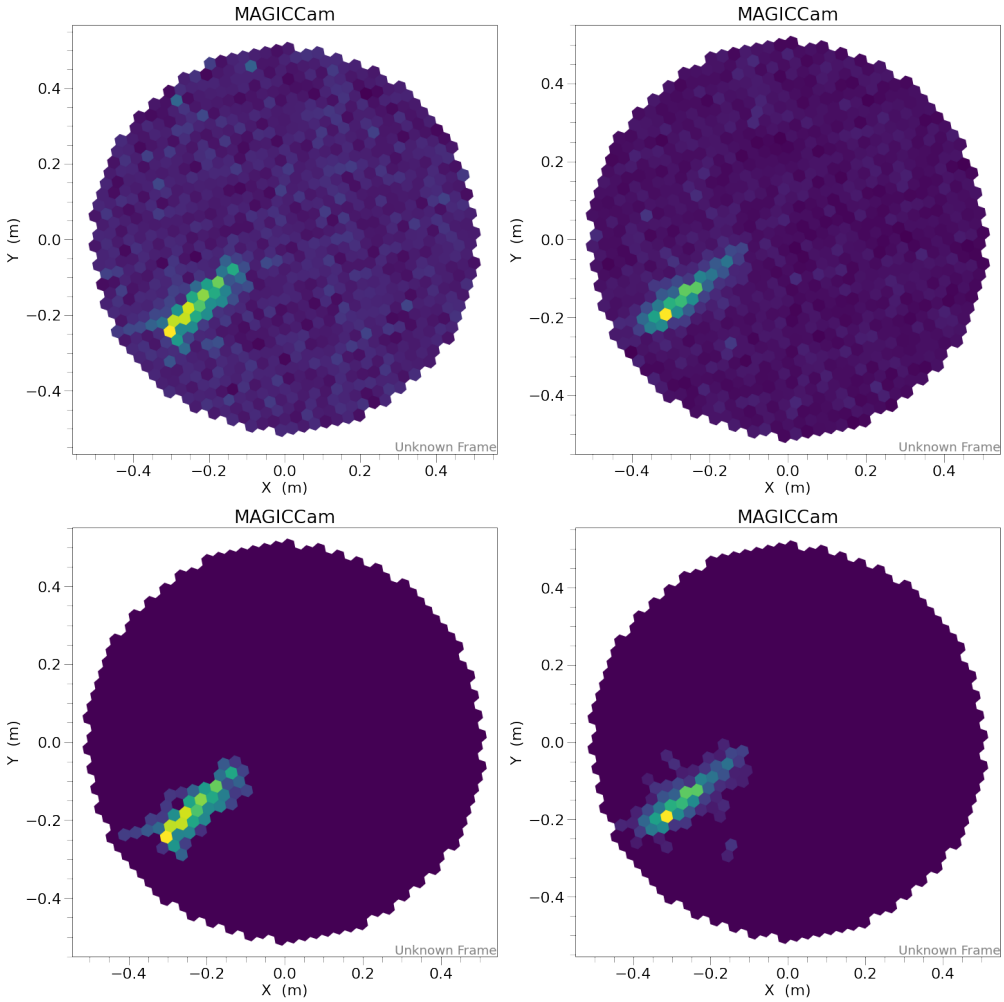


Figure 4.4: Charge maps (hexagonal grid) for raw data (top) and after 6-3.5 cleaning. (bottom) for MAGIC-1 (left) and MAGIC-2 (right).

(Kim et al. (2022)). This is a useful tool, written by the MAGIC collaboration group, to convert images from the hexagonal to square ones. The data-handler allows various types of interpolation: we used the bilinear interpolation one. An example of a hexagonal image and the corresponding square image is shown in Fig. 4.5. It must be noted that the data-handler applies a fixed rotation to all images, since the rotation is the same for all images and cameras are round in first approximation CNN should not be affected by this rotation.

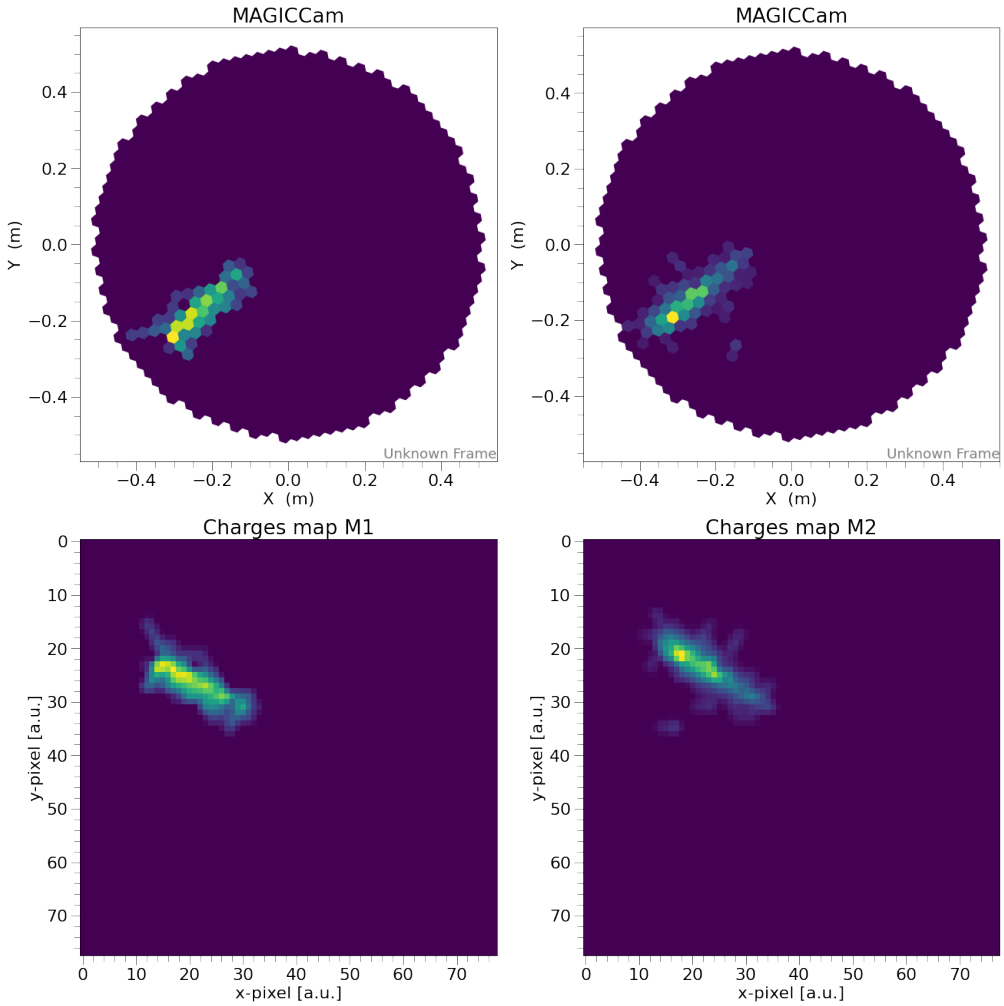


Figure 4.5: Effect of 6-3.5 cleaning for MAGIC 1 (left) and MAGIC 2 (right) for hexagonal (top) and square (bottom) images. It can be noticed from the images that the DL1-Data-handler pads the images.

Selection cuts

Gamma-like events can be filtered in the sample by selecting the events coming from the source direction.

The selection on the arrival direction can be done using the variable θ^2 (Sec. 2.2.2). For the θ^2 calculation the *Disp* parameter was used (Sec: 2.2.1). The process is the same for both telescopes and is shown in Tab. 4.2 for MAGIC 1.

A cut $\theta^2 \leq 0.02$ is typically used to select the ON circle and 3 symmetric OFF circles.

Fig.4.6 shows the hadronness distributions obtained with two different values

Name	Description
$dispx_1$	shower incoming direction x coordinate
$dispy_1$	shower incoming direction y coordinate
$srcposx_1 = srcposx_{1mm} * mm\_to\_deg$	source position on the camera x coordinate
$srcposy_1 = srcposy_{1mm} * mm\_to\_deg$	source position on the camera y coordinate

Direction	θ^2 equation
ON (0°)	$\theta_0^2 = (dispx_1 - srcposx_1)^2 + (dispy_1 - srcposy_1)^2$
OFF (180°)	$\theta_{180}^2 = (dispx_1 + srcposx_1)^2 + (dispy_1 + srcposy_1)^2$
OFF (90°)	$\theta_{90}^2 = (dispx_1 - srcposx_1)^2 + (dispy_1 + srcposy_1)^2$
OFF (270°)	$\theta_{270}^2 = (dispx_1 + srcposx_1)^2 + (dispy_1 - srcposy_1)^2$

Table 4.2: Position of the incoming direction of the shower (disp) on the sky respect the four θ^2 positions (ON and 3 OFF) calculated in degree

of θ^2. The top plot, obtained with $\theta^2 \leq 0.4$, shows that both the ON and OFF samples are still contaminated. The bottom plot, for a $\theta^2 \leq 0.02$ cut, shows that the ON and OFF distributions have distinct peaks, with a clear departure for $hadronness \sim 0.08$. This indicates an excess of gamma-like events towards the source direction.

Although not optimal, this cleaning can be used as an initial training sample for the CNN, the ON data as a gamma-like sample, and the OFF data as hadron-like sample. However, the attempted tests have shown that a *hadronness* cut is necessary although it might induce a bias of the CNN by data coming from the RF of the Standard Analysis 4.3.1.

Hadronness cut

Training a supervised CNN with these samples can bring to wrong results because the model could mix the gamma-like and hadron-like events. Several attempts have been made in this work to reduce the RF bias, from different *hadronness* cuts for the gamma-like and hadron-like samples (≤ 0.3 and ≥ 0.7 respectively), as well as harder cuts. The best value has been found in a symmetrical cut, $hadronness \geq 0.5$ for gamma-like and $hadronness \geq 0.5$ for the hadron-like sample.

