UNIVERSITÀ DI PISA

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI

CORSO DI LAUREA IN FISICA

Anno Accademico 2013/2014

Tesi di Laurea Magistrale

Pass 8: development and science prospects for the new Fermi LAT event-level analysis

Relatore

Dott. Luca Baldini

Candidato

Alberto Manfreda

Contents

1 Introduction **1**

1.1 The Fermi Observatory 1

1.2 Scientific context 2

1.2.1 Dissecting the gamma-ray sky 3

1.2.2 Time domain astronomy 5

1.2.3 Indirect search for Dark Matter 6

1.2.4 Cosmic-ray studies 7

1.2.5 The LAT event-level analysis 8

1.3 Overview of the work 9

2 The Large Area Telescope **11**

2.1 The silicon tracker 12

2.2 The electromagnetic calorimeter 15

2.3 The anticoincidence detector 16

2.4 Orbital environment 17

2.5 Trigger and Data Acquisition 19

2.6 Monte Carlo simulation 22

2.7 Ghost events and Overlays 24

3 The Pass 8 event-level analysis **27**

3.1 Introduction 27

3.2 Event reconstruction 28

3.2.1 Calorimeter 28

3.2.2 Tracker 30

3.2.3 ACD 32

3.3 Event classification 33

3.3.1 Fiducial Cuts 34
3.3.2 Quality Cuts 35
3.3.3 Background Rejection Cuts 36
3.4 Detector Performance 36
3.4.1 Instrument Response Functions 36
3.4.2 Effective Area 38
3.4.3 Point Spread Function 40
3.4.4 Energy Dispersion 42

4 Estimating the quality of the energy measurement 45
4.1 Introduction 45
4.2 Data preparation 46
4.3 Classification Trees 48
4.3.1 Training 49
4.3.2 Variables 53
4.3.3 Testing 56
4.4 Using P_E in the search of spectral lines 62

5 High-Energy Electron and Positron Spectrum 69
5.1 Introduction 69
5.2 Event selection 72
5.2.1 Pre-selection 73
5.2.2 Removing alphas and heavy nuclei 73
5.2.3 Template fitting 76
5.2.4 Main selection 77

6 Conclusions 85

Riferimenti bibliografici 87

Chapter 1

Introduction

1.1 The Fermi Observatory

The *Fermi Gamma-ray Space Telescope (Fermi)* is a space observatory dedicated to the measure of the high-energy electromagnetic radiation. The LAT [Atw09] is the primary instrument of the *Fermi* mission. It is an imaging, wide field-of-view, high-energy γ-ray telescope, covering the energy range from 20 MeV to more than 300 GeV. The LAT is complemented on *Fermi* by the Gamma-ray Burst Monitor (GBM) [Mee09][Bis09], an array of 12 NaI(Tl) and 2 BGO scintillation detectors, distributed around the *Fermi* spacecraft, which extends the sensitivity of the telescope in the soft γ-ray and X-ray energy range (from 8 Kev to 40 MeV). *Fermi* was launched by NASA on 2008, June 11 and began nominal science operation on August 2008.

Figure 1.1 shows a sketch of the *Fermi* timeline, with the two vertical dashed lines representing the prime phase of the mission (i.e., the first five years of operation). Over the first six years of operation the LAT has been collecting data flawlessly at a rate of 60 billion triggers per year, with more than 50 million γ-ray candidates per years made publicly available to the community.

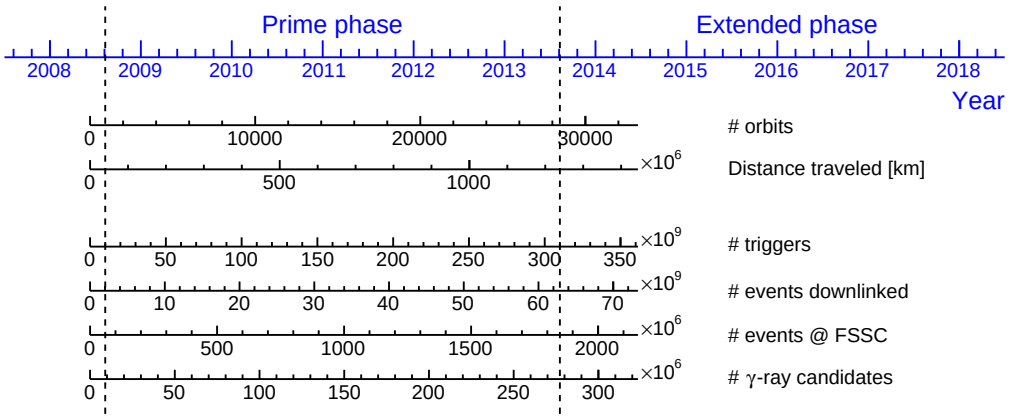


Figure 1.1. Timeline of the *Fermi* experiment.

1.2 Scientific context

One possible way about the taxonomy of the *Fermi* science topics is that of arranging them by energy, time and distance, as schematically shown in Figure 1.2. While the exact choice of items is largely arbitrary—as is the definition of some of the bounds—this kind of representation gives a sense of the enormous dynamic range of the observatory. Fermi is effectively studying the gamma-ray sky over 8 decades in energy (including the GBM) and over time scales spanning 13 orders of magnitudes—from the tens of microseconds characterizing the pulsar substructures to binary systems with orbital periods of years. In terms of distances, the *Fermi* science menu includes such diverse objects as Terrestrial Gamma-ray flashes and the gamma-ray emission of the Earth atmosphere (originating at a few hundred km from the spacecraft), solar system bodies such as the Moon and the Sun, Galactic and extra-galactic objects all the way up to GRBs at redshifts of 5 or more.

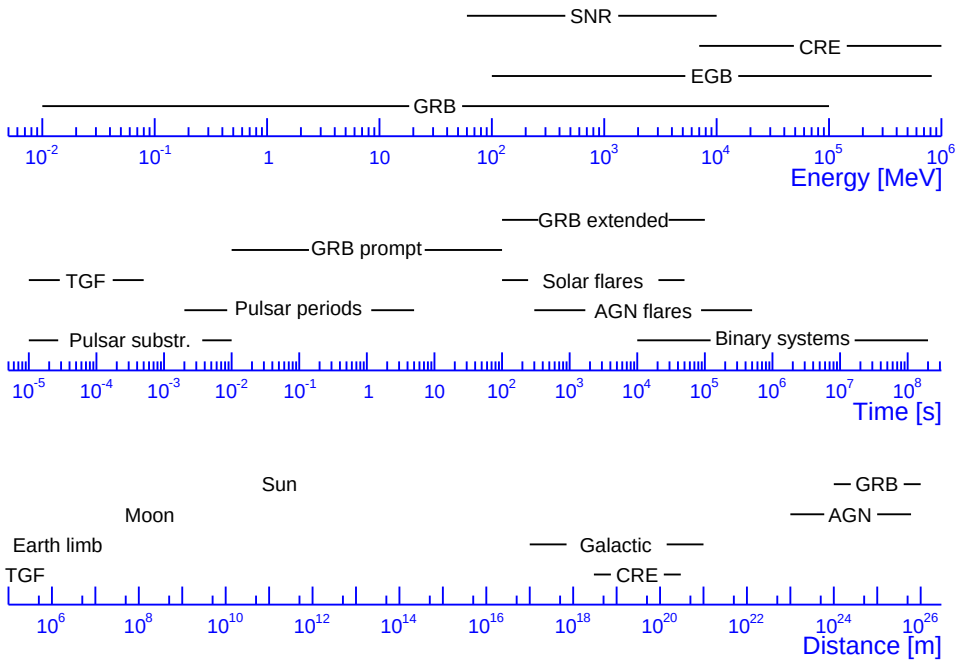


Figure 1.2. Scientific topics investigated by *Fermi* arranged by their energy, time and distance scales.

1.2.1 Dissecting the gamma-ray sky

Most of the celestial γ rays reaching the Earth are associated either with individual astrophysical sources or with a diffuse galactic emission (DGE), produced by cosmic-ray particles interacting with the gas and radiation fields in the interstellar medium (ISM). In addition, an isotropic Gamma-ray background (IGRB) permeates the whole observable sky. In the course of six years of mission, *Fermi* has provided remarkable contributes to the study of all these components.

Using the experience accumulated by its direct predecessor, the Energetic Gamma-Ray Experiment Telescope (EGRET), and provided with a better spatial resolution, an improved sensitivity and a wider energy range (from $\sim$ 20 MeV to more then 300 GeV), the LAT has allowed a high-statistics observations, with unprecedented accuracy, of γ-ray sources of known classes, as well as a discovery of new emitters.

The main products of this effort are the Fermi-LAT Source Catalogs

(FGL), which are periodically updated by the *Fermi* Collaboration. The most recent published version [Nol12] (2FGL) is based on 24 months of data and contains 1873 detected sources, including association with Galaxies, Active Galactic Nuclei (AGN), Pulsars, Pulsar Wind Nebulae (PWN), Supernova remnants (SNR), Globular clusters, Starburst galaxies and High-mass Binary systems (HMB), as well as others still unassociated sources (see Figure 1.3). The *Fermi* collaboration is working on a new global catalog (3FGL), based on 4 years of data, featuring ~ 3000 unique γ-ray sources. Dedicated catalog have also been published for pulsars [Abd13], AGN [Ack11b], variable sources [Ack13c] and high-energy sources above 10 GeV [Ack13d].

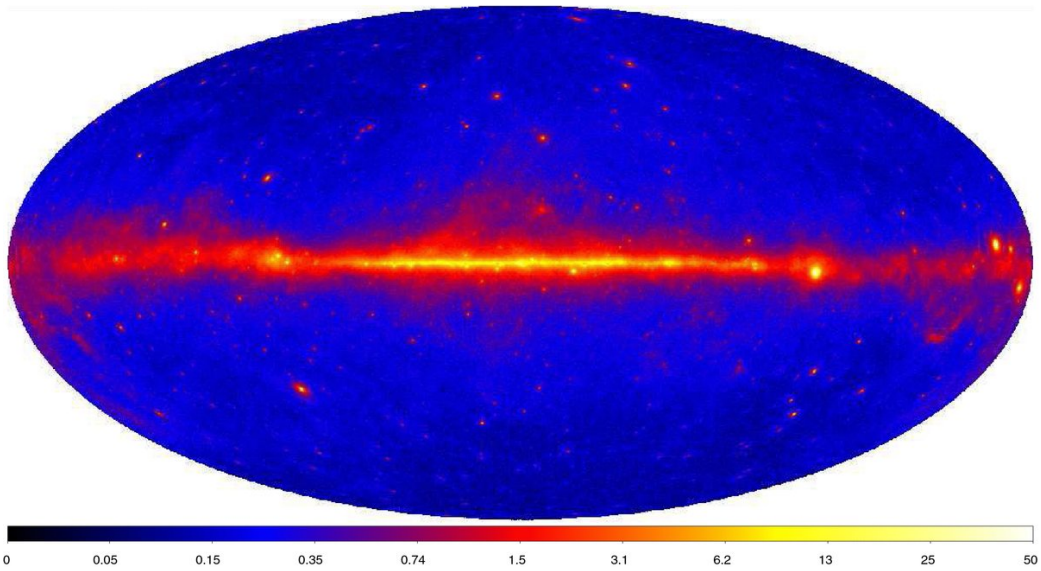


Figure 1.3. Full γ-ray sky map derived from the LAT data of the Second Fermi Catalog (first two years of data), Aitoff projection in Galactic coordinates. The image shows γ-ray energy flux for energies between 100 MeV and 10 GeV, in units of 10^{-7} erg cm^{-2} s^{-1} sr^{-1}.

The LAT has also performed a detailed study of the diffuse galactic emission [Ack12c]. As said, such emission is the result of cosmic rays (CRs) interactions in the galaxy; in particular the primary production process is pion decay from CR proton-nucleon collisions with gas of the interstellar medium, with other contributions coming from inverse Compton scatter-

ing (on the interstellar radiation field) and Bremsstrahlung scattering of CR electrons. In this sense, the data collected by *Fermi* have been used to constrain CR production and propagation models in the Milky Way.

Once one has accounted for all the known sources and the GDE, what is left is the isotropic γ-ray background, whose nature is still debated in literature. *Fermi*'s measure of the IGRB [Abd10b], has provided valuable information to understand the origin of such emission,

1.2.2 Time domain astronomy

The LAT's wide acceptance and field of view, togheter with its capability to continuously monitor the sky (see 2.4), makes it an ideal instrument for the study of transient phenomena, such as Gamma Ray Bursts (GRB), AGN and Novae.

GRB are isotropically distributed high-energy events, of typical durations spacing from a few milliseconds up to a few hours, characterized by an intense emission of γ-rays followed by a fading "afterglow" emission at higher wavelengths. GRBs have been observed for the first time in 1967 and their cosmological nature have been proved in 1997 with the first measure of redshift in the optical component of the afterglow [Par97].

One of the key science objectives of the *Fermi* mission is to study the nature and behavior of GRBs. The *Fermi* collaboration has published and keeps updated a catalog of GRB detected by the GBM [Kie14], currently including 953 transient phenomena identified as GRBs in the first four years of mission (Figure 1.4), complemented by an associated spectral GRB catalog [Gru14]. A dedicated catalog of 35 high-energy GRB observed by the LAT has been published by the collaboration as well [Ack13b].

On April, 27 2013 the GBM triggered on one of the brightest GRB ever registered [Ack13e], GRB 130427A, which was measured by the LAT as well. It had the largest fluence, highest-energy photon (95 GeV), longest γ-ray duration (20 hours), and one of the largest isotropic energy releases ever observed from a GRB.

In addition, GRB have also been used to test Lorentz invariance on cosmological scales [Abd09c].