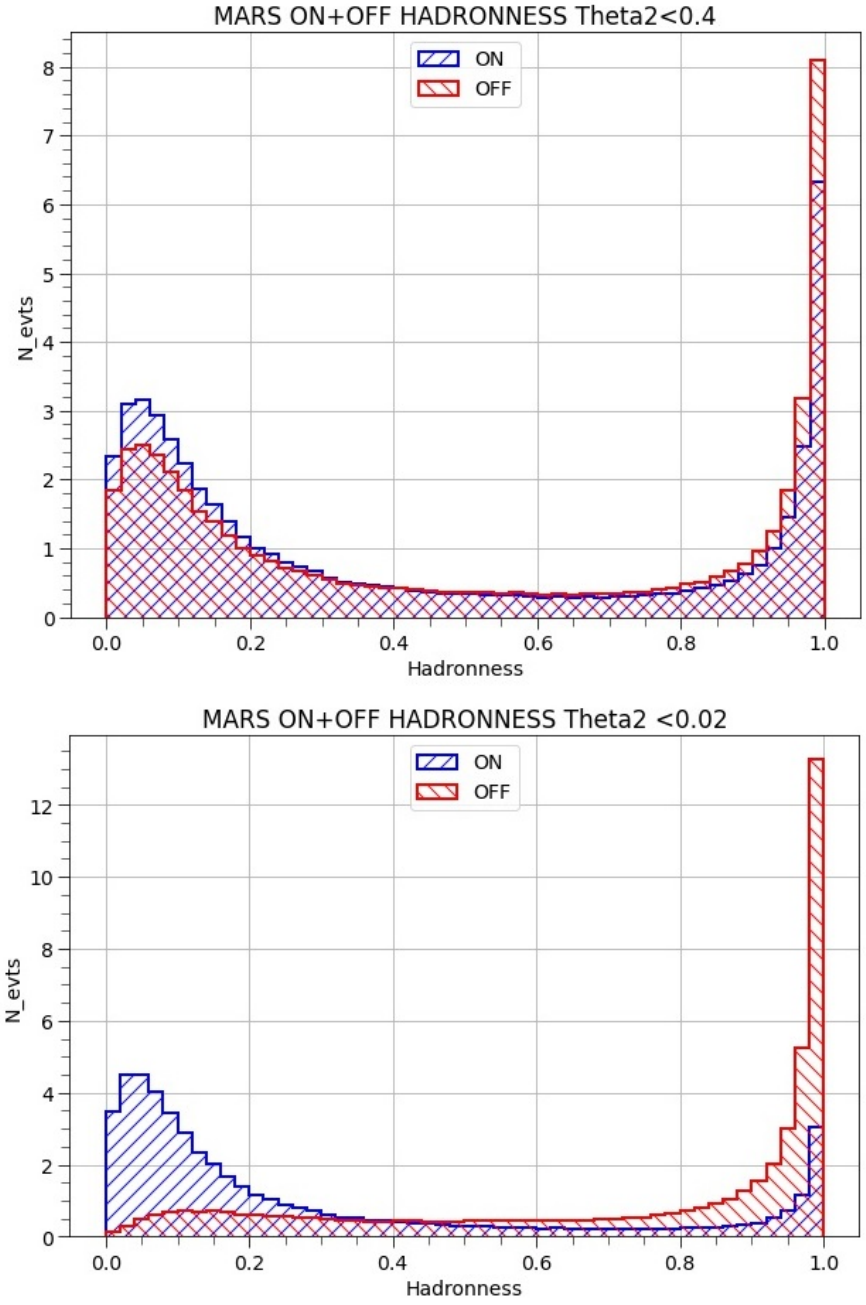


Figure 4.6: Standard analysis hadronness distributions for training set, with $\theta^2 \leq 0.4$ (top) and with $\theta^2 \leq 0.02$ (bottom)

Data augmentation

Data augmentation (Sec. 3.4.9) technique was implemented with the ImageDataGenerator from the Tensorflow `keras` package (Abadi et al. (2016)). The random

rotation of (180 deg) and the random flipping horizontal/vertical of the images have been chosen, in this way only one dataset can generate many different images configuration (ON and OFF samples) that are isotropically distributed on a ring (given that $\theta^2 \leq 0.02$) in the MAGIC FoV.

The images are augmented during the training phase, the original training set is augmented in different ways for every epoch and the same rotation/flipping is applied to both image sets. The image's brightness is not modified otherwise the total energy in the images could be affected. Furthermore, the images are not translated to avoid border and padding effects (Fig. 4.7).

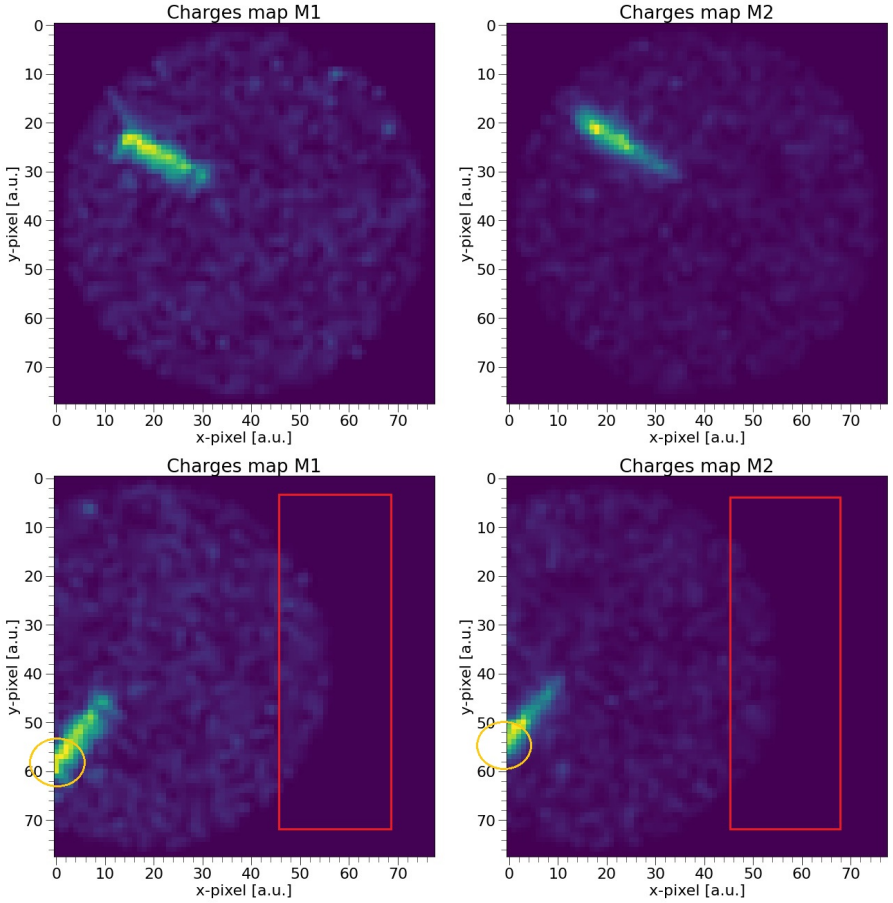


Figure 4.7: Example of roto-translational augmentation applied to a pair of raw images (originals in upper row, augmented images in lower row) for MAGIC-1 (left) and MAGIC-2 (right). The red rectangles show that, without proper image cleaning, the edges do not smoothly connect with the background of the original image. The yellow circles show that there can be a crop effect when the shower is near the camera edges.

4.4 The Convolutional Neural Network model

.

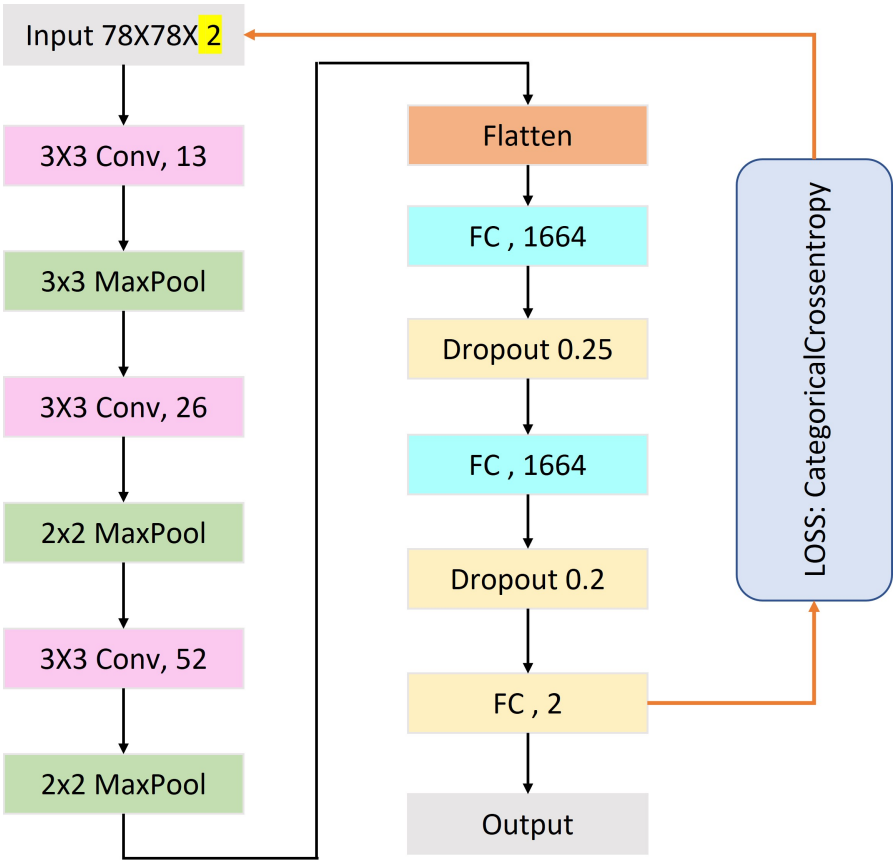


Figure 4.8: The custom CNN used to classify gamma-like/ hadron-like events. This CNN has been built to manage charge information only. To consider the arrival time maps too, the dimensions of the channel have to be doubled (yellow sign on the top `input` block: 2 to 4)

In the initial phase of this work, we built a custom CNN model (Sec. 4.8), on the basis of that described in Piccardo et al. (2021). In a second phase, the Inception Resnet V2 (Szegedy et al. (2016)) network schema was implemented.

The Inception Resnet V2 is a very deep CNN, its convolutional part can be used with its pre-trained weights to extract the features. Finally, a dense network classifies the extracted features.

We used the pure schema only, without the pre-trained weights. The input structure of the ResnetV2 was adapted to our data and, as for the output, dense layers for classification were attached to it. The Resnet V2 model requires a 3-

channel matrix as input, one for each color for each object of the sample. We used 2 of the 3 channels, one for each telescope camera. The structure of the input dataset has resulted: $id_{img}, X_{axis}(78), Y_{axis}(78), telescope(1\,or\,2)$. In this structure, at this time, I am considering the charges only, and I am not leveraging the arrival time information. Subsequently, to include the arrival time information, I modified the structure setting 4 channels in the following sequence: $id_{img}, X_{axis}(78), Y_{axis}(78), telescopes(1\,and\,2)charges, telescopes(1\,and\,2)arrival-time$.

4.4.1 Training and Validation

To training sample for the network has been selected from the whole dataset obtained with these cuts:

θ^2	≤ 0.02
$size$1&2	≥ 50
Start date	10-04-2013
End date	16-04-2013

We used the train_test_split function from `sklearn` package (Pedregosa et al. (2011)) that allows to randomly split the dataset in two sub-sets of data. We chose to split in training 0.8 (80%) and test 0.2 (20%). The validation dataset instead is extracted directly during the training, and it consists of a 20% of the training dataset, in this case too.

As we have discussed in Sec. 4.3.1 about Data Augmentation, we chose to use the ImageDataGenerator function available in the Tensorflow package. During the training cycle, for each "epoch", augmented images are generated from the original ones. In this way, the images are always different for each epoch, which helps in avoiding overfitting issues. We also used the `shuffle = True` option to randomize the order of images within our dataset on each epoch and the "class_weight" option to balance the number of events of the classes. This is particularly critical when the events in the classes are dramatically imbalanced, as in our case. The weight of each class is obtained by dividing by the number of elements in that class and multiplying by half of the total population.