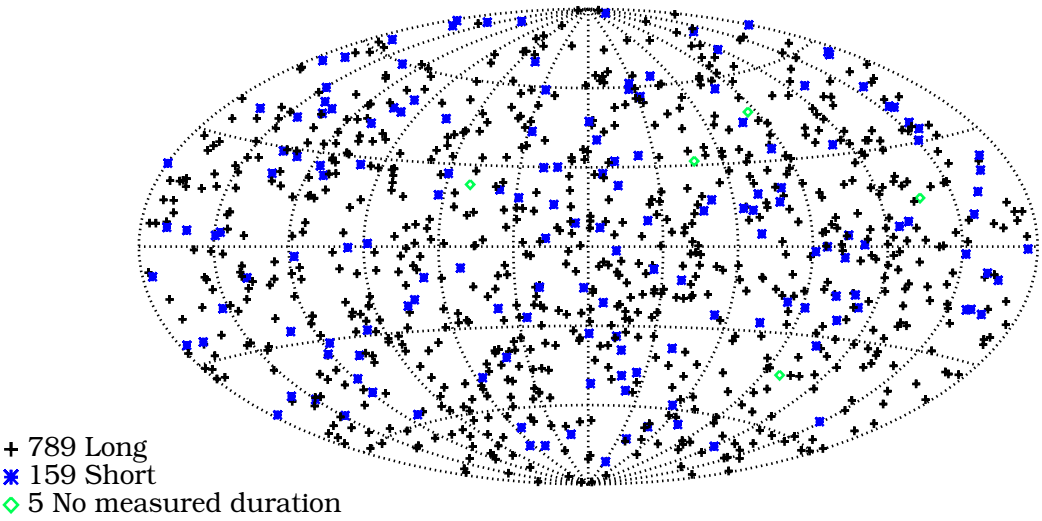


Figure 1.4. Sky distribution of GBM triggered GRBs in celestial coordinates (first four years of operation). Crosses indicate long GRBs (> 2 s); asterisks indicate short GRBs.

1.2.3 Indirect search for Dark Matter

According to cosmological models and to the most recent Planck observations [Ade13] $\sim 27\%$ of energy density in the Universe is constituted by nonbaryonic matter which does not absorb nor emit light, usually referred to as Dark Matter (DM).

Identification of DM with a Weakly Interactive Massive Particle (WIMP) has been proposed in many theoretical models [Mar11]. Depending on the specific model assumed, it is possible that the LAT can indirectly detect the presence of DM revealing the products of its interactions. The production of γ rays can happen in two forms: directly, if the WIMP particle is coupled to photons through loop diagrams [Jun96], leading to a non zero cross section for annihilation processes like $\chi\chi \rightarrow \gamma X$ or for decay processes like $\chi \rightarrow \gamma X$ (χ here indicates a hypothetical WIMP particle and X a generic other particle produced by the reaction, as adequate for the specific WIMP model); or indirectly, if they are produced in a later stage by other particles generated from WIMPs' interactions.

In the latter case the emission would have the form of a continuum spectrum, which could be really hard to distinguish from the background

of astrophysical sources. The case of a direct production, on the other hand, despite the low flux expected due to its second-order nature, would probably be more favorable for a detection. In fact, in the center-of-mass frame of those reactions, the photons produced are monochromatic. If we assume that the DM distribution producing the signal is non relativistic (with velocity v of the order of $\sim 10^{-3}$ c) then this would be approximately true in the laboratory frame as well [1]. As a consequence it is possible that DM can be revealed by a detector like the LAT in the form of a monochromatic line in the γ-ray spectrum.

An extensive search for similar signals has been performed by the *Fermi* collaboration in regions of the sky associated with the Galactic Center [Ack13] (see also section 4.4) and with Milky Way satellites Galaxies [Ack11], imposing upper limits on the cross section of these process.

A detailed knowledge of the instrument's energy resolution is a key ingredient for such analyses. In Chapter 4.4 we will describe a way to improve the sensitivity of the LAT to a monochromatic line-like signal through the characterization of the quality of the energy measurement on a event-by-event basis.

1.2.4 Cosmic-ray studies

Fermi has contributed to the study of Cosmic Rays (CRs) in several different ways [Tho12]:

- Measuring the Galactic diffuse emission, which imposes constraints on the production and diffusion of CRs in the Milky Way.
- Directly observing SNRs, one of the most accredited candidate sources of the galactic cosmic ray.
- Providing direct observation of high-energy cosmic-ray electrons (CRE) and positrons and measuring their spectra [Ack10]. More details about the use of the LAT as a CRE detector will be given in Chapter 5.

[1] From a practical point of view we can consider monochromatic each signal whose intrinsic spectrum is much narrower than the instrumental resolution.

Even if the instrument is not equipped with an on-board magnet, separate cosmic-ray electron and positron spectra have been measured [Ack12d] exploiting Earth's shadow, which is offset in opposite directions for opposite charges due to Earth's magnetic field.

1.2.5 The LAT event-level analysis

Together with its effort in carrying on the described scientific analysis, the *Fermi* collaboration is actively and continuously working to improve the reconstruction and event selection stage of the experiment, in order to reach the full realization of the Fermi-LAT scientific potential.

The current LAT event-level analysis was largely developed before launch in a series of iterations called Passes. After the launch, the accumulated updates have been applied in periodical releases of data, corresponding to reprocessing the entirety of LAT data to make use of the latest available analysis.

Pass 6, the iteration released at launch, was based exclusively on informed estimates of the cosmic-ray environment at the orbit of *Fermi* and on an evaluation of the LAT performance through Monte Carlo (MC) simulations.

Pass 6 was followed in August 2011 by Pass 7 [Ack12a], an improved version of his predecessor, for which parts of the data reduction process were updated to account for some previously neglected on-orbit effects, primarily the effect of chance coincidences with cosmic rays (often referred to as *Ghost events*, see section 2.6), by making use of the large number of real events the LAT collected in the first 2 years of operation. The event reconstruction and the overall analysis design were not modified, but the event classification was re-optimized on simulated datasets including all known on-orbit effects.

The LAT collaboration is now working on a radical revision of the entire event-level analysis, that is expected to significantly improve the instrument performance and greatly extend the LAT science capabilities. Clear improvements have been identified in all the main areas, including the Monte Carlo simulation of the detector, the event reconstruction and the

background rejection. This new iteration of the event-level analysis framework will go under the name of Pass 8.

The first step of this process, the new event reconstruction, is complete and frozen. All mission data have been reprocessed with the new reconstruction. The current focus is on the development and validation of event classes, optimized for different scientific cases. The present work fits into this stage, in particular the characterization of the topological information available on an event-by-event basis and its use for the background rejection. The connections with some relevant science topics and the corresponding prospectives are described in detail in this thesis.

1.3 Overview of the work

In Chapter 2 I will describe briefly the LAT and all its subsystems, including the trigger mechanism and some of the on-board operating software which will be relevant for the prosecution of the work. I will also provide useful information about the LAT orbital environment and the Monte Carlo simulations used by the *Fermi* collaboration both for studying the instrument performance as for estimating the background contamination.

In Chapter 3 I will give an overview of the new Pass 8 event-by-event reconstruction and of the standard selection algorithms developed by the *Fermi* collaboration in order to select genuine celestial γ-ray over the charged particle background. A report on the instrument performance, stressing the difference with Pass 7 results, will be given in this chapter as well.

In Chapter 4 I will present the results of a multivariate analysis, made using the variables produced by the various reconstruction algorithms, aimed at characterizing the quality of the energy reconstruction. This is achieved by an algorithm which, based on the topology of the event in the detector, tries to estimate the probability that its energy measurement will be in the core of the energy dispersion distribution.

The algorithm is created by training several Classification Trees [Bre84] on simulated data. Various possible approaches to the task are examined, as well as a series of different design choices, and their results are compared

to determine the one with the best performances.

The resultant estimator will become part of the mentioned standard selection algorithms as a variable onto which operate quality cuts finalized to remove the tails of the energy dispersion. Furthermore, it can be used to select event sample with a narrower energy dispersion, which can potentially enhance the sensitivity in several different science analyses.

A natural application of this first stage of the work is the search of possible monochromatic lines in the γ-ray spectrum. Through a series of Monte Carlo simulations I show that, taking into account the information given by the quality of energy estimator, the sensitivity to a line can improve up to 15% in interesting regions of the phase space.

Finally, in Chapter 5 I will present my work on various stages of an ongoing analysis of the high-energy cosmic ray electrons measured by the LAT with the new Pass 8 event level analysis. My main contribution was the development of a custom event selection, made of both hand-made cuts and specifically trained Classification Trees, aimed at discriminating electrons from the other charged species. This event selection will be characterized in terms of acceptance of the detector and of residual background after all the cuts. A preliminary electron spectrum is included, as final conclusion of this work.

Chapter 2

The Large Area Telescope

The LAT is a pair-conversion telescope for high-energy γ rays. It is composed by three main subsystems: a converter/tracker (TKR) for promoting conversions in e^+e^- pairs and measuring the direction of incident photons; a calorimeter (CAL), which provides energy measurement; and a segmented anticoincidence detector (ACD) for charged cosmic ray (CR) background rejection. Both the CAL and the TKR are divided in 16 modules (also referred to as *towers*) arranged in a 4×4 scheme and supported by an aluminum grid structure. A programmable trigger and data acquisition system (DAQ) is responsible for the selection and recording of the candidate events before their transmission to the ground.

A schematic of the LAT, reporting also the coordinate system which will be used in the following, is shown in Figure 2.1. Figure 2.2 shows a cutaway of the LAT including all its components.

The typical signal event registered by the LAT consists of a γ ray converting in a e^+e^- pair while traversing the TKR, with one or both the generated charged particles leaving a measurable track into it, before producing an electromagnetic shower in the CAL. There are, however, many other possiblities, involving events with radically different topologies as well as background events from different species of charged CR. This great variety, together with the wide phase space over which the LAT operates (in terms of both energy and incidence angle of the incoming particles), makes the event reconstruction a challenging task.

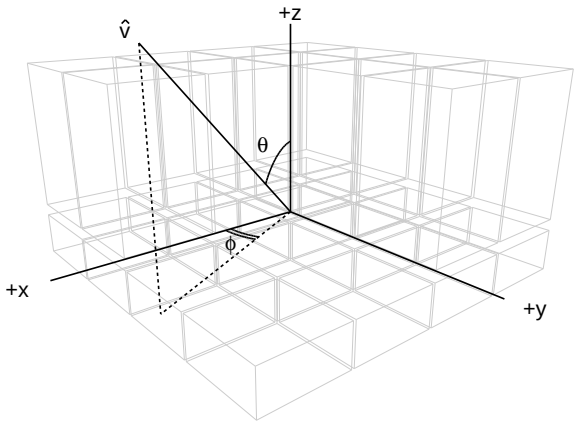


Figure 2.1. LAT scheme showing the CAL and TKR modular scheme (in trasparency) and the standard coordinates sytem.

It is crucial, for the purpose of the present study, to give a brief description of the subsystems listed above and of how they are designed to achieve the multiple goals of registering the highest possible number of events, measuring their energy and direction with a high accuracy and keeping the background contamination at an acceptable rate. The trigger mechanism and some other relevant topics concerning the LAT design and their influence on the instrument performance will be discussed in this chapter as well.

2.1 The silicon tracker

The tracker [Atw07] is responsible for promoting the conversion of photons into e^+e^- pairs, measuring their incident direction and providing the primary trigger for the instrument. It also contributes to the energy measurement at low energy (below a few hundred MeV).

Each of the 16 modules (*towers*) of the TKR is 37.3 cm wide and 66 cm tall and is composed by 18 x–y tracking planes, consisting of two layers of single-sided silicon strip detectors (SSDs). Subsequent layers are mutually orthogonal, ensuring measurement of both x and y coordinates. SSDs have 384 parallel strips spaced at 228 μm pitch over an area of 9.85×9.85 cm^2, and a thickness of 400 μm. Sets of 4 SSDs are bonded to form "ladders".

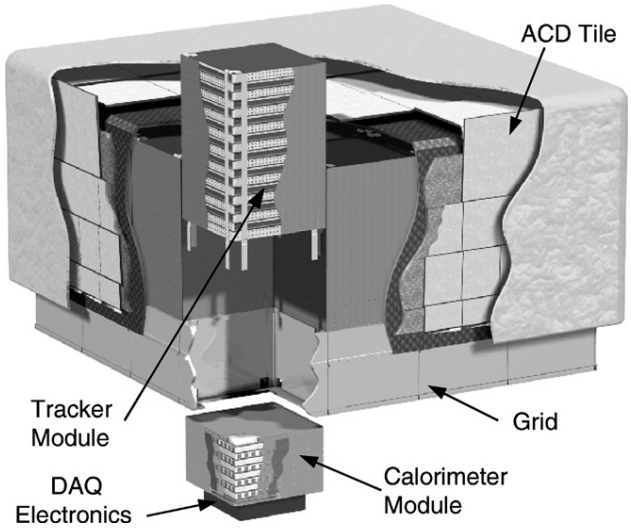


Figure 2.2. Schematic diagram of the Large Area Telescope. The telescope' s dimensions are $1.8 \times 1.8 \times 0.72$ m^3.

There are 4 ladders in each plane, for a total of 16 SSDs per plane (576 per tower).

In order to enhance the chance of a γ ray converting in a pair, the first 16 x–y layers starting from the top are preceded by a converter foil of high-Z material (tungsten). The last two layers have no converter, because the trigger of the TKR, requiring hits in 3 adjacent x–y layers, would be insensitive to photons converting here (see section 2.5). The upper twelve planes of tungsten are each 2.7% of a radiation length (RL) in thickness (0.095 mm), while the final four are each 18% RL (0.72 mm).

This choice is a compromise between two somewhat conflicting needs: reducing the effect of multiple scattering on the angular resolution, especially at low energy, requires minimizing the material traversed by the photon from the conversion point to the first tracking layer; maximizing acceptance, in particular for the relatively rare high-energy photons, which are less affected by multiple scattering, requires increasing it. Thus, the TKR can be thought as effectively being divided into two different instruments: the 12 thin layers form the so called "front" section, while the rest of the layers form the "back" section. Overall approximately 63% of photons above 1 GeV are converted at normal incidence.

Table 2.1. Summary of tracker characteristics and performance metrics.

Metric / Characteristic	Measurement
Active area [1]	1.96 m^2
Gamma-ray conversion probability [1]	63%
Active area fraction within a Tracker module [1]	95.5%
Overall Tracker active area fraction [1]	89.4%
SSD strip spacing	0.228 mm
Power consumption per channel	180 μW
Tower-module mass	33.0 kg
Single-plane hit efficiency in active area	$> 99.4\%$
Dead channel fraction	0.2%
Noisy channel fraction	0.06%
Noise occupancy	$< 5 \cdot 10^{-7}$

[1]At normal incidence

The TKR readout is binary, with a single threshold discriminator for each channel, and no pulse height information is collected at the strip level. The logical OR of all the discriminated strip signals on the same detector plane[1] is used for trigger purpose and the system measures and records also the time-over-threshold (TOT) of this layer-OR signal, which provides charge deposition information that is useful for background rejection.

The individual electronic chains connected to each SSD strip consist of a charge-sensitive preamplifier followed by a simple CR-RC shaper with a peaking time of ~ 1.5 μs (which is the relevant time interval for trigger purposes). The discriminated output remains high for ~ 10 μs for Minimum Ionizing Particles (MIPs) at the nominal $\sim$ 1/4 MIP threshold setting. An important effect of this permanence will be discussed in section 2.7

Some of the relevant characteristics and performance metrics of the TKR are summarized in Table 2.1.

[1]More precisely the logical OR is taken separately for half of a plane. More details can be found in [Atw07].

2.2 The electromagnetic calorimeter

The calorimeter [Gro10] is the main instrument for measuring the energy of the incoming particles. It also provides an image of the shower development, which is used both for background rejection and, to some extent, for measuring the incident direction.

Each CAL module has 96 CsI(Tl) crystals, each $2.7 \times 2.0 \times 32.6$ cm^3 in size . The crystals are optically isolated and are arranged in 8 layers of 12 crystals, with a total vertical depth of 8.6 radiation lengths (for a total instrument depth of 10.1 radiation lengths at normal incidence, including the TKR). Crystals in a layer are arranged orthogonally to crystals in the neighbor layers, so that they form an hodoscopic array.

A single crystal is read out with four photo-diodes, two at each end: a large photo-diode, with 147 mm^2 area, covering the energy range from 2 MeV to 1.6 GeV, and a small photo-diode, with 25 mm^2 area, covering the energy range from 100 MeV to 70 GeV. Each photodiode is connected to a charge-sensitive preamplifier whose output drives a slow (~ 3.5 μs peaking time) shaping amplifier for spectroscopy and a fast shaping amplifier (~ 0.5 μs peaking time) for trigger purposes. A zero-suppression discriminator eliminates all single-crystal signals with an energy <2 MeV.

Each crystal effectively provides a position measurement in all the three spatial coordinates. Two of them derive directly from the position of the crystal in the hodoscope. The third is estimated from the asymmetry in light levels between the two ends of the crystal, from which is possible to determine the position of the energy deposition along the crystal with an accuracy which varies from a few millimeters up to a fraction of millimeter [Atw09] (see also [Abd09a] for more details about the on-orbit calibration of the light asymmetry measurement).

The imaging capability of the CAL is fundamental for reducing the loss in energy resolution due to shower leakage, which for events above a few GeV is the dominant source of inaccuracy. It also plays a key role in distinguish and reject hadronic showers and, for high energy events, can be used to help the reconstruction algorithm in determining the right direction of track.

2.3 The anticoincidence detector

The purpose of the ACD [Moi07] is the rejection of background from charged particles. The key characteristics of the ACD are its high efficiency, greater then 0.9997 on average for a MIP, and its segmented nature.

In fact, instead of being a monolithic object, the ACD is segmented into several tiles of plastic scintillator: 25 of these tiles cover the top of the LAT, while other 64 cover the side faces (16 each). The reason behind this design choice was reducing the loss of effective area due to the so called "back-splash effect": isotropically distributed secondary particles coming from electromagnetic showers in the CAL (mostly 100–1000 KeV photons) which produce false vetoes in the ACD via Compton scattering. Segmentation allows to consider only ACD tiles nearby the photon candidate, effectively reducing the possibility of a *self-veto*.

The scintillation light from each tile is collected by wavelength shifting fibers (WLS) that are embedded in the scintillator and are coupled to two photomultiplier tubes (PMTs) for redundancy.

In order to keep the highest possible hermeticity, scintillator tiles are overlapped in one spatial direction, while the gaps in the other direction are covered by flexible scintillating fibers (*ribbons*) read out by a PMT at each end. These gaps, where the efficiency is inferior, account for $< 1\%$ of the total area. In addition, there are also gaps in the corners of the sides of the ACD which are not covered by ribbons and must be accounted for during reconstruction and classification of events.

The tile threshold is at 0.45 MIP for onboard use in rejection of background particle (see section 2.5). A zero-suppression theresold at ~ 100 keV is used to sparsify the signal. The output of each PMT is connected to a fast shaping amplifier (with ~ 400 ns shaping time) for trigger purposes and two separate slow electronic chains (with ~ 4 μs shaping time and different gains) to measure the signal amplitude.

To minimize the chance of light leaks due to penetrations of the light-proof wrapping by micrometeoroids and space debris, the ACD is completely surrounded by a low-mass micrometeoroid shield (0.39 g cm^{-2}).

2.4 Orbital environment

The *Fermi* spacecraft follows an orbit at 565 km of altitude, with an inclination angle of 26.5° and a period of ~ 96 minutes. In standard sky-survey mode the spacecraft rocks north and south about the orbital plane on alternate orbits with a characteristic (adjustable) rocking angle. This rocking angle was set to 35° at the moment of the launch. On 2009 September 3 it was increased to 50°, in order to lower the temperature of the spacecraft batteries and thus extend their lifetime.

Fermi's survey observations are regularly interrupted during passages through the South Atlantic Anomaly, a region of high particle flux, where the instruments cease data-taking and are commanded into a protected state. In addition, sky-survey is occasionally interrupted by Autonomous Repoints of the observatory to observe gamma-ray bursts deemed to be "of-interest" by algorithms running in the instruments' flight software. Finally, on rare occasions, the planned survey mode may be halted to allow a Target of Opportunity (Too) observation of a particularly interesting astrophysical event.

The LAT is characterized by a wide Field of View (FoV, see section 3.4) with approximately one fifth of the sky (~ 2.4 sr) observed at each instant by the detector. When combined with the LAT observing strategy, the result is that a full scan of the sky is performed every ~ 3 hours.

The rate of celestial γ rays triggering the instrument is only a minimal fraction of the full rate of events registered by the LAT, with the most prominent component being constituted by CR protons (Figure 2.3). In fact, there are several species of background particles which needs to be rejected, in order to isolate the contribute of γ rays of genuine astrophysical origin:

1. *Primary charged CR*: protons, antiprotons, electrons, positrons, alpha particles and heavier ions. Due to geomagnetic cutoff, they become a relevant component of the background above a few GeV.

2. *Secondary charged particles and neutrons*: neutrons, protons, electrons and positrons produced by CR interactions in the atmosphere. Among

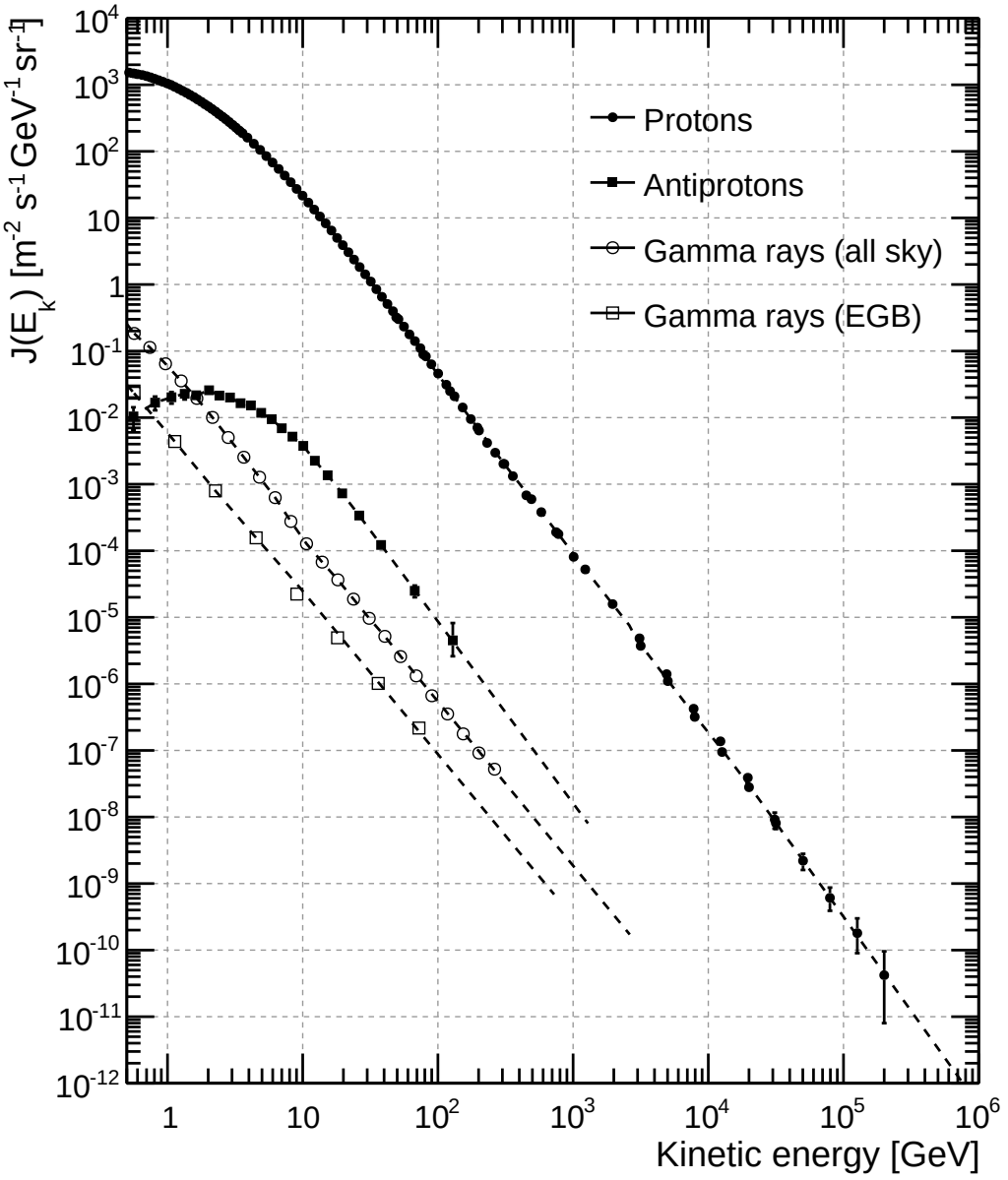


Figure 2.3. Comparison between the differential spectra of primary CR protons, primary CR antiprotons, astrophysical γ rays (integrated on all the sky) and the extragalactic diffuse emission (EGB).

them, neutrons are essentially negligible, being easily rejected by the selection algorithms.

3. *Secondary gamma*: the Earth is the brightest source in the sky due to its proximity. The γ-ray emission from the Earth is produced by cosmic-ray interactions with the Earth's atmosphere, and it is often called the γ-ray albedo. In particular, a bright limb at the Earth's horizon (Earth Limb) is produced by grazing-incidence CR showers coming directly towards the LAT, whose forward-moving γ rays can penetrate the thin atmospheric layer. Such photons typically present a much greater flux compared to those of astrophysical origin and can constitute a relevant source of background when the Earth Limb enters the LAT field of view.

   The standard strategy to exclude albedo γ rays from scientific analysis consists simply in performing a zenith cut on the data, based on the relative position of the Limb with respect to the LAT boresight, finalized at removing the contaminated regions of the LAT field of view. However, there are two different ways in which some of these photons can pass through the selection: when their direction is very poorly reconstructed, so that they no more result in the angular excluded region; and when, entering the LAT from behind, they are wrongly interpreted as coming from the opposite direction. Even if both possibilities are rather unlikely, the flux of these γ rays is so intense that they can represent a relevant source of background.

2.5 Trigger and Data Acquisition

As said, the flux of charged particle traversing the LAT is some orders of magnitude grater then that of celestial γ rays. In addition, the average downlink available to the LAT is limited to ~ 1 Mb/s (daily average). The trigger and data acquisition (DAQ) system is deputed to initiate the readout of the three subsystems of the LAT in presence of a candidate γ ray, while the *on-board filter* keeps the data sent to ground at an acceptable rate,

providing a first stage of background rejection. This process can be thought as being made of 4 steps:

1. *hardware trigger request*: at least one of the three subsystems of the LAT sends a trigger request via the Tower Electronic Modules (TEMs).

2. *hardware trigger accept*: the Global-Trigger Electronic Module (GEM) checks if the event in question generates an acceptable trigger pattern and starts the read out of all the instrument.

3. *on-board filter*: the event is passed to the on-board filter, that selects those with the highest probability of being a γ ray and transmits them to the ground.

4. *ground analysis*: a more refined analysis is performed off-line. Events are divided in *event classes* of increasing purity with respect to background contamination.

There are several condition that can cause a subsystem to send a trigger request to the GEM:

TKR (*three-in-a-row*): each time that a channel in the tracker is over threshold (nominally 0.25 MIP), a trigger request is sent to the corresponding TEM, which then checks if at least three consecutive x–y layers pairs in that tower have a signal over threshold and, if that is the case, send a trigger request to the GEM. Note that, even if a trigger condition is referred to a single tower, the readout of the LAT is always global.

CAL_LO : issued when the signal from any of the CAL crystals is above a fixed threshold (nominally 100 MeV)[2].

CAL_HI : issued when the signal from any of the CAL crystals is above a fixed threshold (nominally 1 GeV).

[2]In the nominal configuration for science data taking the CAL_LO condition is inhibited from opening a trigger window.

VETO : issued when any of the ACD channels is over threshold. Note that the GEM can logically group tiles and ribbons to form Regions Of Interest (ROIs). An ROI can be defined through a series of configuration registers as any combination of the ACD tiles and ribbons. The ROI signal is simply whether any one of the tiles that define the ROI is asserted. An ROI signal is assembled when a TKR primitive in a tower happens in concomitance with a VETO.

CNO : issued when the signal in any of the ACD tiles is above the CNO threshold (nominally 25 MIPs). This trigger is intended to signal the passage of a heavy nucleus, mainly for calibration purposes.

PERIODIC : a trigger which runs with a frequency of 2 Hz and is used for diagnostic and calibration purposes.

Trigger requests are collected by the Central Trigger Unit (CTU). All 256 possible combinations of the eight trigger requests listed above are mapped into so-called trigger engines (see Table 2.2). Such set of primitives is compared to a table of allowed trigger conditions and, in case a trigger condition is satisfied, a global trigger is issued and event acquisition starts.