4.4.2 Test results

We initially tested the network with the split part of the main dataset Mrk421 (10-16 April Mrk421), obtained as described in the previous section. Then, to obtain a more general result, we applied the trained model to different runs. We chose another Mrk421 sample, acquired ten days before the main dataset. During this run, an Mrk421 flare was observed (Sec. 4.5). In addition to that, we also

used events from the 1ES0927+500 run (Sec. 4.5.3) regarding the Crab Nebula (Sec. 4.5.2) and BL Lac (Sec. 4.5.4).

The results obtained from all the tests we executed and shown in Tab. 4.3 lead to affirm that the best combinations of cuts and cleaning for the training sample is:

- $\theta^2 \leq 0.02$
- $hadroness \leq 0.5$ (gamma-like/ON)
- $hadroness \geq 0.5$ (hadron-like/OFF)
- charge cleaning 6-3.5 plus arrival times

For evaluating the CNN performance, a useful variable is $Excess = ON - OFF$ which should show a monotonous rising trend as a function of the number of OFF events since its value cannot be < 0, apart from small fluctuations. If the monotonous behavior is not observed, it means that an overestimation of ON events is probably happening.

During the training with the Mrk421 dataset, we observed that the presence of noise in the images could have an impact on the CNN performance. For example, we obtained good results with raw images on the flare dataset, but when compared to other datasets, it was clear that the more effective configuration seemed the cleaning level 6-3.5. In this condition, the ROC curve still reaches an excellent value of 98% and a very low rate of False Negative/True Positive events (Fig. 4.9).

In Tab.4.3 we show a summary of all tests.

4.5 Cross-Check on Control Datasets

After the Mrk421 test, we selected other datasets for cross-check testing: samples coming from observation of the Crab Nebula, which is the reference source for gamma-rays experiments (2.4.2), from the 1ES0927+500, a source not detected by MAGIC and used as a dark source, and from the BLLac source.

The test on 1ES0927+500 is particularly interesting to verify if the network is influenced by the arrival direction of the events. As we discussed in the previous sections, for the training, we chose to consider the events toward the source as gamma-like events and all the others as hadron-like events. This way of building the dataset is based exclusively on the arrival direction, and for this reason, it could introduce a bias in the network. By randomly rotating the arrival direction in data augmentation, we successfully managed to avoid this kind of bias.

During the test procedures, we followed a specific workflow to consider different aspects. First, we checked the θ^2 plot of Standard Analysis and the CNN

Id	γ-like	h-like cut	Time	Cleaning	LOSS	Accuracy
1	No	No	No	raw	0.416	0.825
2	No	No	No	8-4-0	0.472	0.785
3	$\leq$0.5	$\geq$0.5	No	raw	0.086	0.979
4	$\leq$0.5	$\geq$0.5	No	8-4-0	0.173	0.943
5	$\leq$0.5	$\geq$0.5	No	6-3.5-0	0.159	0.963
6	$\leq$0.5	$\geq$0.5	Yes	raw	0.079	0.978
7	$\leq$0.5	$\geq$0.5	Yes	8-4-0	0.191	0.948
8	$\leq$0.5	$\geq$0.5	Yes	6-3.5-0	0.133	0.956
9	$\leq$0.5	$\geq$ 0.5	Yes	4-2-0	0.0937	0.9763
10	$\leq$0.5	$\geq$ 0.5	No	4-2-0	0.1142	0.9736

Table 4.3: Training results of the Inception Resnet V2 with training data cut at $\theta^2 \leq 0.02$ as a function of γ/h-like cuts and levels of absolute cleaning, and with or without the arrival-time maps. Although the best result seems to be the 6 "raw (no-cleaning)", when we tested the model on several data samples, we concluded that the optimal configuration is 8 "6-3.5-0".

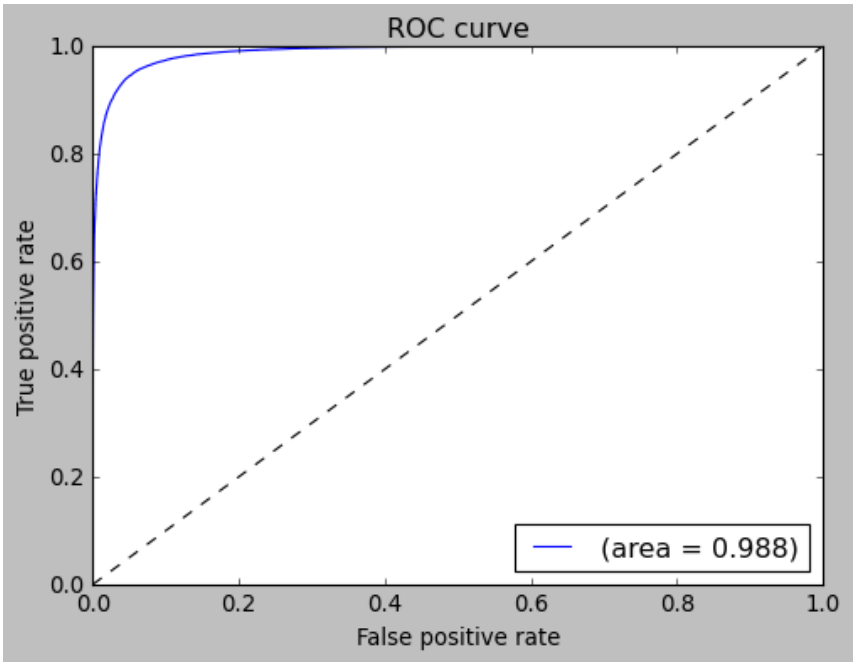


Figure 4.9: Typical ROC curve of the Mrk421 test dataset, obtained using the Inception Resnet V2, 6-3.5 absolute cleaning, charge maps, and arrival-time maps data.

	6-3.5 cleaning + arrival time	raw + arrival time
Total test events	48410	48410
True Negative	19460	20272
True Positive	26370	26968
False Negative	1193	595
False Positive	1387	575
Classification error	0.0533	0.0242
Sensitivity	0.957	0.978
False Positive Rate	0.066	0.0276
Precision	0.950	0.979

Table 4.4: Mrk421 test results obtained with the model based on Inception Resnet V2 and the sample acquired between April 10th and April 16th. As from the classification error values, the results for 6-3.5 absolute cleaning + arrival time maps (left column) are less performant than those obtained from raw images + time (right column).

to evaluate the γ-ray excess in the direction of the source. We calculated the $Excess(OFF)$ to check if the plot is monotonously growing or not (apart from small fluctuations). If this does not happen, it means that the number of OFF is steadily increasing more than the number of ON, and this is against the hypothesis of pointing in the source direction. In fact, the number of OFF should always be lesser or equal to the number of ON, due to the isotropic distribution of the background events. In figure 4.10 an example of "good" $Excess(OFF)$ and "bad" $Excess(OFF)$ plot is shown.

The plot of $\frac{Excess}{OFF^{1/2}}$ vs OFF is used as the method sensitivity and to compare it with the Standard Analysis. Finally, $LiMa(OFF)$ and the $LiMa(hadronness)$ plots show the significance. Also, the CNN vs Standard Analysis confusion matrix is plotted in (Fig. 4.11) to compare the classification performed by the two methods. All these results are obtained for LE cuts (Tab.2.1).

4.5.1 Markarian 421

The θ^2 plots are compared in figure 4.12, LE Standard Analysis (top) versus this work (bottom).

We used the Standard Analysis options `Odie.nWobbleOff=3` and `Odie.skipNormalization=TRUE` we can notice that the statistics is very similar.

We remind that, before this test, we had executed the Mrk421 cross-check, i.e. that the Excess (Sec.4.5) was actually steadily growing as expected.

Figure 4.13) summarizes the cross check on Mrk421. We want to point out that

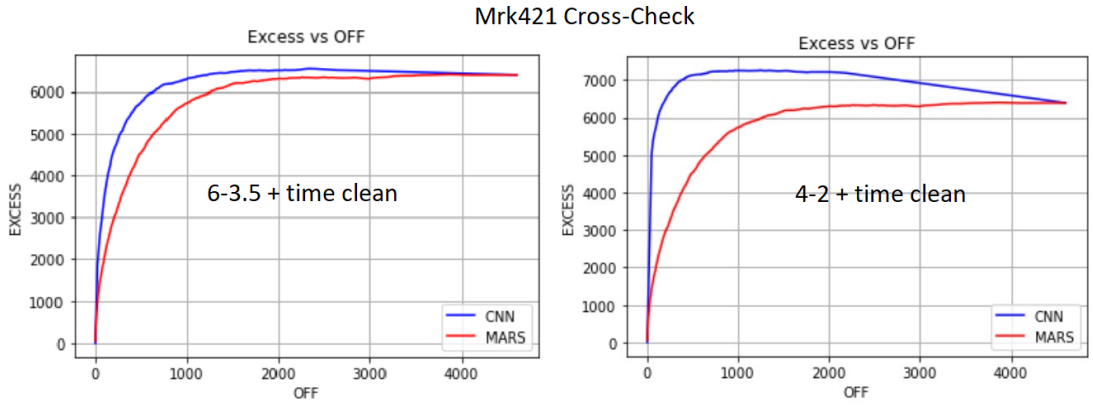


Figure 4.10: Cross-check on Mrk421 using the training sample with 6-3.5 charge and time cleanings (left plot). Same plot but using 4-2 charge cleaning plus arrival times cleaning (right plot).

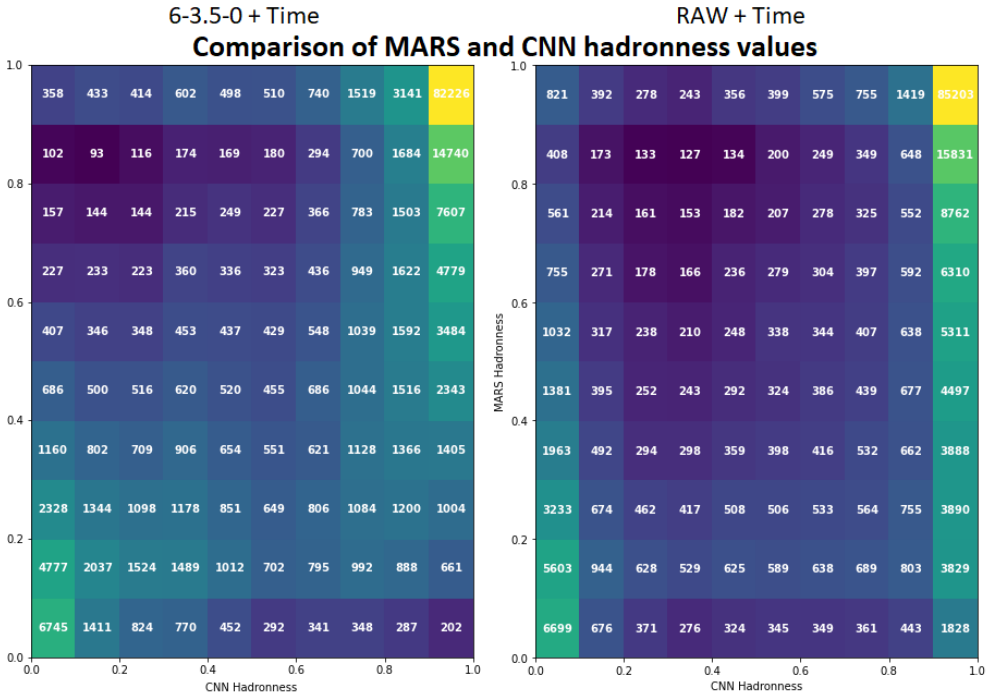


Figure 4.11: Differences between CNN (x) and Standard Analysis (y) algorithms for the Mrk421 cross check sample. The 6-3.5 cleaning plus arrival time maps has been used both for the test and the training datasets (left) versus the "no cleaning" case (right).