Each engine in the table (a part from the periodic trigger) is associated with a specific physical event. For example, engine 4 typically signal the passage of a heavy ion. Engine 7 is the most relevant for γ rays, requiring a trigger signal in the TKR without associated ROI vetoes. Engine 9 is intended to recover those events for which the veto in the ACD may be caused by backsplash from the CAL. Engine 6 ensures that almost every event presenting a high energy deposition in the CAL is accepted.

Trigger engines are scalable: for each trigger condition a prescale is specified, corresponding to the number of valid trigger requests necessary to issue a single global trigger (obviously no prescale is applied to engines intended for γ-ray collection).

The minimum dead-time associated with the read out of the LAT is 26.5 μs, so one of the goals of the stage discussed above was to prevent the global fraction of dead-time from exceeding $\sim 10\%$. The residual rate, which is still of $\sim 2-4$ kHz, needs to be further reduced to ~ 400 Hz before

Table 2.2. Definition of standard trigger engines in terms of primitives used (1: required, 0: excluded, x: either). Engine 0,1 and 2 are not reported, as they are used for other purposes then data acquisition. Engine 8 is disabled in normal configuration for data-taking.

Engine	PERIODIC	CAL_HI	CAL_LO	TKR	ROI	CNO	Prescale
3	1	×	×	×	×	×	0
4	0	×	1	1	1	1	0
5	0	×	×	×	×	1	250
6	0	1	×	×	×	0	0
7	0	0	1	×	0	0	0
9	0	0	1	1	1	0	0
10	0	0	0	1	1	0	50

being downliked to the ground. This task is accomplished by the on-board filter.

More in detail, there are three different kind of filters: a HIP filter, designed to select heavy ions, a GAMMA filter, designed to accept γ rays and a DIAGNOSTIC filter, used to collect data from the PERIODIC trigger plus a small unbiased sample of all kinds of trigger. The most relevant for scientific purpose is the γ filter. It is composed by a sequence of tests (performed in hierarchical order) that an event must go through in order to get accepted and which are designed, for example, to exclude events not presenting at least one rudimentary track in the TKR, or whose track points to a hit ACD tile. The use of the ACD veto signal is disabled in case of an event for which the energy deposited in the CAL exceed a given threshold (adjustable, currently 20 GeV). A more detailed description of the γ filter can be found in [Ack12a]

2.6 Monte Carlo simulation

Both for the stage of event reconstruction and for that of background rejection, a heavy use was made of a detailed simulation of the LAT [Bal06], including information about the detector geometry and materials, as well as calibration data (thresholds, gains etc..). The simulation is written in

the framework of the Geant 4 Monte Carlo (MC) toolkit [Ago03] [All06] and is used to predict how particles interact in the instrument. It is also able to replicate the instrument response and the Trigger and On-board Filter behavior. Simulated data undergo the same event reconstruction and classification as real data.

The Monte Carlo simulation can be used, starting from astrophysical models, to reproduce fluxes from specific γ-ray sources as seen by the LAT. For the purpose of the present study, however, another MC product will be of particular importance: a simulation of a uniform γ-ray field (often referred to as the *allGamma* simulation) which is used to explore the response of the LAT across the entire range of energy and incident inclination angle. The characteristics of the *allGamma* simulation are:

- the incoming flux is made of pure γ rays.
- the input energy spectrum is a power-law with index -1. The distribution is uniform when logarithmically binned.
- γ rays are generated randomly on a sphere of 6 m^2 centered in the origin of the reference frame shown in Figure 2.1, so that the whole detector is enclosed in the simulation.
- the direction of each γ ray is randomly sampled from a isotropic distribution restricted to downward going photon. Consequently, the distribution of events is uniform when binned in $\cos(\theta)$ and ϕ, (with θ and ϕ being the angles introduced in Figure 2.1).

The same LAT model was used also to produce a simulation of the background from CR primaries and secondaries. This *Background* simulation was essential for developing the selection and classification algorithms and for estimating the residual background in the various event classes. A comparison of the various background fluxes, based on the same model used for the simulation (described in [Miz04]), is shown in Figure 2.4.

Since the intensity of the various kinds of background depends on the position on the LAT in the terrestrial orbit, in order to appropriately reproduce their distribution the simulation samples on uniformly spaced inter-

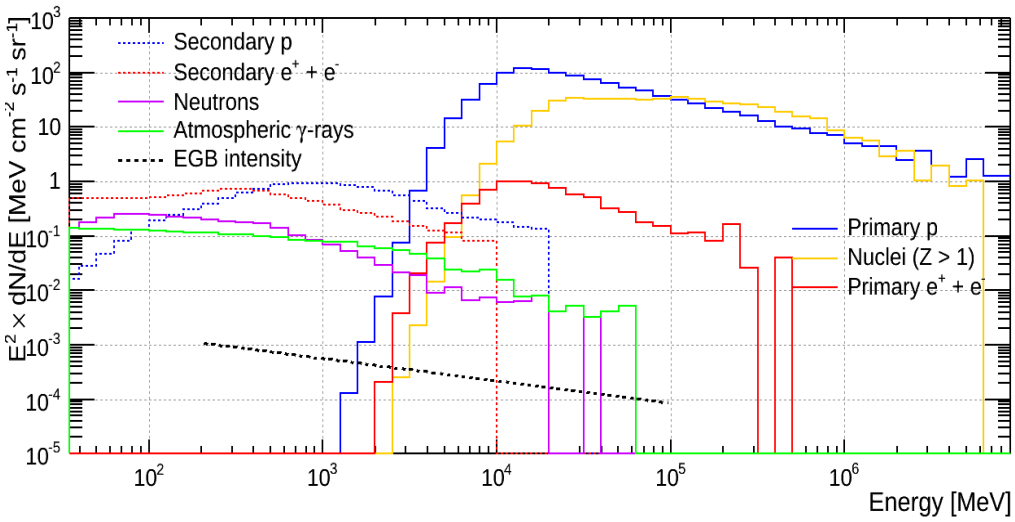


Figure 2.4. Background model, showing the fluxes of the various components

vals of times (such as 4 s every 4 min) across long periods of acquisition time (at least grater then the Fermi orbital full procession period, 53 days).

2.7 Ghost events and Overlays

A non negligible source of noise affecting LAT's measurements consists of remnants of electronic signals which persist in the various subsystems for a few μs after the passage of a particle[3]. If a particle traverses the LAT a few μs before or after the instrument has entered in read out mode, it is possible that the signal it leaves in the various electronic channels are erroneously read out together with the event data.

Such noise, usually referred to as "ghost events", can confuse the event reconstruction and lead to a degraded energy or direction measurement, or even to incorrectly reject a good γ-ray event as background. An example of ghost events is given in Figure 2.5. The entity of this effect varies according to the trigger rate (i.e. mostly the rate of charged CR).

In order to take into account ghost events in the MC simulations, we made use of the PERIODIC trigger (see section 2.5) which provides a sam-

[3]The persistence time can reach a few hundreds μs if the particle is a high-Z ion.

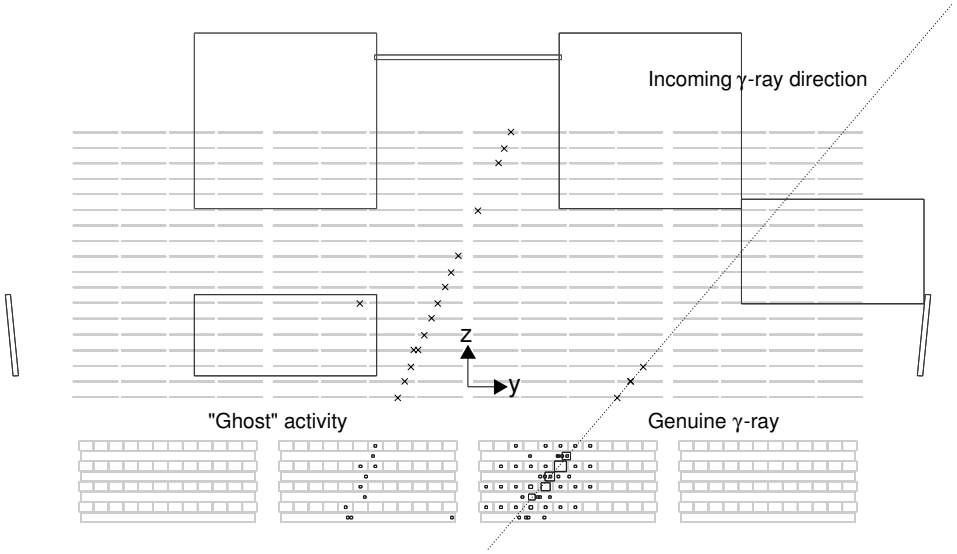


Figure 2.5. Example of a ghost event in the LAT (y-z orthogonal projection). In addition to an 8.5 GeV back-converting γ-ray candidate (on the right) there is additional activity in all the three LAT subsystems, with the remnants of a charged-particle track crossing the ACD, TKR and CAL. The small crosses represent the clusters (i.e., groups of adjacent hit strips) in the TKR, while the variable-size squares indicate the reconstructed location of the energy deposition for every hit crystal in the CAL (the side of the square being proportional to the magnitude of the energy release). The dashed line indicates the γ-ray direction. For graphical clarity, only the ACD volumes with a signal above the zero suppression level are displayed.

ple of such noise. The channel-by-channel signal from a randomly chosen periodic trigger event, conventionally called “overlay event” is merged into each simulated event to reproduce the effect of the ghost activity. Overlays are appropriately chosen from those registered in similar orbital conditions (in terms of charged CR rates) as the simulated event they are added to.

In Pass 8 the problem of ghost events has been addressed with a new global approach. The improved event reconstruction (see section 3.2) makes use of clustering information from the CAL to minimize the effect of such spurious signals during the event analysis.

Chapter 3

The Pass 8 event-level analysis

3.1 Introduction

Upon transmission to ground, raw data undergo a complex analysis from a series of specifically designed algorithms, deputed to attempt *event-by-event* a reconstruction of the event full development in the detector. In the process, a few hundreds of variables (*figures of merit*) are produced, providing a high-level description of the topology of each event in all the subsystems.

After that, a multivariate analysis is performed on these quantities to find the best possible estimates for the particle's energy and incoming direction and to produce a set of variables quantifying the quality of the reconstruction and estimating the probability that the particle is indeed a celestial γ ray.

As anticipated in section 1 this stage of the work has been improved significantly since the beginning of the mission. The initial scheme of analysis was developed before the launch of the LAT and was called Pass 6. A first upgrade was released in 2011 under the name of Pass 7, which incorporates all the information and the experience accumulated in the first years of mission. Pass 7 is the current publicly released framework and is used for most of the ongoing scientific analysis performed on LAT data.[1]

[1] In 2013 the Fermi collaboration has reprocessed all the data collected since the beginning of the mission using an updated set of calibration constants [Bre13]. Unless other-

Reconstruction in Pass 7 does not differ substantially from what was done in Pass 6, the main change being a small adjustment in how the energy of the photon is assigned. A much more radical change is in preparation for the next iteration of the event analysis scheme, Pass 8. In this chapter we will describe Pass 8 and its difference with Pass 7. The detail of the events reconstruction and the performance of the instrument at various stages of selection will be discussed here, as well.

3.2 Event reconstruction

Event reconstruction has been subjected to a great amount of improvement during the developing of Pass 8. The most relevant change from the current implementation is probably represented by the abandonment of the "single-particle paradigm"; in Pass 7, in fact, the reconstruction was always attempted under the hypothesis that the incoming particle was a single photon going downward (in the LAT reference frame).

A more refined approach, especially designed to reduce the effect of ghost events (see section 2.7) and to generally improve the instrument performance, has now been studied for all the three main LAT subsystems. Here we give a brief summary of how these new reconstruction algorithms work, stressing the major differences with Pass7.

3.2.1 Calorimeter

The starting point for the energy evaluation are the measured energy depositions in the crystals and their positions along the crystals themselves, which form a 3D array of energies and locations. The centroid of the energy deposition is determined and the principal axes of the shower are evaluated by means of a principal moment analysis, in which the inertia tensor (with energy in place of mass) is diagonalized. In Pass 7 the energy deposition was treated as a single quantity, with no attempt to identify contamination from ghost signals. In Pass 8, as said, the computation of shower centroid

wise specified, in this work we will always refer to this reprocessed dataset (often called "P7_REP") when referring to Pass 7.

and axis has been preceded by a clustering-stage, with the aim of isolating and removing such ghost activity. Figure 3.1 shows an exampe of the benefits of this approach on the event reconstruction.

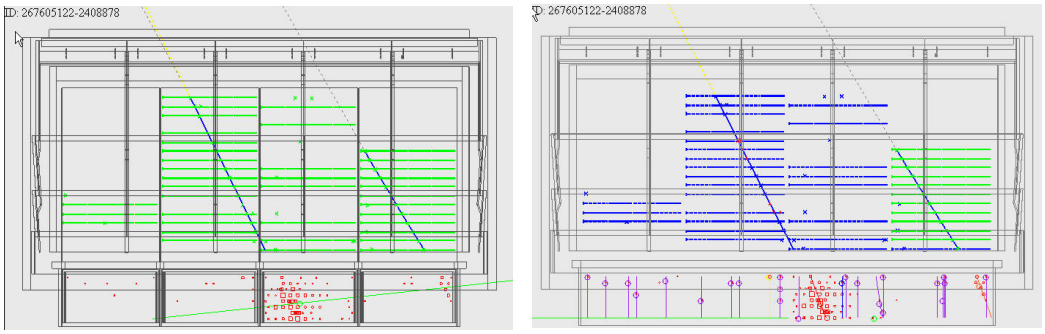


Figure 3.1. Effect of the CAL clustering stage on the event reconstruction. The picture shows the signal prodiced by a gamma ray crossing the detector in presence of a ghost particle (black dotted line). Current reconstruction is on the left panel: a single particle in the LAT is assumed. The CAL axis (green line on the bottom) is flipped by ghost activity far from the gamma ray. The effect of the clustering is on the right panel: the gamma ray is isolated and the event correctly reconstructed.

Two different algorithms are applied to estimate the actual energy of an event: a Parametric Correction (PC) and a fit of the Shower Profile (SP).

The PC algorithm was found to give better results at low energy (roughly below a few GeV). Initially, the overall energy is taken to be simply the sum of the crystal energies. Starting from this quantity, a series of corrections are made to account for the energy loss due to leakage out the sides and back of the CAL and through the internal gaps between CAL modules. The amount of energy deposited in the TKR is evaluated by treating the tungsten-silicon detector as a sampling calorimeter, where the number of hit silicon strips in a tracker layer provides the estimate of the energy deposition at that depth. This “tracker” energy is added to the corrected CAL energy and contributes with an important correction at low energies (reaching $\sim 50\%$ on average at 100 MeV).

For higher energies, where the fraction of energy released in the tracker is negligible and the energy loss is mostly due to the leakage of the shower out of the CAL, a full three-dimensional fit of the SP has proven to be more

effective. The fit requires a starting direction, which is provided by the TKR. This method has been partially modified, in the context of the Pass 8 development, in order to improve the handling of the two main factors degrading the energy resolution at high energy: saturated crystals in the CAL and poor shower containment. Thanks to that, the energy reach has been extended above the TeV scale. A complete description of the algorithm and its performance can be found in [Bru12].

A dedicated multivariate analysis is performed to decide, event by event, which of these two algorithms is more likely to produce the best estimate of the energy, which is then picked as the energy of the photon. In a small number of cases, the energy is assigned as a weighted average of the estimate of the two algorithms. This happens mostly in the transition region where their performance are comparable ($\sim$ a few GeV), in order to avoid introducing spectral feature with a sharp transition.

3.2.2 Tracker

The current Pass 7 tracker reconstruction code uses a track-by-track combinatoric pattern recognition algorithm to find and fit up to two tracks, representing the electron-positron pair. If more then one track is found, it combines them in the attempt to form a vertex, representing the photon conversion point.

The algorithm starts by considering nearby pairs of TKR cluster in adjacent layers as candidate tracks and proceed using a Kalman filtering technique ([Kal60],[Fru87]) to add additional TKR clusters to each of the candidate tracks.

A limitation of such approach is that this makes the efficiency and quality of the track-finding intrinsically dependent on the accuracy of the CAL reconstruction. In fact, the algorithm is based on the assumption that the energy centroid lies on the trajectory of the original γ ray, so it uses the reconstructed CAL energy centroid and axis to choose the initial hits. Furthermore, it makes use of the CAL information both for estimating the effect of multiple scattering (which is the most relevant factor degrading the direction measurement at low energy) and for disfavoring candidate tracks

for which the TKR and CAL estimated directions differ significantly (especially above 1 GeV).

In addition to that, the track-finding algorithm can be confused by a too high number of hits, in particular in presence of:

- multiple hits produced by electrons and positrons interacting readily in the material comprising the tracker (in particular the converter foils and silicon);
- backsplash particles which move upwards from the calorimeter, causing a large number of randomly hit strips in the lower planes of the tracker (particularly for high-energy events).

with the result of both a loss of events (whose reconstruction fails at all or which are mislabeled as background) and a degraded direction measurement.

The Pass 8 reconstruction addresses these issues by introducing a global approach to track-finding that tries to model the shower development with one or more tree-like structures. In this process tracker hits are linked together and the Tree structures are built by attaching links that share a common hit.

The head of the Tree represents the assumed gamma-ray conversion point. For each tree, the primary and secondary branches, defined as the two longest and straightest, represent the primary electron and positron trajectories (if unique) and sub-branches represent associated hits as the electron and positron radiate energy traversing the tracker. The Tree axis is evaluated and used to associate the tree to a particular cluster in the calorimeter, which allows an estimate of the energy associated with the tree.

Once an energy is available, up to two tracks are extracted from the hits along the primary and secondary branches and fitted with a Kalman Filter technique which accounts for multiple scattering. Tests with Monte Carlo simulations and flight data show that the new tracker pattern recognition significantly reduces the fraction of mis-tracked events.

At the end of the process, the primary track (when available) is used to

estimate the direction of the γ-ray. If a vertex has been found, the contribution of the two track is combined in producing the estimate.

3.2.3 ACD

In Pass 8 the ACD phase of the energy reconstruction has been fully rewritten. It starts, as before, by estimating the energy deposited in each of the tiles and ribbons. Subsequently, these energy depositions are associated to incident particle directions: here a major improvement occurred, since in Pass 7 only tracks derived from the TKR were used, while now directional information derived from calorimeter clusters is propagated as well. This additional CAL information is particularly important for identifying background events at high energies or large incident angles, which are more susceptible to tracking errors. In these cases, the CAL provides the more robust directional information.

For each track is calculated whether its projection intersects an ACD tile or ribbon with non-zero energy deposition; if not so, the distance of closest approach is computed between the track projection and the nearest such ACD element. Track-tile associations are used when considering whether the event should be identified as charged particle and rejected in later analysis stages. Previously, a simple energy scaling dependence was used to characterize the robustness of such associations. However, widely varying event topologies can lead to large differences in the quality of directional reconstruction for events of the same energy. In Pass 8, ACD reconstruction utilizes also event-by-event directional uncertainties to capture this information and to provide better background rejection.

The last major improvement comes from utilizing the fast ACD signals, provided to the LAT hardware trigger, to remove out-of-time signals from the ACD and mitigate the impact of ghost signals in the slower ACD pulse-height measurements.

3.3 Event classification

After the reconstruction stage, a series of algorithms, operating on the various figures of merit produced, are deputed to select those events which are most likely genuine γ rays of astrophysical origin.

Events are analyzed to determine the accuracy of the energy determinations, the directional accuracy, and whether they are γ rays. Such estimates try to capture the information from the single-event topology using a machine-learning approach based on Classification Trees (CTs) generated probabilities. The building of an estimator of the quality of the energy measurement is described in full details in chapter 4.

Since the initial contamination and the required data purity can vary significantly for different scientific analyses, a choice was made to divide events in classes, characterized in terms of residual background contamination. The definition of the classes, as well as the residual contamination in each of them, was studied with the help of MC simulations and validated through comparison with physical source of known properties.

In Pass 7 such classes are hierarchically arranged, so that each class is a subset of the previous one obtained with the addition of more stringent cuts. There are 4 standard classes:

- P7TRANSIENT. This is the one with the lower level of background rejection and the highest signal efficiency. It is used for studying transient phenomena, such as Gamma Ray Burst (GRB), for which the background rate is already suppressed by the small temporal window of integration.
- P7SOURCE. This is the standard event class recommended for point source analysis on long term integration.
- P7CLEAN. This class is thought especially for the study of the galactic diffuse emission.
- P7ULTRACLEAN. The purest class. It is used for the study of the weak extragalactic diffuse emission.

further details on the definition of these classes can be found in [Ack12a].

In Pass 8 the whole class structure has been revisited. Here we will describe briefly the new organization, addressing the main differences with Pass 7. Please note that this stage of the selection is still preliminary and possibly subjected to development from the *Fermi* collaboration. By design each event class is constructed from three components:

- Fiducial Cuts: Select events in an effective fiducial volume based on the γ-ray conversion point and the track projection through the TKR and CAL subsystems.
- Quality Cuts: Selections studied to remove events with poorly reconstructed energy or direction.
- Background-Rejection Cuts: Selections studied to remove CR-background and back-entering photons.

3.3.1 Fiducial Cuts

In Pass 7 a single fiducial cut is applied, common to all classes, selecting events with at least one reconstructed track, with the track's projection crossing at least 4 X_0 of material in the CAL and with at least 5 MeV of energy deposited in the CAL. Pass 8, on the other side, allows for a greater variability, including definition of classes without any information from the TKR or the CAL. In this sense the Fiducial Cuts effectively divide the event classes into three categories:

CalTkr : Events that convert in the TKR and have a track that projects into the CAL. In these events both the CAL and TKR subsystems can be used for event reconstruction and background discrimination. At least one track identified in the TKR, 4 X_0 of material traversed by the track's projection in the CAL and (at high energy) a minimal agreement between the CAL and TKR reconstructed direction are required.

CalOnly : Events that convert in the CAL or have a compromised track reconstruction. In these events the CAL is used for direction reconstruction.

TkrOnly : Events that convert in the TKR, but deposit no energy in the CAL. At least one track identified in the TKR is required.

CalTkr is the standard category of events, designed to be used for the vast majority of the scientific analysis, in continuity with the above described Pass 7 classes. The other two categories have been recently developed in an attempt to expand the scientific capability of the LAT.

In particular, TkrOnly events can become relevant at low energies ($<$ 100 MeV) where many events range out in TKR without reaching the CAL, and where the information from the TKR can be sufficient to provide a useful measurement of the particle energy. Recovering a fraction of these events, now discarded by the standard event selection, can be especially useful for the study of transient phenomena.

CalOnly events, instead, are relevant at high energies ($>$ 50 GeV), where a significant fraction of the particles have no usable tracker information (either because they convert in the calorimeter or due to mis-tracking), but the disabling of the on-board GAMMA filter (see section 2.5) becomes fully efficient. Even in the absence of tracker, at these energies the CAL provides a directional capability at the level of a few degrees or better, so CalOnly events constitute a very promising event category for those analyses where the pointing accuracy is not critical.

However the rejection of particle backgrounds in the absence of usable tracker information must still be studied in detail and at the actual stage of the work no event classes have been defined which falls into this category.

3.3.2 Quality Cuts

The quality selection is used to remove events with poorly reconstructed direction or energy. It is based on the value of quality estimators of the energy and direction measurement, which are produced at the end of the reconstruction stage. At the current stage the quality cuts are identical for all the event classes and are designed to be quite conservative, removing only events with a high chance of being mis-reconstructed.

3.3.3 Background Rejection Cuts

The background rejection cuts are designed to remove events with a high probability of being cosmic rays or back-entering gamma rays. For the rejection of these two background sources Pass 8 uses two different CT parameters to define energy-dependent selections for the various classes. These selections vary, due to the different requirements on the residual background-rate designed for each class.

A detailed description of all the event classes in Pass 8 is beyond the purpose of this study. We limit to say that the CalTkr category contains the rough equivalent of the Pass 7 SOURCE, CLEAN and ULTRACLEAN classes, while the TRANSIENT class, now existing both for the CalTkr and TkrOnly categories, has been split in multiple subclasses, hierarchically sorted by residual contamination. A brief description of the most relevant performance relative to the the new classes will be presented in section 3.4.2, 3.4.3 and 3.4.4

3.4 Detector Performance

3.4.1 Instrument Response Functions

The *instrument response functions* (IRFs) are specific parameterization of the instrument performance allowing to convert the raw counts registered by the detector into physically meaningful quantities such as fluxes and spectral indices. In general, the IRFs are not intrinsic characteristics of the detector: they always subtend a specific event selection and a detector may very well have different response functions in the context of different analyses.

In this section we provide a discussion of the LAT performance, often referring for illustrative purpose to the P8_SOURCE class, stressing the emphasis on how the IRFs tie to the detector characteristics and the basic interaction processes of particles and radiation into it. First we introduce a few preliminary definitions:

1. *Effective Area*, $A_{eff}(E, \hat{v}, s)$, is the product of the cross-sectional ge-

ometrical collection area, γ ray conversion probability, and the efficiency of a given event selection (denoted by s) for a γ ray with energy E and direction $\hat{v}$ in the LAT frame;

2. *Point Spread Function* (PSF), $P(\hat{v}'; E, \hat{v}, s)$, is the probability density to reconstruct an incidence direction v' for a γ ray with $(E, \hat{v})$ in the event selection s;

3. *Energy Dispersion*, $D(E'; E, \hat{v}, s)$, is the probability density to measure an event energy E'' for a γ ray with $(E, \hat{v})$ in the event selection s.

Although theoretically the energy measurement depends on the measurement of the direction and vice versa, in practice it is usually assumed that one can factorize the instrument response into those three pieces. This is possible, since the correlation between the two measures is small (for a more detailed discussion of this correlation see [Ack12a]). Under this hypothesis one can write, for a given distribution $S(E, \hat{p})$ of γ rays:

$$M(E', p', \hat{s}) = \int\int\int S(E, \hat{p}) A_{eff}(E, \hat{v}(t, \hat{p}), s) P(\hat{v}'; E, \hat{v}(t, \hat{p}), s) \\ D(E'; E, \hat{v}(t, \hat{p}), s) dE d\Omega dt \tag{3.1}$$

where $M(E', p', \hat{s})$ is the measured distribution.

Other useful quantities are:

- *Acceptance*, $\mathcal{A}(E)$, the integral of the A_{eff} over the solid angle.

$$\mathcal{A}(E) = \int A_{eff}(E, \theta, \phi) d\Omega \tag{3.2}$$

- *Field Of View*, (FoV), the ratio between the acceptance and the on-axis ($\theta = 0$) effective area at given energy:

$$\text{FoV}(E) = \frac{\mathcal{A}(E}{A_{eff}(E, \theta = 0)} \tag{3.3}$$

3.4.2 Effective Area

An accurate knowledge of the effective collecting area of the LAT is of great importance in order to correctly evaluate the spectra of astrophysical γ-ray sources. This is a task of high complexity, since A_{eff} depends on the geometrical cross section of the LAT as well as the efficiency for converting and correctly identifying incident γ rays. The *allGamma* simulation (see section 2.6) is the main tool for evaluating A_{eff}, which is then corrected, if needed, based on comparison with flight data.