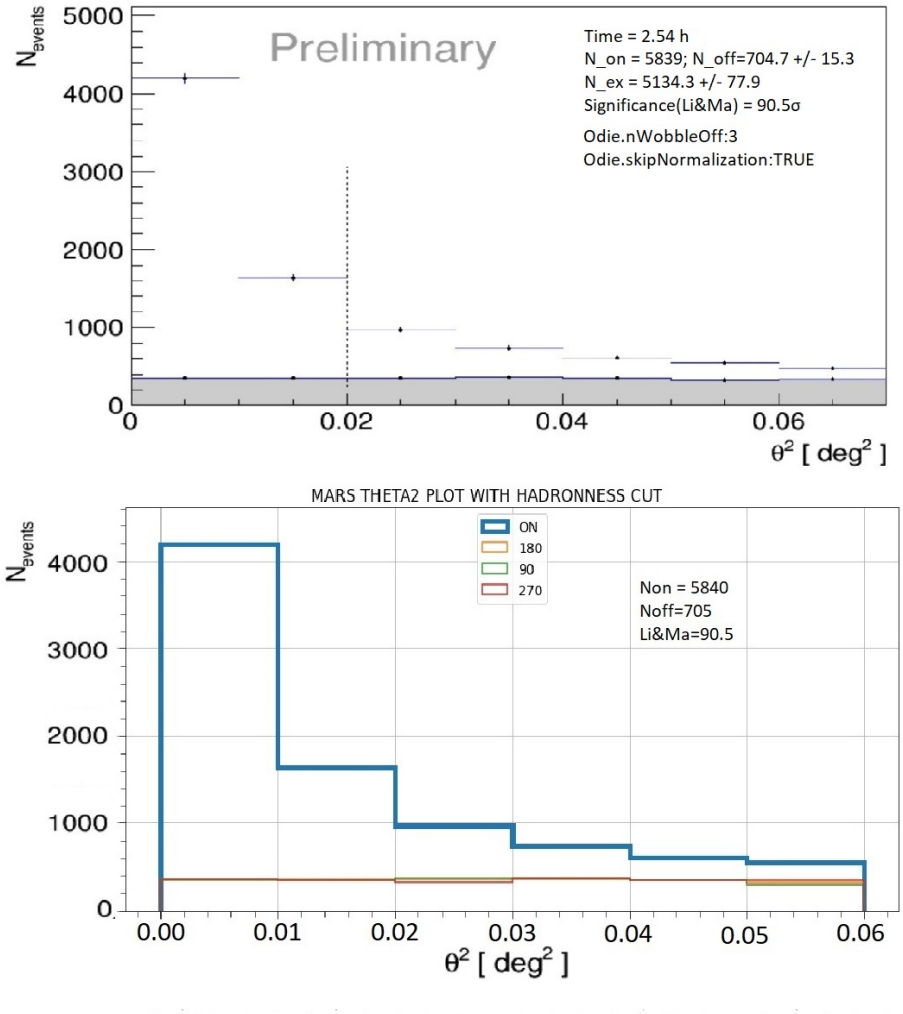


Figure 4.12: Mrk421 cross-check sample, θ^2 plots comparison after LE cuts. Standard Analysis plot (top) versus this work (bottom). Standard Analysis options: Odie.nWobbleOff=3 and Odie.skipNormalization=TRUE

the CNN analysis shows a significant improvement with respect to the Standard Analysis. The efficiency comparison shows also that, using the cleaning previously discussed and the arrival time maps, the efficiency is slightly higher than the counterpart.

4.5.2 Crab Nebula

The results of the Crab Nebula Cross Check are shown in Fig. 4.14 and 4.15.

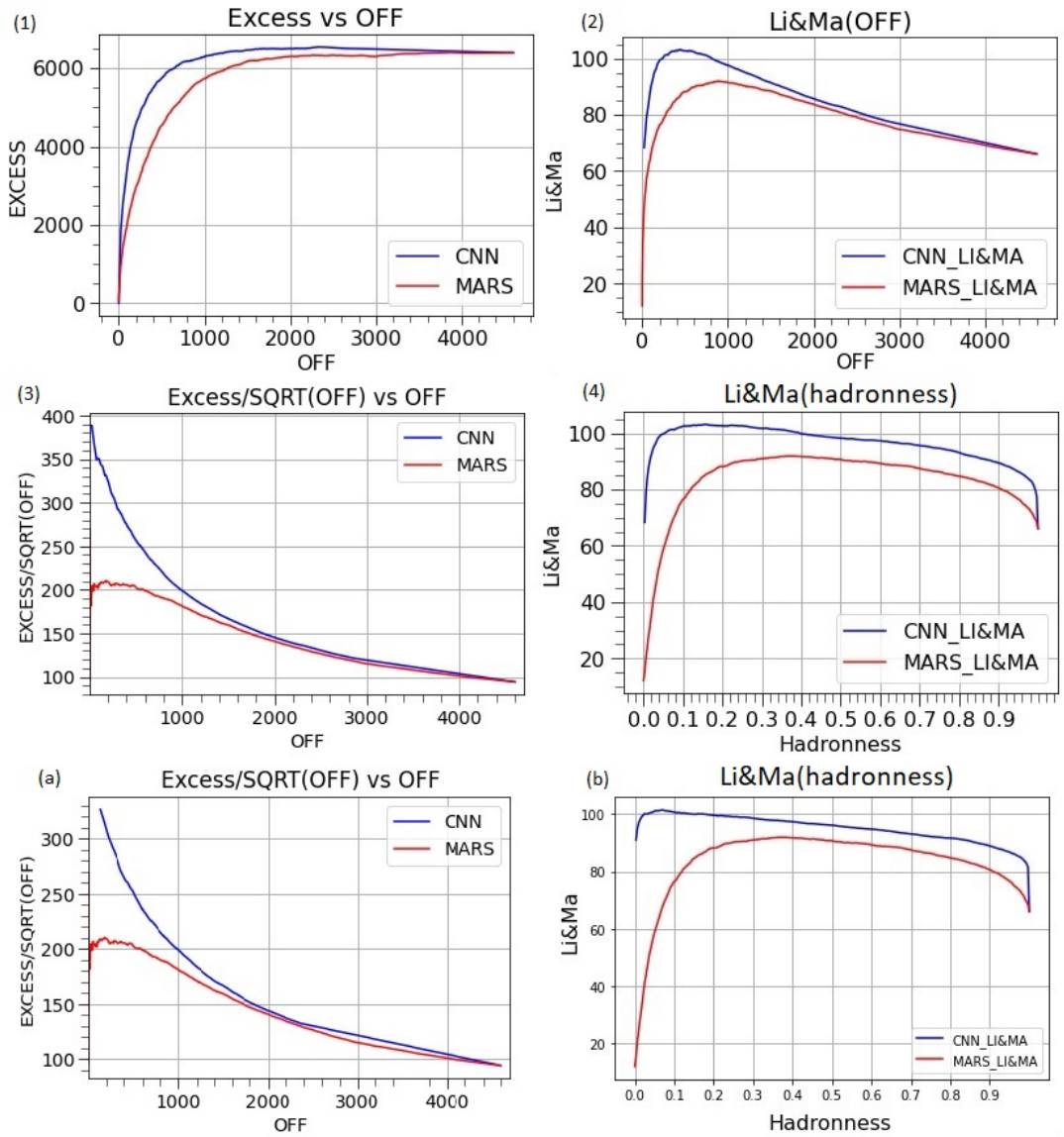


Figure 4.13: Mrk421 cross-check results. The plots show the performance of the CNN versus Standard Analysis as a function of selected variables. Plots (1) through (4) are obtained with Inception Resnet V2 row id=8 (6-3.5 absolute cleaning including the arrival time maps) 4.3. Plots (a) and (b) are obtained using charge maps cleaning only.In this case, the $Excess/SQRT(OFF)$ shows for the CNN a better performance then the SA

An unexpected outcome of the Crab Nebula cross-check has been that the CNN was performing worse than the Standard Analysis with arrival times cleaning (Fig.

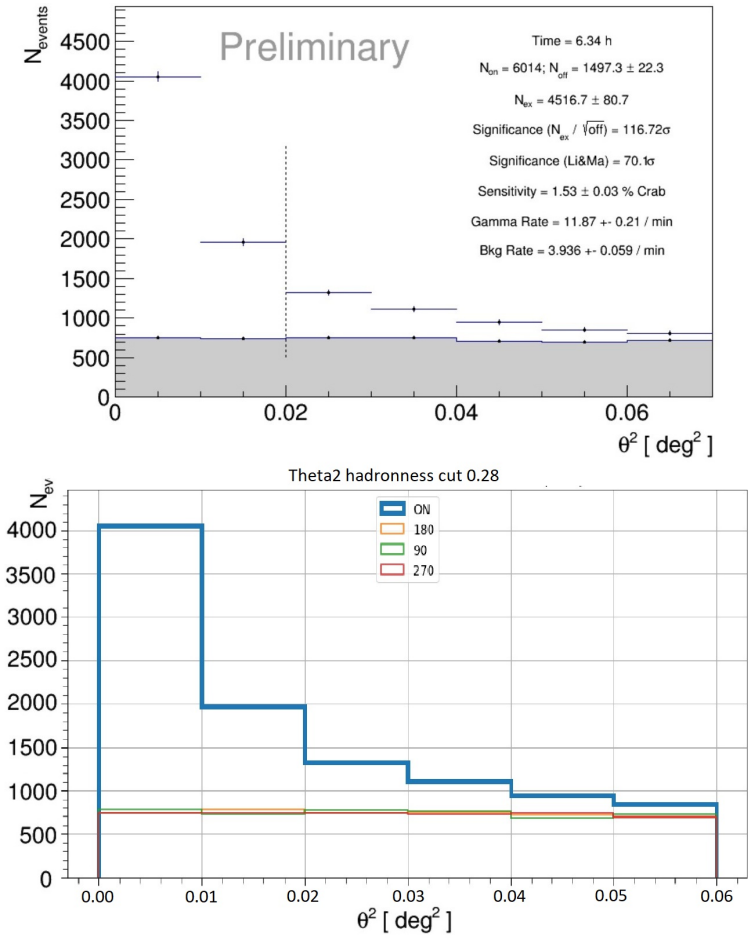


Figure 4.14: Crab Nebula cross-check sample, θ^2 plots comparison after LE cuts. Standard Analysis plot (top) versus this work (bottom). Standard Analysis options: Odie.nWobbleOff=3, Odie.skipNormalization=TRUE

4.15.(2)) compared when it was not applied (Fig. 4.15.(b)).