Since in the smulation γ rays are generated uniformly in $\log(E)$ and solid angle, the effective area in any of the bins in which the parameter space is partitioned can be expressed in terms of the total number of generated events N_{gen} and the number of events $n_{i,j,k}$ passing the γ-ray selection within the specific bin centered at $E = E_i$, $\theta = \theta_j$ and $\phi = \phi_k$:

$$A_{eff}(E_i, \theta_j, \phi_k) = (6\mathrm{m}^2) \left(\frac{n_{i,j,k}}{N_{gen}}\right) \left(\frac{2\pi}{\Delta\Omega_{j,k}}\right) \left(\frac{\log_{10} E_{max} - \log_{10} E_{min}}{\log_{10} E_{max,i} - \log_{10} E_{min,i}}\right) \tag{3.4}$$

where $\Delta\Omega_{j,k}$ is the solid angle subtended by the bin j, k in θ and ϕ, E_{min} and E_{max} give the energy range of the allGamma sample and $E_{min,i}$ and $E_{max,i}$ are the boundaries of the i^{th} energy bin. In practice the ϕ dependence of the effective area is small and it is usually averaged in standard scientific analyses, so Eqn:3.4 can be rewritten as:

$$\begin{aligned} A_{eff}(E_i, \theta_j, \phi_k) =& (6\mathrm{m}^2) \left(\frac{n_{i,j}}{N_{gen}}\right) \left(\frac{2\pi}{\Delta\Omega_{j}}\right) \left(\frac{\log_{10} E_{max} - \log_{10} E_{min}}{\log_{10} E_{max,i} - \log_{10} E_{min,i}}\right) \\ & \times R(E_i, \theta_j, \phi_k) \end{aligned} \tag{3.5}$$

where $R(E_i, \theta_j, \phi_k)$ is a small factor (typically <10%) whose average is 1 by definition. The effect of ignoring the ϕ dependence is completely negligible for standard long-terms ($\sim$ years or months) source analysis, with some correction required only for short transient phenomena.

Figure 3.2 shows the dependence of the effective area from energy and incidence angle for the P7SOURCE class. At very low energy a combination

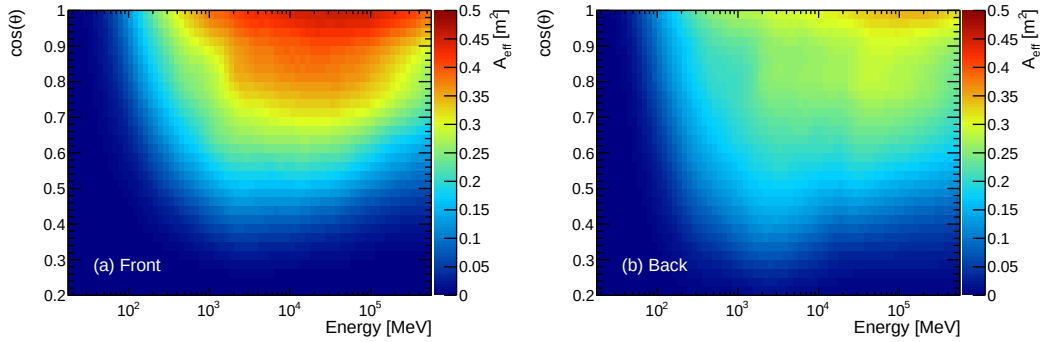


Figure 3.2. Effective area table for P7SOURCE class event, front and back section.

of factors makes the effective area go quickly to zero:

- the cross-section for pair-production decreases;
- most of the electrons and positrons produced range out in the TKR;
- the effects of multiple scattering may prevents from successfully reconstruct a track;
- an increasing fraction of photons, instead of converting into a pair, interacts via Compton scattering, which the reconstruction is not optimized to deal with;
- scarce information from the CAL makes hard to distinguish γ rays from background particles, causing misidentification;
- since the quality of the reconstruction is generally lower, ghost activity is more likely to not being correctly identified and removed.

At high energy, instead, the small decrease in effective area is mostly due to mis-tracking caused by the high number of backspalsh hit in the TKR.

Figure 3.3 shows a comparison between the acceptance of the SOURCE class event in Pass 8 and Pass 7. The recover of previously lost photons, thanks to the improved capability of removing ghost events, is clearly evident at low energy (especially below ~ 100 MeV). This recover of acceptance is one the most remarkable results of Pass 8 and it has already proved

of being effectively capable to enhance the LAT capability to measure transient phenomena like GRB [Atw13].

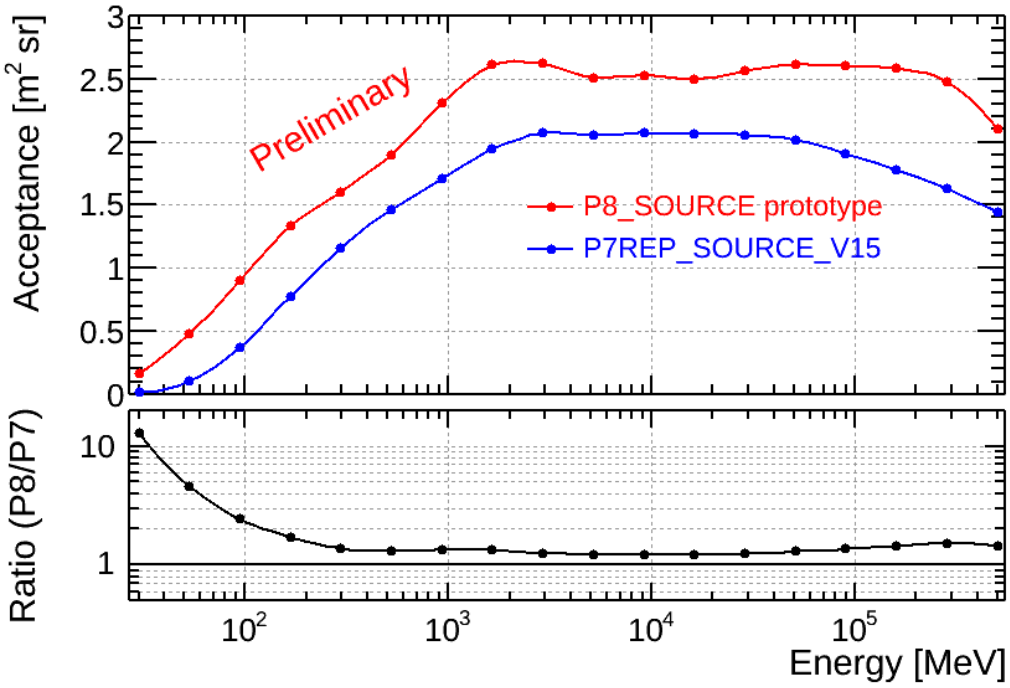


Figure 3.3. Acceptance as a function of energy for P7 and P8 SOURCE classes.

The θ dependence of the effective area can be visualized in Figure 3.4. As the incidence angle increases, the fraction of particles the LAT is capable of register progressively decreases for a combination of merely geometrical factors (for example because they miss the CAL or the TKR) and the elevate difficulty of correctly reconstruct such events.

3.4.3 Point Spread Function

The PSF is a fundamental measure of the LAT capability to correctly associate incoming photons with their direction in the sky (and consequently to their source of provenience). A standard way to summarize the angular resolution in a certain region of the phase space is to compute the angular width in the space containing a certain fraction (typically 68% or 95%) of the events. This kind of information is easily extracted from the PSF.

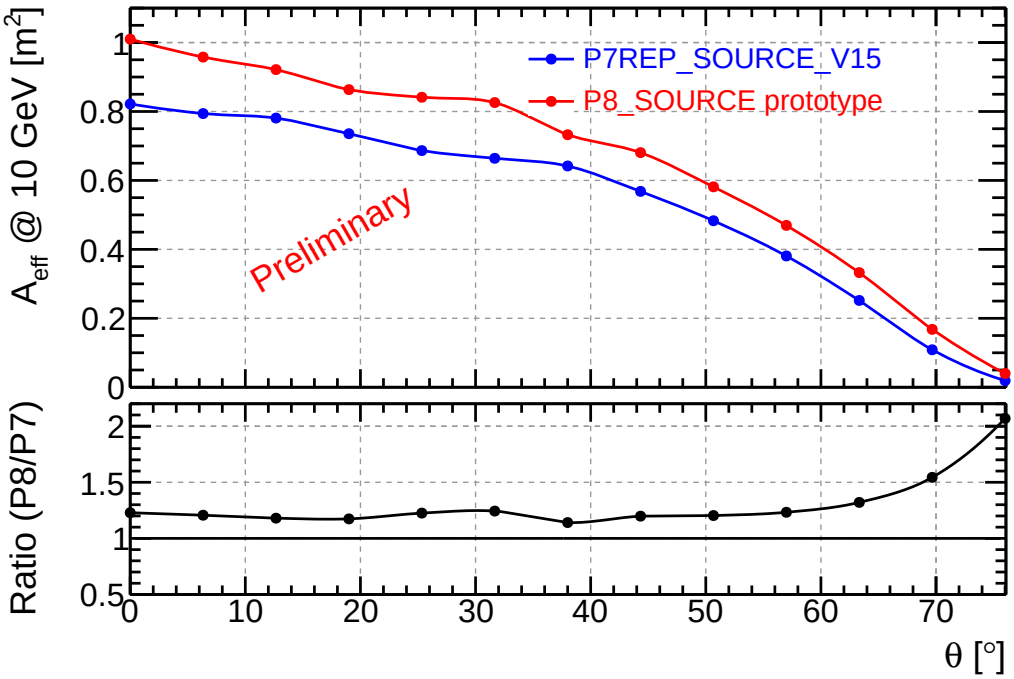


Figure 3.4. Effective area at 10 GeV as a function of θ in Pass 7 and Pass 8 source classes.

As for the Effective Area, the *allGamma* simulation is used to study the dependence of the 68% containment radius from energy and incident zenith angle. Again, the dependence from ϕ has been found to be much weaker then that from θ, so it is usually neglected in this analysis.

As anticipated in section 3.2, the most relevant factor in determining the PSF at low-energy is the multiple scattering of the primary electron-positron pair in the high-Z foils of converter, which scales with the energy as $\sim E^{-1}$. At high energy, on the other hand, the angular resolution is mostly limited by the ratio between the strip pitch (228 μm) and the lever arm between the point of conversion and the first hit in a layer (in the TKR). The transition between the two regimes occurs near a few GeV. It has been found that most of the energy dependence can be parametrized as:

$$S_P(E) = \sqrt{\left[c_0 \cdot \left(\frac{E}{100\text{MeV}}\right)^{-\beta}\right]^2 + {c_1}^2} \tag{3.6}$$

with $\beta \sim 0.8$ and the other parameters varying across the LAT phase space. Figure 3.5 shows the 68% containment radius as a function of energy for the Pass 7 and Pass 8 SOURCE class. The PSF generally benefits of the improved reconstructed algorithms, especially at high energy. At very low energy the gain is mitigated by the huge increase in acceptance.

The PSF computed from the *allGamma* simulation has been validated, and corrected when needed, through the analysis of pointform sources in the sky whose localization is known with great precision.

3.4.4 Energy Dispersion

Equivalently to what we did in section 3.4 we define the energy dispersion as being the distribution of the quantity:

$$\frac{\delta E}{E_{true}} = \frac{E_{meas} - E_{true}}{E_{true}} \tag{3.7}$$

where E_{true} is the true energy of an event and E_{meas} the energy measured by the instrument for that event. This distribution is not, generally speak-

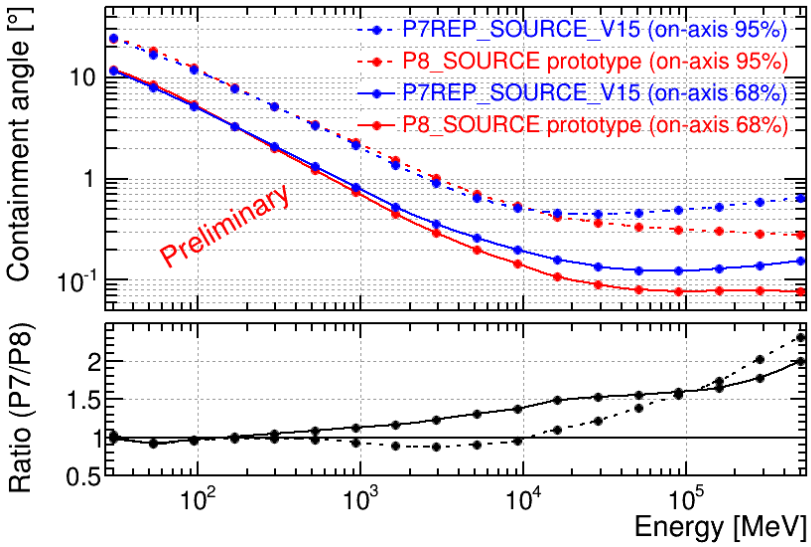


Figure 3.5. PSF 68% and 95% containment radius as a function of energy for the Pass 7 and Pass 8 SOURCE class as a function of energy.

ing, Gaussian, or even symmetric; however, it generally features at least one definite peak in proximity of zero (or, in other words, the energy estimate is not dramatically biased), in most of the LAT operating phase space. As a consequence, it is still meaningful to define the Energy Resolution as the width of the minimum 68% containment interval of the energy dispersion.

We will discuss the main factors influencing the quality of the energy measurement in Chapter 4. Here we limit to compare the performance of the Pass 8 and Pass 7 SOURCE classes (Figure 3.6). Overall the performance are similar. A minor gain is visible at high energy, due to the new reconstruction algorithm. As said for the PSF (see section 3.4.3), the loss at low energy is mostly a consequence of the increase in effective area.

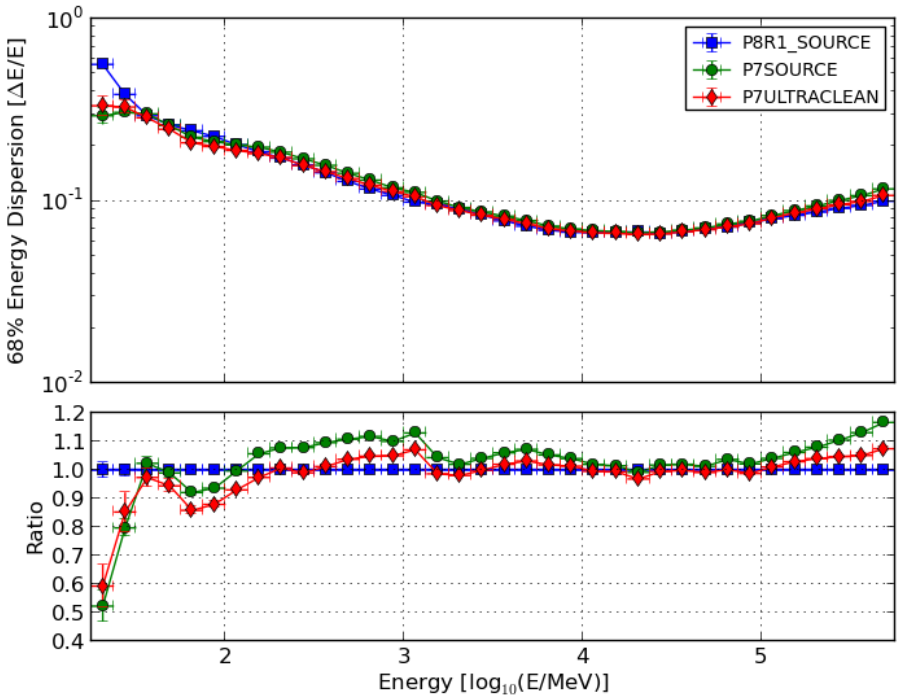


Figure 3.6. Energy Dispersion 68% containment window for the Pass 7 and Pass 8 SOURCE classes as a function of energy. P7ULTRACLEAN is shown for comparison.