This effect can be caused by the ZTauri star presence in the Crab Nebula FoV that could influence the image classification, a bit enhanced when time cleaning is applied. Another cause of the discrepancy is that the Crab is much closer to the galactic plane than the training source (Mrk421), and its nature is very different since it is a Pulsar Wind Nebula. This is a case of extreme interest because it enhances the features of training and image classification and is worth future studies.

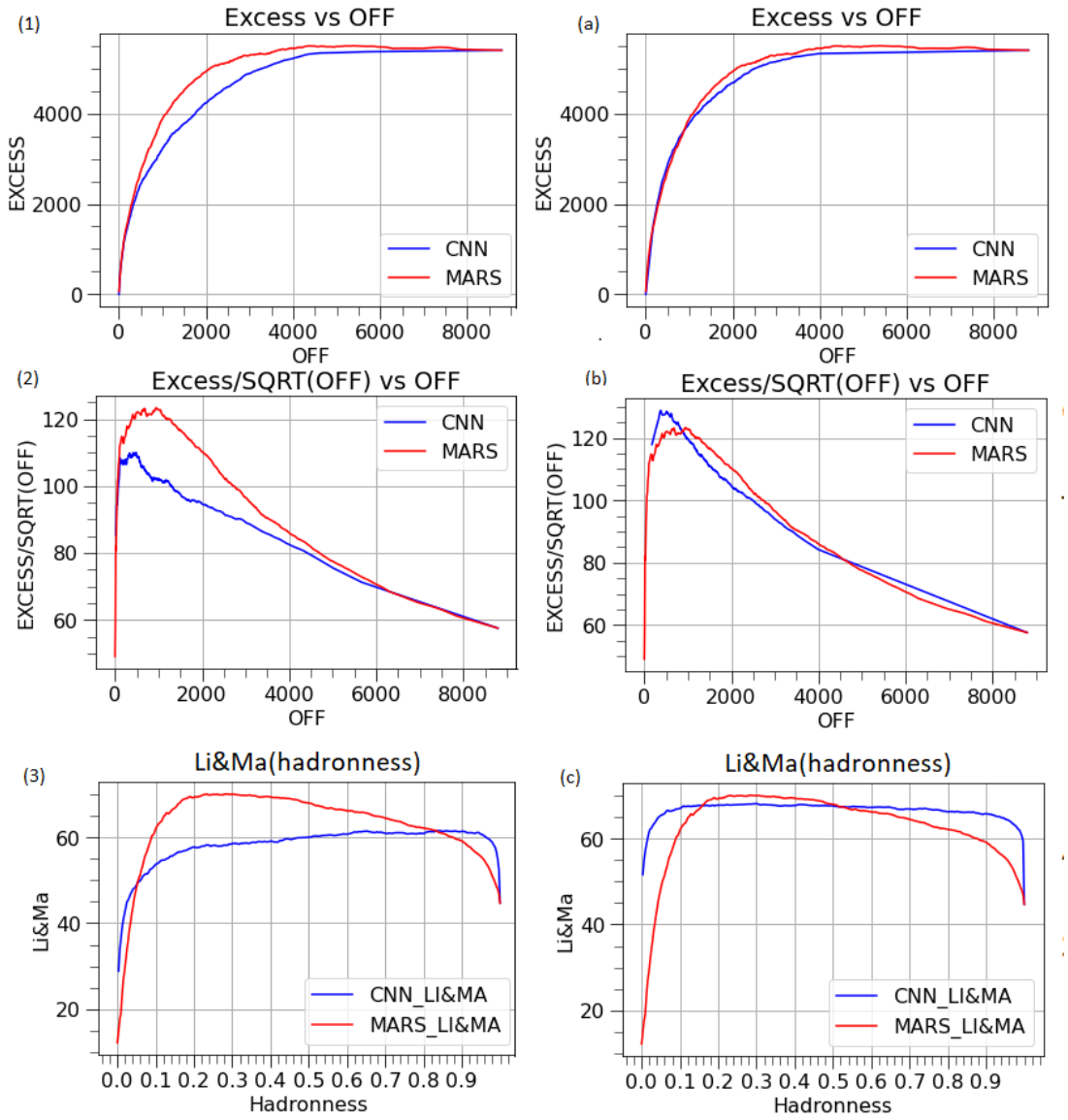


Figure 4.15: Performance of the CNN versus Standard Analysis as a function of selected variables on the Crab Nebula cross-check dataset with Inception Resnet V2 row id=8 (left, 6-3.5 absolute cleaning including the arrival time maps) versus Inception Resnet V2 row id=5 (right, no arrival time maps) 4.3.

4.5.3 1ES0927+500 Dark Patch

A useful test to check the good functioning of the CNN consists in considering one dark VHE source, like 1ES0927+500 (described in Sec. 2.4.4). "Dark source" means that during the MAGIC observation, it did not produce gamma-like flux.

Consequently, the CNN θ^2 plots must show a trend of the ON data that equals the trend of OFFs ones. This was actually verified, and it is shown in Fig. 4.16.

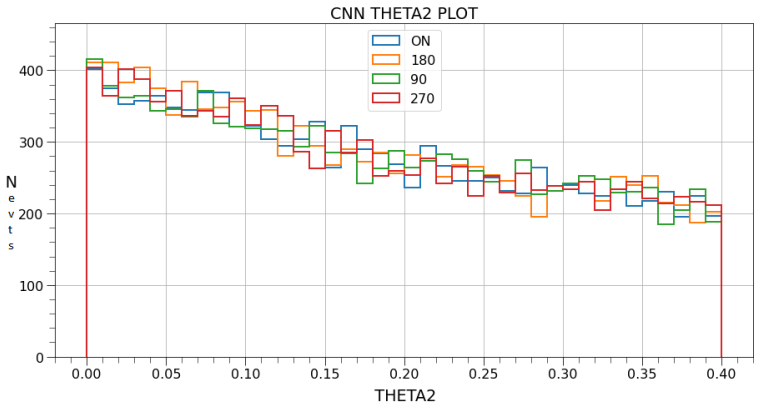


Figure 4.16: θ^2 obtained with CNN id=8 (table: 4.3) applied to 1ES0927+500 dark patch sample (same cleaning+time map), the plots show a similar trend for the ON and the 3 OFF samples (180, 90, 270 degrees).

The second check regards the hadronness histogram of the CNN. It should be concentrated in the right zone because the major part of the events of the 1ES0927+500 are hadron-like. Both the hadronnes distribution of the CNN and the Standard Analysis (which is always performed to compare the CNN results with) showed a correct distribution of the hadronness (Fig.4.17).

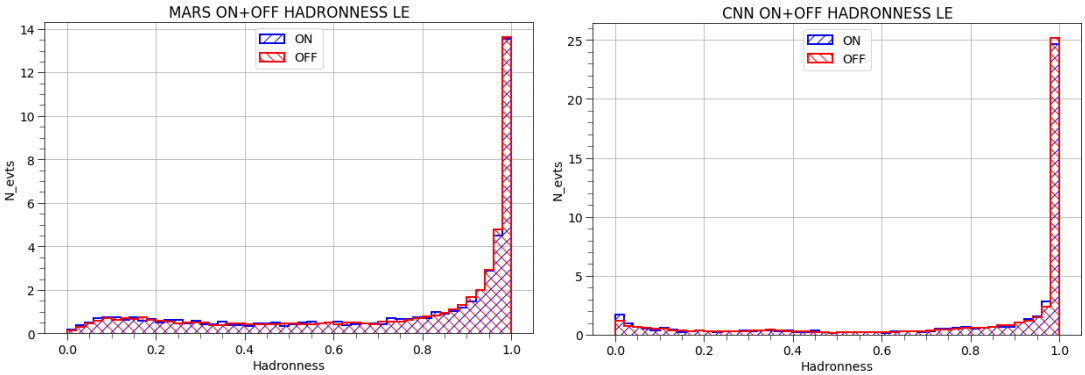


Figure 4.17: Normalized hadronness distributions for 1ES0927+500 dark patch. Standard Analysis (left) versus CNN (right) for ON data (blue) and OFF data (red).

4.5.4 BL Lac

The results of two CNN networks are summarized in Fig. 4.18 where the results regarding the CNN id=8 are shown in plots (1) to (4), while the results of CNN id=5 on pots (a) and (b) (Tab. 4.3). The two CNN show similar behaviour with a slight better performance for CNN id=8 and an overall better performance with respect to the Standard Analysis. Considering that the BL Lac dataset and the training samples are from different periods, the results are very positive.

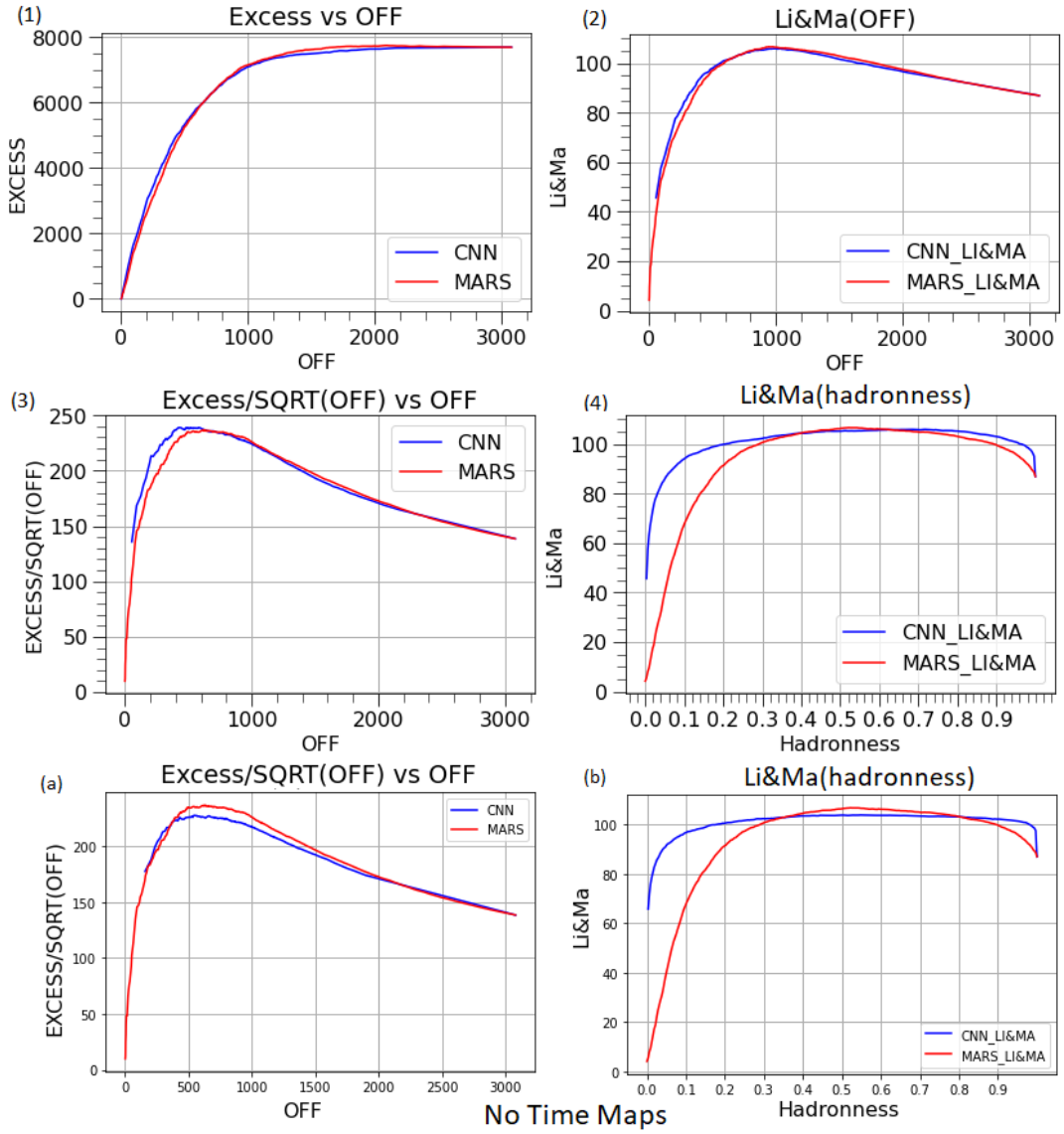


Figure 4.18: BL Lac cross check. Plots (1) through (4) are obtained when the CNN has been trained with charge maps and arrival time maps. For the bottom plots (a) and (b) the CNN has been trained with charge maps only. In this case, the $Excess/SQRTOFF$ shows for the CNN a performance similar to the SA.

4.6 Conclusions and future work

In this thesis a new method to perform the gamma/hadron separation on the data of the MAGIC telescopes was explored, based on CNN trained on real gamma-ray events. Several similar methods exist but generally rely on MC simulated data to train the CNN. This introduces potential systematics due to discrepancies between observed and simulated gamma-ray events.

Training a CNN with a sample of observed events is not an easy task. First, because CNNs are supervised algorithms that require well-known data during the training. Moreover, there are other issues to consider such as the amount of observed events. During a standard observation the telescopes acquisition rate amounts to $\sim 250\,\mathrm{Hz}$, implying that $\sim 9 \cdot 10^5$ total events are collected in a single hour of observation, mostly generated by a primary hadron. As a consequence, selecting a sample of gamma-like events suitable for training from an observed sample requires both a large storage system and a huge memory, in addition to high-end performing GPUs to process the data.

In this thesis, I have shown that CNNs provide a valid alternative to the standard method for classifying events in MAGIC. Unlike other approaches, I have studied in particular the possibility of applying CNN on the images before any parametrization is applied, thus exploiting the whole information embedded in the charge distribution recorded on the focal plane of the IACT. In addition to this, I have successfully started to involve in the analysis also the information carried by the distribution of arrival times.

Provided that, the method certainly has room for various further improvements. The competitive performance reached so far indicates promisingly that new analysis methods, significantly more efficient than the current MAGIC Standard Analysis, could become available in the next future. All these topics move beyond the scope of this thesis and will be the subject of further work in the future.

There are several potential further improvements to be investigated. The most used and developed DL techniques are the supervised ones and the unsupervised networks have greater room for improvement.

So, the first area to explore is the implementation of an unsupervised NN, to classify the observed event sample before the CNN training and avoid RF. This would be the starting point towards the implementation of an autoencoder. This would allow the separation of the observed events in classes in an unsupervised way and, potentially, also the recognition of other classes of events, like events initiated by $e^{\pm}$ versus γ-initiated ones, or even finding anomalous, or rare, events.

A very intriguing topic worth to investigate is a full event reconstruction with the measurement of two more quantities: the energy of the primary particle and its arrival direction. The task of energy estimation has relevant similarities with the task of event classification I tackled in my PhD work: for instance, existing

algorithms use MC simulated data, for which the true energy of the primary particle is given. Instead, the energy estimated from the observations derives from an indirect measurement, mainly related to the size that is the charge contained in the event image. Evaluating the event energy could be performed by an unsupervised NN via regression, analogous considerations hold for the determination of the arrival direction of the events.

Another relevant topic deals with the kind of network adopted and the algorithms that are implemented. For instance, a network exploiting not only the event images but also the image parameters can further improve the identification of the gamma-like events. In addition to this, my research was based on the Inception Res Net V2 network and without using the pre-trained weights, which could improve the network performance.

Moreover, the transfer learning and fine tuning techniques that reduce the amount of data and the time required to train other networks, could be studied. The transfer learning makes this algorithm very portable and, with small modifications, it would be possible to apply this algorithm to other situations and even to other topics. For example, the weights of the Mrk421 network can be stored and a new classifier layer can be added to classify the Crab Nebula or a BL Lac object. Eventually, the Inception Resnet V2 can be replaced with another CNN model with even better performances.

In conclusion, there are many things that can still be done to improve this work. Most of these look very promising and capable of improving significantly the analysis performance; for this reason I am willing to pursue this research path in future works.

Acknowledgment

Durante tutto il percorso di dottorato molte persone mi hanno aiutato in moltissimi modi diversi, alcune mi hanno aiutato a sviluppare parte del lavoro o mi sono state vicino e mi hanno incoraggiato. Vorrei quindi utilizzare queste righe per ringraziare tutte le persone che mi hanno permesso di realizzare questo lavoro.

Per primo vorrei ringraziare il mio supervisor Riccardo Paoletti per avermi consigliato in modo preciso e concreto durante tutto il dottorato e, per avermi proposto questo tipo di lavoro che fonde l'analisi di dati astrofisici con l'uso delle reti neurali che si accosta molto alle mie competenze professionali. Senza di lui, non avrei conosciuto Roberto Cappuccio, che ringrazio per avermi aiutato molto nell'apprendimento delle rete neurali e per avermi sempre incoraggiato. Ringrazio ancora Giacomo Bonnoli per il suo prezioso aiuto nella parte di analisi astrofisica e per avermi sempre consigliato ottimi ristoranti in ogni parte del mondo. E Tjark Miener che mi ha assistito in presenza durante il lavoro. Un ringraziamento va inoltre ad Abelardo Moralejo e Rubén López Coto grazie dai quali è partita l'idea originale di utilizzare dati di osservazioni reali. Infine vorrei ringraziare tutta la collaborazione di MAGIC per avermi aiutato e perché mi ha permesso di partecipare agli shift, durante i quali ho avuto modo di imparare a gestire i telescopi MAGIC.

Tra i miei affetti vorrei per prima ringraziare Neda per avermi sempre supportato e sopportato durante tutti questi anni di dottorato. Il mio caro amico Lorenzo che mi ha consigliato di fare domanda a Siena e con cui ho condiviso buona parte dell' esperienza di dottorato. Mia madre e mio padre per aver sempre creduto in me ed avermi permesso di scegliere la strada che preferivo. Infine vorrei ringraziare tutti i miei amici di Torino, Fossano, Cuneo, Siena, Como, Genova e Marsciano che mi sono stati vicini.

Grazie a tutti.

Bibliography

Abadi, Martín et al. (2016). "Tensorflow: Large-scale machine learning on heterogeneous distributed systems". In: *arXiv preprint arXiv:1603.04467*.

Abbott, B. P. et al. (Feb. 2016). "Observation of Gravitational Waves from a Binary Black Hole Merger". In: 116.6, 061102, p. 061102. DOI: 10.1103/PhysRevLett.116.061102. arXiv: 1602.03837 [gr-qc].

Abbott, B. P. et al. (Oct. 2017). "Multi-messenger Observations of a Binary Neutron Star Merger". In: *ApJL* 848.2, L12, p. L12. DOI: 10.3847/2041-8213/aa91c9. arXiv: 1710.05833 [astro-ph.HE].

Abdo, A. A. et al. (Jan. 2010). "Fermi Large Area Telescope Observations of the Crab Pulsar And Nebula". In: *ApJ* 708.2, pp. 1254–1267. DOI: 10.1088/0004-637X/708/2/1254. arXiv: 0911.2412 [astro-ph.HE].

Abdo, A. A. et al. (June 2011). "Observations of the Young Supernova Remnant RX J1713.7-3946 with the Fermi Large Area Telescope". In: *ApJ* 734.1, 28, p. 28. DOI: 10.1088/0004-637X/734/1/28. arXiv: 1103.5727 [astro-ph.HE].

Abdo, Aous A et al. (2009). "Fermi large area telescope observations of the Vela pulsar". In: *The Astrophysical Journal* 696.2, p. 1084.

Abraham, Ajith (2005). "Artificial neural networks". In: *Handbook of measuring system design*.

Acciari, V. A. et al. (June 2020). "Unraveling the Complex Behavior of Mrk 421 with Simultaneous X-Ray and VHE Observations during an Extreme Flaring Activity in 2013 April". In: *ApJSS* 248.2, 29, p. 29. DOI: 10.3847/1538-4365/ab89b5. arXiv: 2001.08678 [astro-ph.HE].

Acciari, VA et al. (2019). "A fast, very-high-energy γ-ray flare from BL Lacertae during a period of multi-wavelength activity in June 2015". In: *Astronomy & Astrophysics* 623, A175.

Acciari, Victor A et al. (2020). "Detection of the Geminga pulsar with MAGIC hints at a power-law tail emission beyond 15 GeV". In: *Astronomy & Astrophysics* 643, p. L14.

Acharya, BS et al. (2013). "Introducing the CTA concept". In: *Astroparticle physics* 43, pp. 3–18.

Aharonian, F. et al. (2006). "Observations of the Crab nebula with HESS". In: *Astronomy &amp Astrophysics* 457.3, pp. 899–915. DOI: 10.1051/0004-6361:20065351. URL: https://doi.org/10.1051%2F0004-6361%3A20065351.

Albert, J. et al. (2007). "Discovery of Very High Energy -Ray Emission from the Low-Frequency-peaked BL Lacertae Object BL Lacertae". In: *The Astrophysical Journal* 666.1, pp. L17–L20. DOI: 10.1086/521550. URL: https://doi.org/10.1086%2F521550.

Albert, J. et al. (2008a). "Implementation of the Random Forest method for the Imaging Atmospheric Cherenkov Telescope MAGIC". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 588.3, pp. 424–432. DOI: 10.1016/j.nima.2007.11.068. URL: https://doi.org/10.1016%2Fj.nima.2007.11.068.

Albert, J. et al. (2008b). "VHE -Ray Observation of the Crab Nebula and its Pulsar with the MAGIC Telescope". In: *The Astrophysical Journal* 674.2, pp. 1037–1055. DOI: 10.1086/525270. URL: https://doi.org/10.1086\%2F525270.