Chapter 4

Estimating the quality of the energy measurement

4.1 Introduction

The accuracy of the energy measurement for a single event depends heavily on its topology in the detector. So it is possible, at least in principle, to make predictions about the quality of such measurement based on the high-level description of the event produced by the reconstruction algorithms. One way to accomplish this task is by creating an estimator of the probability that the energy measured for an event is "in the core of the energy dispersion", that is in the smallest 68% containment window of the Energy Dispersion, as defined in section 3.4.4.

We chose to compute the width of this window, which from now on will be indicated with Δ_{68}, in several bins of energy and zenith inclination angle, in an attempt to increase the consistence of the performance of the resulting estimator across the LAT phase space. This design choice will be discussed more in detail in section 4.2.

Making use of this definition, it is possible to perform a multivariate analysis in order to divide the "good" or "signal" events (those in the core of the distribution) form the "bad" or "background" ones (all the other events). The tool chosen for the work is a CT, trained on simulated data. Full technical details on the training of the CT, including the software used,

all the options selected, a comparison between various possible approaches tested and an evaluation of the results, are presented in the rest of this chapter.

Note that, in this particular case, the meaning of "good" and "bad" is intrinsically different from the typical case, where one is trying to distinguish between a signal and a background coming from two different populations. Here all events considered are γ rays and the splitting of the sample in signal and background is largely arbitrary. In fact, another possible definition of signal, as the content of a 95% window, has been tested (see 4.3), but proved to produce a less useful estimator.

4.2 Data preparation

The base for the study is the *allGamma* simulation presented in section 2.6. The standard *allGamma* covers a range of energy from $\sim$ 17 MeV to $\sim$ 560 GeV. Since Pass 8 has extended the high-energy reach of the LAT, we made use also of a dedicated simulation of high-energy photons, analogous to the *allGamma* but covering an energy range up to 10 TeV[1].

We required our sample to satisfy certain minimum quality cuts:

1. At least one successfully reconstructed track in the tracker.

2. At least 4 RL of intersection between this track and the CAL.

3. A reconstructed energy greater then 10 MeV.

Overall the simulated sample consists of more then $5 \cdot 10^5$ events. Tests have shown that a greater number of events does not improve the performance of the resulting estimator.

The first step of the analysis consists in splitting the sample into the signal and background classes. We used the energy of the simulated event (E_{mc}) as the E_{true} of Eqn. 3.7 for the purpose of calculating the energy dispersion.

[1]Note that, while the simulation extends up to 10 TeV, the real perfomance of the LAT does not allow for any sensible measurement beyond 1 TeV

As a design choice, we require our estimator to classify events according to how well their energy is measured not in an absolute sense, but relatively to the region of the phase space they belong to. The reason is that the accuracy of the energy measurement can vary a lot across the LAT phase space and, as a consequence, if Δ_{68} were computed only once for the whole data set, the selection of the signal class would have presented a strong bias towards those regions where the energy dispersion is intrinsically narrower, causing a heavy dependence of the classifier performance from the event coordinates.

In order to reduce this dependence we chose to divide the sample in bins of simulated energy and incident angle; more precisely, we made 30 bins of $\log(E_{mc})$, uniformly spaced from $\sim$ 17 MeV to $\sim$ 560 GeV, and 5 additional bins of energy up to ~ 7.5 TeV. The bins in the high-energy range are slightly larger to compensate for lower statistics in the Monte Carlo simulation. For the incident angle, we made 8 equally spaced bins in $\cos(\theta)$, from 0.2 to 1.0. Overall, the sample was dived in 280 bins, covering the full operative phase space of the LAT.

In each bin we computed Δ_{68} as defined in Eqn. 4.1. Note that, if the energy dispersion were a Gaussian, Δ_{68} would be exactly equal to 2σ.

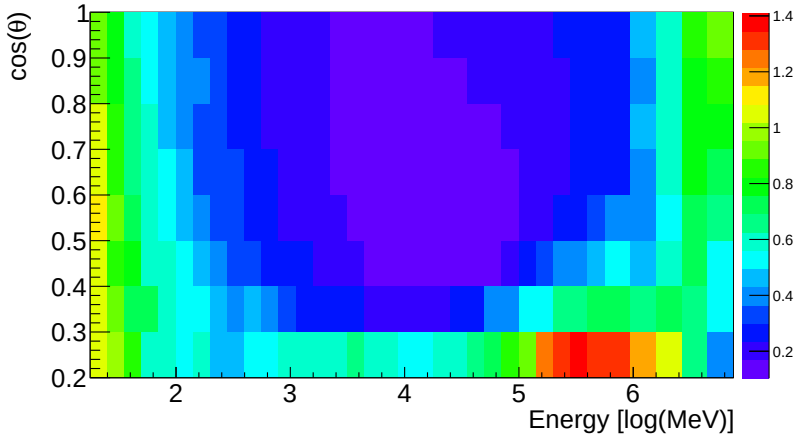


Figure 4.1. Δ_{68} computed in bins of energy and incidence angle for a filtered *allGamma* distribution (see text).

Then we use this 2-dimensional table (Figure 4.1) to associate to each

event a *Scaled Energy Dispersion* S_E defined as:

$$S_E = \frac{E_{meas} - E_{mc}}{E_{mc}} \cdot \frac{2}{\Delta^i_{68}} \tag{4.1}$$

where Δ^i_{68} is obtained for each event with a bilinear interpolation from the above mentioned table. With the factor 2 at the numerator we are considering half the width of the 68 % containment window, so that S_E can be seen as the deviation of the measured energy from the true value expressed in units of a (local) analog of the standard deviation. The distribution of S_E for the sample used is shown in fig. 4.2

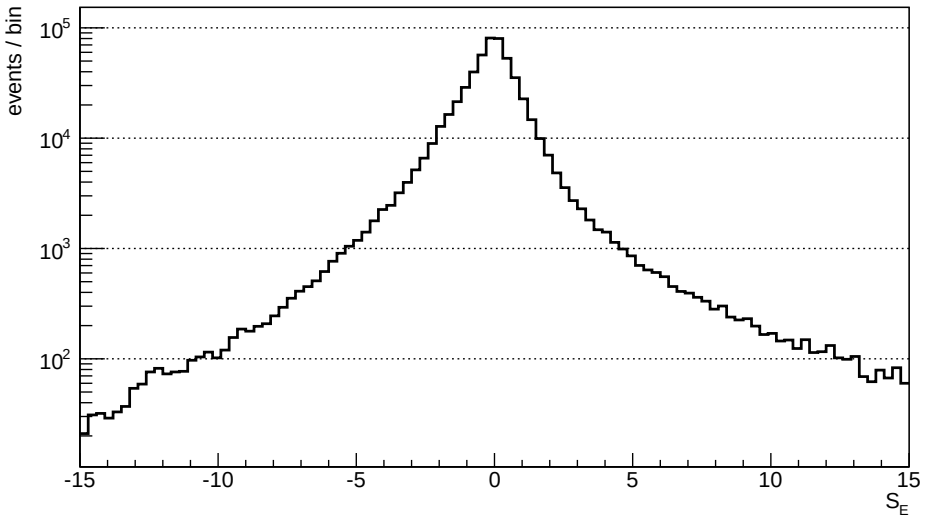


Figure 4.2. Scaled Energy Dispersion for the simulated data sample, integrated in energy and angle.

4.3 Classification Trees

As anticipated in section 4.1 we made use of CTs to perform the analysis. CTs were preferred to other available tools for their capability to deal with a high number of non linearly correlated input variables. TMVA[2] [Hoe09],

[2]Version 4.2.0. Further information can be found at http://tmva.sourceforge.net/

a software toolkit interfaced with the ROOT[3] data analysis framework, was used for both training and testing the CTs.

Basically a CT consists of a series of nodes, hierarchically arranged to form a tree-like structure, with a single "root" node on the top and with each node connected to other two below. The terminal nodes are the "leaves" of the tree. When an event is classified, the whole tree is traversed, starting from the root and stopping into one of the leaves; in each node a decision test is performed, on the basis of which one of the two following nodes is chosen for the subsequent step. The process is iterated until the bottom of the structure is reached, where the event is classified according to which terminal leaf it has been assigned to. It is important to note that in each node the decision test is based on the value of a single input variable.

The final classification results is given in terms of probability, for the event, to belong to the "Signal" or "Background" class. The case of a multiple choice has not been considered in this work.

The building of the CT involved two distinct phases:

- Training phase, during which the structure of the CT, as well as the classification probability assigned to each terminal leaf, are decided.
- Testing phase, during which the performance of the resulting CT are evaluated on an independent sample.

Half the simulated sample described in section 4.2 was used for the training phase and half for the testing phase, with the two mutually exclusive subsamples randomly selected from the initial sample.

4.3.1 Training

We decided to restrict the analysis to those region of the phase space where the energy dispersion presents a single definite peak in proximity of zero, otherwise our definition of Δ_{68} would become unrepresentative of the energy resolution of the instrument in that region. More in detail we eliminated from the analysis:

[3] http://root.cern.ch/drupal/

1. events with $(E_{mc}) \lesssim 56$ MeV
2. events with $(E_{mc}) \gtrsim 3.16$ TeV
3. events with $\cos(\theta) < 0.3$

Figure 4.3 shows how the distribution in the above mentioned regions differs from the typical case (a). Note that they all are (in particular the high-energy one of point 2) near or beyond the limits of the LAT operating range in terms of either energy or incidence angle, where the reconstruction algorithms are more likely to fail. Since these are also regions in which the acceptance of the instrument is very low (see section 3.4.2) we can confidently neglect them in the rest of our analysis.

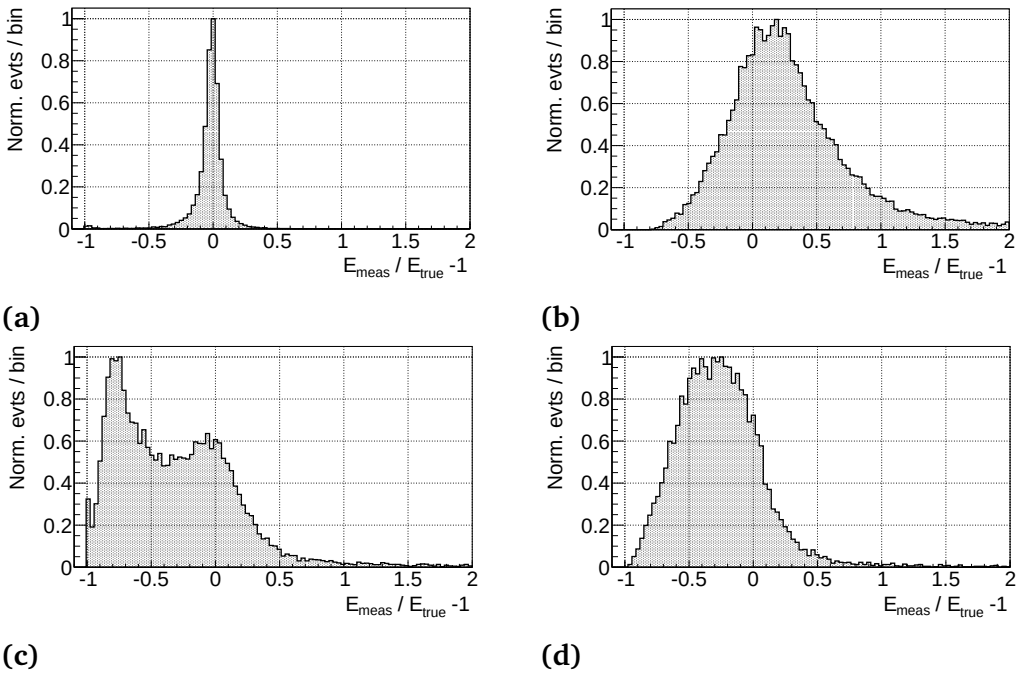


Figure 4.3. Energy Dispersion (a) at 10 GeV, (b) for $E_{mc} < 56$ Mev, (c) for $E_{mc} > 3.16$ TeV, (d) below 1 GeV, for $\cos(\theta) < 0.3$

The training phase of a CT consists in defining, at each node, a splitting criterion for the sample, under the form of a cut on a single input variable. Such criterion is given in terms of a suitable *impurity function*. Among the

available impurity functions we chose the Gini index, which for our purpose is defined at a node t as:

$$i(t) = p_S(t)(1 - p_S(t)) + p_B(t)(1 - p_B(t)) = 2p_S(t)p_B(t) \tag{4.2}$$

where $p_S(t)$ and $p_B(t)$ are the relative fractions of signal and background events at the node t, and where we made use of $p_B(t) + p_S(t) = 1$. Note that the Gini index is symmetric with respect to signal and background classes, has a maximum for an equally composed sample and a minimum for a purely signal (or background) one.

At each node the training procedure selects the variable and cut value that maximizes the increase $\Delta i(t)$ in the Giny index between the parent node and the sum of the Gini indices of the two daughter nodes, weighted by their relative fraction of events; in formula that means maximizing the quantity:

$$\Delta i(t) = i(t) - f_R i_R(t) - f_L i_L(t) \tag{4.3}$$

where f_R and f_L are the fraction of events which after the cut are assigned respectively to the right and let node and $i_R(t)$ and $i_L(t)$ are the Gini indices of that nodes. In other words, that means finding the cut that maximize the separation between the two subsamples generated with the split. The process of splitting is iterated until one or more stopping criteria are satisfied. Another criterion is necessary to assign a classification probability to the terminal leaves. In our case we made use simply of a "majority vote" rule (i.e. compared the number of signal and background events in the leaf).