Aleksić, J. et al. (Dec. 2010). "Search for an extended VHE γ-ray emission from Mrk 421 and Mrk 501 with the MAGIC Telescope". In: *A&A* 524, A77, A77. DOI: 10.1051/0004-6361/201014747. arXiv: 1004.1093 [astro-ph.HE].

Aleksić, J. et al. (2012). "Performance of the MAGIC stereo system obtained with Crab Nebula data". In: *Astroparticle Physics* 35.7, pp. 435–448. DOI: 10.1016/j.astropartphys.2011.11.007. URL: https://doi.org/10.1016%2Fj.astropartphys.2011.11.007.

Aleksić, J. et al. (2015). "Measurement of the Crab Nebula spectrum over three decades in energy with the MAGIC telescopes". In: *Journal of High Energy Astrophysics* 5-6, pp. 30–38. DOI: 10.1016/j.jheap.2015.01.002. URL: https://doi.org/10.1016%2Fj.jheap.2015.01.002.

Aleksić, J. et al. (Apr. 2015). "The 2009 multiwavelength campaign on Mrk 421: Variability and correlation studies". In: *A&A* 576, A126, A126. DOI: 10.1051/0004-6361/201424216. arXiv: 1502.02650 [astro-ph.HE].

Aleksić, J. et al. (Jan. 2016). "The major upgrade of the MAGIC telescopes, Part I: The hardware improvements and the commissioning of the system". In: *Astroparticle Physics* 72, pp. 61–75. DOI: 10.1016/j.astropartphys.2015.04.004. arXiv: 1409.6073 [astro-ph.IM].

Aleksić, J. et al. (2016). "The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula". In: *Astroparticle Physics* 72, pp. 76–94. DOI: 10.1016/j.astropartphys.2015.02.005. URL: https://doi.org/10.1016\%2Fj.astropartphys.2015.02.005.

Aliu, E. et al. (Jan. 2009). "Improving the performance of the single-dish Cherenkov telescope MAGIC through the use of signal timing". In: *Astroparticle Physics*

30.6, pp. 293–305. DOI: 10.1016/j.astropartphys.2008.10.003. arXiv: 0810.3568 [astro-ph].

Ansoldi, S. et al. (Jan. 2016). "Teraelectronvolt pulsed emission from the Crab Pulsar detected by MAGIC". In: *A&A* 585, A133, A133. DOI: 10.1051/0004-6361/201526853. arXiv: 1510.07048 [astro-ph.HE].

Antonucci, Robert (Jan. 1993). "Unified models for active galactic nuclei and quasars." In: *ARA&A* 31, pp. 473–521. DOI: 10.1146/annurev.aa.31.090193.002353.

Atwood, WB et al. (2009). "The large area telescope on the Fermi gamma-ray space telescope mission". In: *The Astrophysical Journal* 697.2, p. 1071.

Bahcall, John N. and Eli Waxman (2003). "Has the GZK suppression been discovered?" In: *Physics Letters B* 556.1, pp. 1–6. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(03)00105-9. URL: https://www.sciencedirect.com/science/article/pii/S0370269303001059.

Berti, Alessio (2018). "Study of astrophysical transients with the MAGIC telescopes". PhD thesis. Trieste U.

Biteau, Jonathan (Feb. 2013). "A window on stochastic processes and gamma-ray cosmology through spectral and temporal studies of AGN observed with H.E.S.S." Theses. Ecole Polytechnique X. URL: https://pastel.archives-ouvertes.fr/pastel-00822242.

Borla Tridon, D. et al. (June 2009). "Performance of the Camera of the MAGIC II Telescope". In: *arXiv e-prints*, arXiv:0906.5448, arXiv:0906.5448. arXiv: 0906.5448 [astro-ph.HE].

Bretz, T et al. (2005). "Comparison of On-Off and Wobble mode observations for MAGIC". In: *Part 3: MAGIC Detector and Analysis Details*, p. 11.

Brun, Pierre (2012). "Seeking particle dark matter in the TeV sky". In: *AIP Conference Proceedings*. AIP. DOI: 10.1063/1.4732717. URL: https://doi.org/10.1063%2F1.4732717.

Buckley, J. H. et al. (May 1997). "Multiwavelength observations of Markarian 421". In: *Proceedings of the Fourth Compton Symposium*. Ed. by Charles D. Dermer, Mark S. Strickman, and James D. Kurfess. Vol. 410. American Institute of Physics Conference Series, pp. 1381–1385. DOI: 10.1063/1.54061.

Carreto Fidalgo, David (2019). "Cherenkov Telescopes and MAGIC". In: *Revealing the Most Energetic Light from Pulsars and Their Nebulae*. Cham: Springer International Publishing, pp. 49–81. ISBN: 978-3-030-24194-0. DOI: 10.1007/978-3-030-24194-0_3. URL: https://doi.org/10.1007/978-3-030-24194-0_3.

Čerenkov, Pavel Alekseevič (1937). "Visible radiation produced by electrons moving in a medium with velocities exceeding that of light". In: *Physical Review* 52.4, p. 378.

Dargan, Shaveta et al. (2020). "A survey of deep learning and its applications: a new paradigm to machine learning". In: *Archives of Computational Methods in Engineering* 27.4, pp. 1071–1092.

Davis, Jesse and Mark Goadrich (2006). "The relationship between Precision-Recall and ROC curves". In: *Proceedings of the 23rd international conference on Machine learning*, pp. 233–240.

De Angelis, Alessandro (2012). *L'enigma dei raggi cosmici*. Springer Milan.

de Naurois, Mathieu and Daniel Mazin (2015). "Ground-based detectors in very-high-energy gamma-ray astronomy". In: *Comptes Rendus Physique* 16.6. Gamma-ray astronomy / Astronomie des rayons gamma, pp. 610–627. ISSN: 1631-0705. DOI: https://doi.org/10.1016/j.crhy.2015.08.011. URL: https://www.sciencedirect.com/science/article/pii/S1631070515001462.

Dermer, Charles Dennison and Berrie Giebels (June 2016). "Active galactic nuclei at gamma-ray energies". In: *Comptes Rendus Physique* 17.6, pp. 594–616. DOI: 10.1016/j.crhy.2016.04.004. arXiv: 1602.06592 [astro-ph.HE].

DiSciascio, Giuseppe (2019). "Ground-based Gamma-Ray Astronomy: an Introduction". In: *Journal of Physics: Conference Series* 1263.1, p. 012003. DOI: 10.1088/1742-6596/1263/1/012003. URL: https://doi.org/10.1088/1742-6596/1263/1/012003.

Domingo-Santamaria, E. et al. (Jan. 2005). "The DISP analysis method for point-like or extended gamma source searches/studies with the MAGIC Telescope". In: *29th International Cosmic Ray Conference (ICRC29), Volume 5*. Vol. 5. International Cosmic Ray Conference, p. 363.

Doro, M. et al. (2008). "The reflective surface of the MAGIC telescope". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 595.1. RICH 2007, pp. 200–203. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2008.07.073. URL: https://www.sciencedirect.com/science/article/pii/S0168900208009765.

Escolano, Victor (2019). "Study of strategies for the identification of gamma-ray initiated showers with Cherenkov Telescopes". Universitat Autonoma di Barcellona.

Fermi, Enrico (Apr. 1949). "On the Origin of the Cosmic Radiation". In: *Physical Review* 75.8, pp. 1169–1174. DOI: 10.1103/PhysRev.75.1169.

Foffano, L et al. (2019). "A new hard X-ray-selected sample of extreme high-energy peaked BL Lac objects and their TeV gamma-ray properties". In: *Monthly Notices of the Royal Astronomical Society* 486.2, pp. 1741–1762. DOI: 10.1093/mnras/stz812. URL: https://doi.org/10.1093%2Fmnras%2Fstz812.

Fornasa, Mattia and Miguel A. Sánchez-Conde (2015). "The nature of the Diffuse Gamma-Ray Background". In: *Physics Reports* 598, pp. 1–58. DOI: 10.1016/

j.physrep.2015.09.002. URL: https://doi.org/10.1016\%2Fj.physrep.2015.09.002.

Gabici, S., D. Gaggero, and F. Zandanel (Oct. 2016). "Can supernova remnants accelerate protons up to PeV energies?" In: *arXiv e-prints*, arXiv:1610.07638, arXiv:1610.07638. arXiv: 1610.07638 [astro-ph.HE].

Gabici, Stefano and Felix Aharonian (July 2016). "Gamma-ray emission from young supernova remnants: Hadronic or leptonic?" In: *European Physical Journal Web of Conferences*. Vol. 121. European Physical Journal Web of Conferences, 04001, p. 04001. DOI: 10.1051/epjconf/201612104001. arXiv: 1502.00644 [astro-ph.HE].

Galama, T. J. et al. (Oct. 1998). "An unusual supernova in the error box of the γ-ray burst of 25 April 1998". In: *Nature* 395.6703, pp. 670–672. DOI: 10.1038/27150. arXiv: astro-ph/9806175 [astro-ph].

Gaug, Markus (Mar. 2006). "Calibration of the MAGIC telescope and observation of gamma ray bursts". PhD thesis. Autonomous University of Barcelona, Spain.

Ghisellini, Gabriele (2013). *Radiative Processes in High Energy Astrophysics*. Vol. 873. DOI: 10.1007/978-3-319-00612-3.

González Muñoz, Adiv (Jan. 2015). "Measurement of the gamma-ray opacity of the Universe with the MAGIC telescopes". PhD thesis. Autonomous University of Barcelona, Spain.

Häffner, Stephanie (2010). "Studies of cosmic-ray electrons with the H.E.S.S. experiment using an advanced reconstruction technique". MA thesis. U. Erlangen-Nuremberg (main).

Heck, D. et al. (Feb. 1998). "CORSIKA: A Monte Carlo code to simulate extensive air showers". In: *Forschungszentrum Karlsruhe Report FZKA 6019*.

Heitler, W. (1954). *Quantum theory of radiation*. 3rd ed. Oxford University Press.

Hess, Victor (July 2018). "On the Observations of the Penetrating Radiation during Seven Balloon Flights". In: *arXiv e-prints*, arXiv:1808.02927, arXiv:1808.02927. arXiv: 1808.02927 [physics.hist-ph].

Hester, J. J. (Sept. 2008). "The Crab Nebula : an astrophysical chimera." In: *ARA&A* 46, pp. 127–155. DOI: 10.1146/annurev.astro.45.051806.110608.

Hillas, A. M. (Aug. 1985). "Cerenkov Light Images of EAS Produced by Primary Gamma Rays and by Nuclei". In: *19th International Cosmic Ray Conference (ICRC19), Volume 3*. Vol. 3. International Cosmic Ray Conference, p. 445.

IceCube Collaboration et al. (July 2018). "Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A". In: *Science* 361.6398, eaat1378, eaat1378. DOI: 10.1126/science.aat1378. arXiv: 1807.08816 [astro-ph.HE].

Ioffe, Sergey and Christian Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *Proceedings*

of the 32nd International Conference on Machine Learning. Ed. by Francis Bach and David Blei. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, pp. 448–456. URL: https://proceedings.mlr.press/v37/ioffe15.html.

Jacquemont, Mikaël et al. (May 2021). "First Full-Event Reconstruction from Imaging Atmospheric Cherenkov Telescope Real Data with Deep Learning". In: *arXiv e-prints*, arXiv:2105.14927, arXiv:2105.14927. arXiv: 2105.14927 [astro-ph.IM].

Juryšek, Jakub, Etienne Lyard, and Roland Walter (Nov. 2021). "Full LST-1 data reconstruction with the use of convolutional neural networks". In: *arXiv e-prints*, arXiv:2111.14478, arXiv:2111.14478. arXiv: 2111.14478 [astro-ph.HE].

Kanbach, G et al. (1989). "The project EGRET (energetic gamma-ray experiment telescope) on NASA's Gamma-Ray Observatory GRO". In: *Space Science Reviews* 49.1, pp. 69–84.

Kanbach, Gottfried, Volker Schonfelder, and Alex Zehnder (Jan. 2010). "High-energy astrophysics - energies above 100 keV". In: *ISSI Scientific Reports Series* 9, pp. 57–71.

Kim, Bryan et al. (May 2022). *DL1-Data-Handler: DL1 HDF5 writer, reader, and processor for IACT data*. Version v0.10.6. DOI: 10.5281/zenodo.6535834. URL: https://doi.org/10.5281/zenodo.6535834.

Knödlseder, Jürgen (June 2016). "The future of gamma-ray astronomy". In: *Comptes Rendus Physique* 17.6, pp. 663–678. DOI: 10.1016/j.crhy.2016.04.008. arXiv: 1602.02728 [astro-ph.IM].

Kosack, Karl et al. (2022). "cta-observatory/ctapipe: v0. 13.0–2022-04-08". In: *Zenodo*.

Li, T. P. and Y. Q. Ma (Sept. 1983). "Analysis methods for results in gamma-ray astronomy." In: *ApJ* 272, pp. 317–324. DOI: 10.1086/161295.

Lide, David R (2004). *CRC handbook of chemistry and physics*. Vol. 85. CRC press.

Lipari, P. (2006). *Introduction to Neutrino Physics*. http://cds.cern.ch/record/677618/files/p115.pdf.

Lopez-Coto, Ruben (July 2015). "Very-high-energy gamma-ray observations of pulsar wind nebulae and cataclysmic variable stars with MAGIC and development of trigger systems for IACTs". PhD thesis. Autonomous University of Barcelona, Spain.

Lorimer, D. R. and M. Kramer (2004). *Handbook of Pulsar Astronomy*. Vol. 4.

Matthews, J. (2005). "A Heitler model of extensive air showers". In: *Astroparticle Physics* 22.5, pp. 387–397. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2004.09.003. URL: https://www.sciencedirect.com/science/article/pii/S0927650504001598.

Mészáros, Péter et al. (Oct. 2019). "Multi-messenger astrophysics". In: *Nature Reviews Physics* 1.10, pp. 585–599. DOI: 10.1038/s42254-019-0101-z. arXiv: 1906.10212 [astro-ph.HE].

Morlino, G. and D. Caprioli (Feb. 2012). "Strong evidence for hadron acceleration in Tycho's supernova remnant". In: *A&A* 538, A81, A81. DOI: 10.1051/0004-6361/201117855. arXiv: 1105.6342 [astro-ph.HE].

Naurois, Mathieu de (2006). *Analysis methods for Atmospheric Cerenkov Telescopes*. DOI: 10.48550/ARXIV.ASTRO-PH/0607247. URL: https://arxiv.org/abs/astro-ph/0607247.

Ng, Andrew Y (2004). "Feature selection, L 1 vs. L 2 regularization, and rotational invariance". In: *Proceedings of the twenty-first international conference on Machine learning*, p. 78.

O'Shea, Keiron and Ryan Nash (2015). "An introduction to convolutional neural networks". In: *arXiv preprint arXiv:1511.08458*.

Pacini, D. (Dec. 1912). "La radiazione penetrante alla superficie ed in seno alle acque". In: *Il Nuovo Cimento* 3.1, pp. 93–100. DOI: 10.1007/BF02957440. arXiv: 1002.1810 [physics.hist-ph].

Palacio, J., D. Navarro-Girones, and J. Rico (2019). "Pointing optimization for IACTs on indirect dark matter searches". In: *Astroparticle Physics* 104, pp. 84–90. DOI: 10.1016/j.astropartphys.2018.09.002. URL: https://doi.org/10.1016\%2Fj.astropartphys.2018.09.002.

Parsons, R. D. and S. Ohm (May 2020). "Background rejection in atmospheric Cherenkov telescopes using recurrent convolutional neural networks". In: *European Physical Journal C* 80.5, 363, p. 363. DOI: 10.1140/epjc/s10052-020-7953-3. arXiv: 1910.09435 [astro-ph.IM].

Pedregosa, Fabian et al. (2011). "Scikit-learn: Machine learning in Python". In: *the Journal of machine Learning research* 12, pp. 2825–2830.

Piccardo, Arnoldo et al. (2021). "The role of the deep convolutional neural network as an aid to interpreting brain [18F] DOPA PET/CT in the diagnosis of Parkinson's disease". In: *European Radiology* 31.9, pp. 7003–7011.

Pierre Auger Collaboration et al. (Sept. 2017). "Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8×10^{18} eV". In: *Science* 357.6357, pp. 1266–1270. DOI: 10.1126/science.aan4338. arXiv: 1709.07321 [astro-ph.HE].

Piran, T. (June 1999). "Gamma-ray bursts and the fireball model". In: 314.6, pp. 575–667. DOI: 10.1016/S0370-1573(98)00127-6. arXiv: astro-ph/9810256 [astro-ph].

Shilon, I. et al. (Feb. 2019). "Application of deep learning methods to analysis of imaging atmospheric Cherenkov telescopes data". In: *Astroparticle Physics* 105, pp. 44–53. DOI: 10.1016/j.astropartphys.2018.10.003. arXiv: 1803.10698 [astro-ph.IM].

Shorten, Connor and Taghi M. Khoshgoftaar (2019). "A survey on Image Data Augmentation for Deep Learning". In: *Journal of Big Data* 6.1, p. 60. ISSN: 2196-1115. DOI: 10.1186/s40537-019-0197-0. URL: https://doi.org/10.1186/s40537-019-0197-0.

Simonyan, Karen and Andrew Zisserman (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. DOI: 10.48550/ARXIV.1409.1556. URL: https://arxiv.org/abs/1409.1556.

Smith, Andrew J (2015). "HAWC: Design, operation, reconstruction and analysis". In: *arXiv preprint arXiv:1508.05826*.

Spurio, Maurizio (2015). *Particles and Astrophysics*. DOI: 10.1007/978-3-319-08051-2.

Srivastava, Nitish et al. (2014). "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15.1, pp. 1929–1958.

Swordy, S. P. (Oct. 2001). "The Energy Spectra and Anisotropies of Cosmic Rays". In: *Space Sci.Rev.* 99, pp. 85–94. DOI: 10.1023/A:1013828611730.

Szegedy, Christian et al. (2016). *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*. DOI: 10.48550/ARXIV.1602.07261. URL: https://arxiv.org/abs/1602.07261.

Tajbakhsh, Nima et al. (2016). "Convolutional neural networks for medical image analysis: Full training or fine tuning?" In: *IEEE transactions on medical imaging* 35.5, pp. 1299–1312.

Tescaro, Diego et al. (2013). "The MAGIC telescopes DAQ software and the on-the-fly online analysis client". In: *arXiv preprint arXiv:1310.1565*.

The MAGIC Collaboration et al. (Mar. 2022). "Multiwavelength variability and correlation studies of Mrk421 during historically low X-ray and γ-ray activity in 2015–2016". In: *37th International Cosmic Ray Conference. 12-23 July 2021. Berlin*, 866, p. 866.

Torii, Shoji, Pier Simone Marrocchesi, and Calet Collaboration (Dec. 2019). "The CALorimetric Electron Telescope (CALET) on the International Space Station". In: *Advances in Space Research* 64.12, pp. 2531–2537. DOI: 10.1016/j.asr.2019.04.013.

Urry, C. Megan and Paolo Padovani (Sept. 1995). "Unified Schemes for Radio-Loud Active Galactic Nuclei". In: *PASP* 107, p. 803. DOI: 10.1086/133630. arXiv: astro-ph/9506063 [astro-ph].

VERITAS Collaboration et al. (Oct. 2011). "Detection of Pulsed Gamma Rays Above 100 GeV from the Crab Pulsar". In: *Science* 334.6052, p. 69. DOI: 10.1126/science.1208192. arXiv: 1108.3797 [astro-ph.HE].

Vink, Jacco (2011). "Supernova remnants: the X-ray perspective". In: *The Astronomy and Astrophysics Review* 20.1. DOI: 10.1007/s00159-011-0049-1. URL: https://doi.org/10.1007\%2Fs00159-011-0049-1.

Weekes, T. C. et al. (July 1989). "Observation of TeV Gamma Rays from the Crab Nebula Using the Atmospheric Cerenkov Imaging Technique". In: *ApJ* 342, p. 379. DOI: 10.1086/167599.

Wilson-Hodge, Colleen A. et al. (2011). "WHEN A STANDARD CANDLE FLICKERS". In: *The Astrophysical Journal* 727.2, p. L40. DOI: 10.1088/2041-8205/727/2/l40. URL: https://doi.org/10.1088%2F2041-8205%2F727%2F2%2Fl40.

Zanin, Roberta et al. (2013). "MARS, the MAGIC analysis and reconstruction software". In: *Proc. of the 33st International Cosmic Ray Conference, Rio de Janeiro, Brasil*. Citeseer.

Zhu, Wen, Nancy Zeng, Ning Wang, et al. (2010). "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations". In: *NESUG proceedings: health care and life sciences, Baltimore, Maryland* 19, p. 67.

List of Figures

List of Tables

List of Abbreviations and Acronyms

γ-ray gamma-ray

γ ray gamma ray

γ-rays gamma-rays

AE Auto Encoder

AI Artificial Intelligence

ANN Artificial Neural Network

AS anticoincidence system

BP Back Propagation

CMB Cosmic Microwave Background

CNN Convolutional Neural Network

CR Cosmic Rays

DL Deep Learning

DNN Deep Neural Network

EAS Extensive Air Showers

EBL Extragalactic Background Light

EM Electromagnetic

FC Fully Connected

GD Gradient Descent

GPU Graphic processing unit

GZK Greisen–Zatsepin–Kuzmin limit

HBL High Energy Blazar

IACT Imaging Atmospheric Cherenkov Telescope

IR infrared

IRF Instrument Response Functions

LC Light Curve

LIGO Laser Interferometer Gravitational Wave Observatory

MAGIC Major Atmospheric Gamma Imaging Cherenkov telescope

MC Monte Carlo

ML Machine Learning

MRK421 Markarian 421

NN Neural Network

NSB Night Sky Background

PMT Photomultiplier Tube

RF Random Forest

SA Standard Analysis

SED Spectral Energy Distribution

SGD Stochastic Gradient Descent

SNR Supernova Remnant

VHE Very High Energy

Zd Zenith degree