A major concern when building a CT lies in the risk of *overtraining*, i.e. continuing in the splitting process over the point where the statistical fluctuation in the training sample become dominating. A solid way to enhance the performance of a classifier, at the same time reducing the risk of overtraining, comes from the procedure of *boosting* [Sha01].

Generally speaking, boosting consists in sequentially applying a classification algorithm (such as a CT) to reweighted versions of the training data and averaging over the results of the thus produced algorithms. The reweighting is done assigning increasing weights to previously misclassified

Table 4.1. Summary of the training options for the BDT.

Option	Value
Max. vertical depth	5
N. trees	500
Boosting type	$Gradient$
Shrinkage	0.3
Stochastic fraction	0.5

events. In the case of CTs this leads to building a whole *forest* of trees, each with an associated weight. The final classification is given by the weighted average of the single tree results.

The exact details of how the events are reweighted and of which weight is assigned to each tree depend on the specific boosting algorithm used. For our work we chose a *Gradient Boosting* algorithm, a technical description of which can be found in [Fri01]. Since boosting is most effective when applied to weak classifier, like CTs with a limited number of nodes, we fixed the maximum vertical depth of the single tree to 5 nodes. In this sense, boosting is also useful to reduce the risk of overtraining.

Further two standard strategies (described in [Has09]) were used to enhance the robustness of the CT, both part of the default implementation of TMVA. The first one is to artificially reduce the learning rate of the boosting algorithm through an external parameter (called "*Shrinkage*") which can vary between 0 and 1, where a value of 1 means that the learning rate is not modified. The other strategy consists in introducing a resampling procedure, using random subsamples of the training events for growing the trees. Such procedure is called *stochastic gradient boosting*. The sample fraction used in each iteration can be controlled through an external parameter, varying from 0 to 1. A value of 1 means that the whole sample is used in each iteration (and corresponds to disabling this procedure). Both those procedures are intended to stabilize the response of the CT against statistical fluctuations of the training sample.

We performed several test to find the best combination of options in terms of CT performance. The full final configuration used is reported in

Table 4.1. It is important to underline that the CT response is very stable against variations of these parameters and that only marginal difference in performance were found for a large set of their possible combinations. Note also that a low value for the shrinkage parameter demands growing a greater number of trees then otherwise, comporting a slightly greater, but still acceptable, computational cost.

4.3.2 Variables

Another fundamental aspect of the training is the choice of the input variables. An advantage in using CTs when dealing with problem of high dimensionality is that they are almost unaffected by the presence of useless input variables (i.e. variables with little or no discrimination power), since the algorithm simply ignores such variables, never using them to split the sample. As a consequence, we need not to be conservative in selecting our list. A total of 42 variables were selected among the various figures of merit produced by the reconstruction phase. Note that we did not include the reconstructed energy in this list, because test showed that doing so can introduce artificial features in the spectra.

The *TMVA* toolkit provides a ranking of the input variables used, sorted by their importance in the classification process. Though such ranking can not in any way be considered as a reliable metric (since it tends to fluctuate significantly with small changes in the training configuration setup) it provides interesting information about which variables carries the best discrimination power.

Not surprisingly, most of them are related either with the geometry of the event or with its topology in then CAL. As a general rule, energies of events with a greater observable development in the CAL are better measured. At high energy the reconstruction is mainly limited by shower leakages in the gaps between crystals and towers and by the shower containment. At low energy the topology in the TKR becomes more relevant, since events lose a significant fraction of the energy there (remember from section 3.2 that the energy released in the TKR is estimated by treating it as a rudimentary sampling calorimeter).

To give a feeling of how the classification process operates, here we report a list of the most relevant variables, as indicated by the TMVA ranking system, commenting briefly on their connection with the energy measurement:

- Total radiation lengths in the active volume of the calorimeter, integrated along the event axis (line connecting the first hit in the tracker to the CAL energy centroid). Events with a longer development in the CAL provide the reconstruction algorithm with more information to operate on, so they generally benefit of a better energy measurement.
- Distance of the energy centroid from the nearest tower boundary. It is strongly related with the amount of lateral shower leakage in the gaps between the towers.
- z-direction cosine of the best track. The inclination angle influence many aspects of the topology of the events, from the conversion point to the amount of material traversed in the CAL and in the TKR.
- Fraction of the total energy deposited in the CAL which is released in the back half (last 4 layers). This variable is significant in two different ways (see Figure 4.4): a value of zero, in fact, is often associated with a poor energy measurement, following from events with a very small development in the CAL, (typically low-energy events, especially if early converting in the TKR). A value of zero is also assigned by default to events without any energy deposit in the CAL. A great fraction of energy in the back half of the CAL, on the other side, can be index of an event "spoiled" by an extensive shower leakage beyond the back of the instrument, and thus of a poor shower modeling.
- Approximate fraction of the shower volume which falls in inter-tower gaps. The relevance is quite obvious.
- Effective layer-by-layer edge correction (see section 3.2) mainly due to the gaps between Cal modules. The greater is the correction that must be applied to the raw energy measurement, the higher becomes the uncertainty.

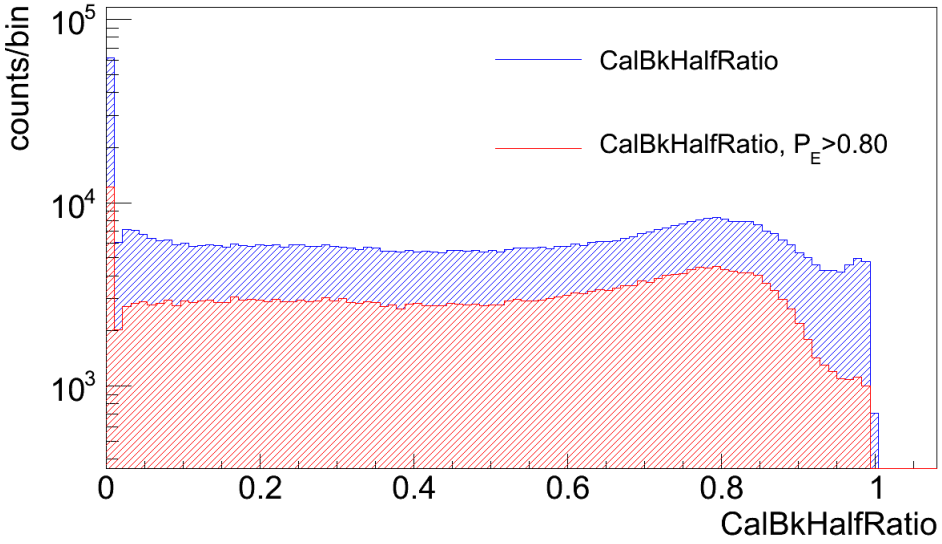


Figure 4.4. Histogram showing the distribution of the back-half fraction of the total energy deposited in the CAL (CalBkHalfRatio) in the *allGamma* simulation before (blue) and after (red) a cut on P_E. After the cut, the high peak at zero is heavily suppressed. There is also a fall near the value of 1.

- χ^2 reported by the full shower profile fit used in the reconstruction stage (see 3.2). This variable provides useful information about the quality of the fit.

- Root mean square of transverse position measurements. The reconstruction algorithm is designed to assign greater weights to the nearest crystals clustered around the principal axis and the centroid of the shower. More "compact" showers are usually better reconstructed then spreader ones.

- Position of the maximum of the shower as reported by the full profile fit (using the cal direction). The position of the maximum is one of the most important parameters in estimating the energy of the shower.

4.3.3 Testing

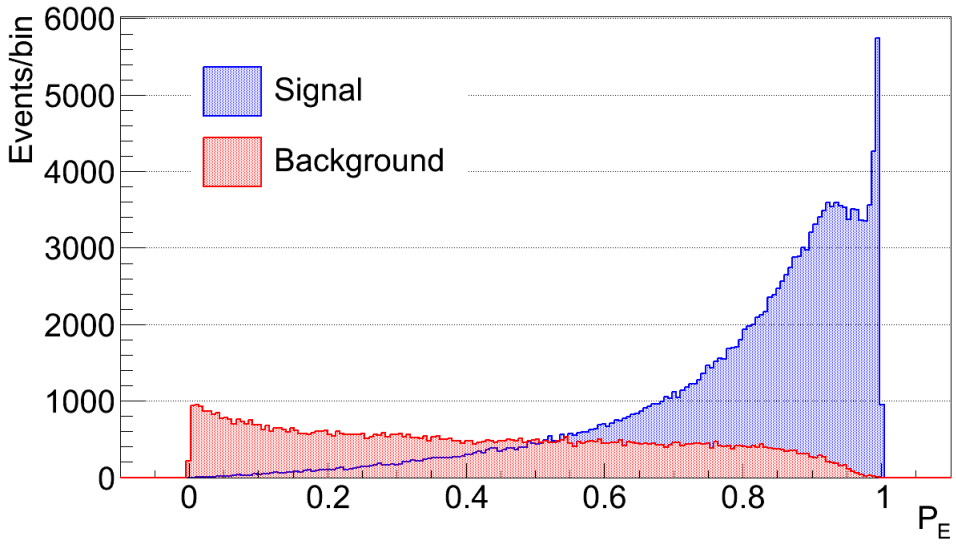
After the boosting procedure, the response of the CT is given in the form of a real number, comprised between 0 and 1, resulting from the weighted average of the individual tree response. Such average can be conveniently interpreted, for each event, as a sort of probability to belong to the signal class. We will use this variable, for shortness named P_E, as an event-by-event measure of the quality of the energy reconstruction. We expect, on average, that events with a value of P_E closer to 1 are more likely to have a better energy measurement compared to events with a value closer to 0.

Fig. 4.5 shows the distributions of P_E for both the training and testing sample. It is important to notice how they look very similar, since a big discrepancy could be an indication of overtraining. As expected, the distribution for signal and background classes is highly peaked near 1 and 0, respectively.

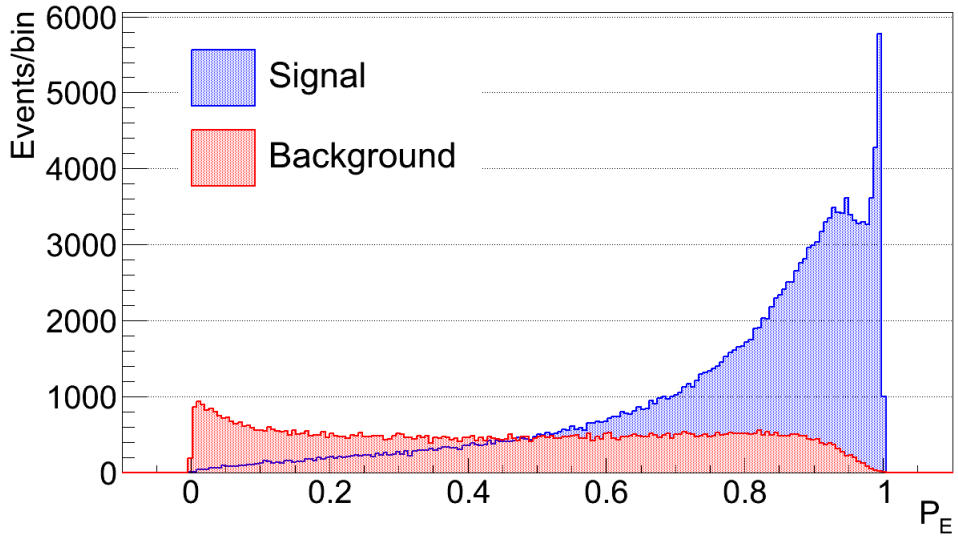
A standard way to evaluate the performance of a classifier is to plot its rejection power, namely the residual fraction of background after a cut on it, as a function of the selection efficiency for the same cut. Such representation is not entirely suitable for our analysis. In fact, as explained in sec. 4.1, our definition of signal and background is purely conventional and all the events analyzed are, in principle, "signal" gamma-ray events.

Remembering that the goal of our classifier is to select samples with a better energy resolution, we found more interesting to plot the value of Δ_{68}, computed on a sample selected with a cut on P_E, as a function of the *global* efficiency of the same cut. Here, "global" means that it is computed including both signal and background classes. Fig. 4.6 shows the above described performance plot for various regions of the phase space.

The first thing that can be noticed from the plot is that cutting on the classifier effectively improves the energy resolution of the sample. Not surprisingly, the algorithm is more efficient at high energy, where the shower development in the CAL is more extended and, consequently, there is more information for the CT onto which operate. For the same reason the dependence on the energy is much more accentuated for events with a high inclination angle, since they traverse more RL on average.

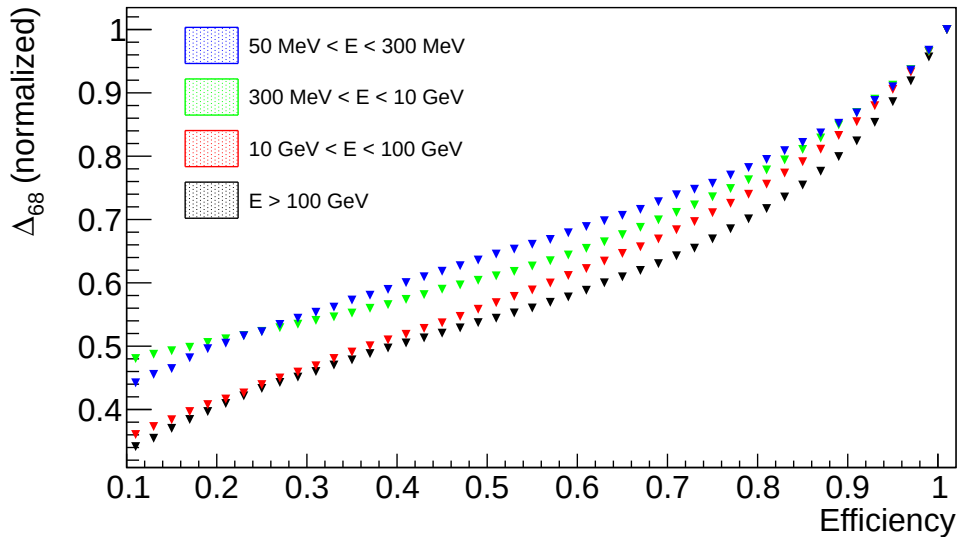


(a)

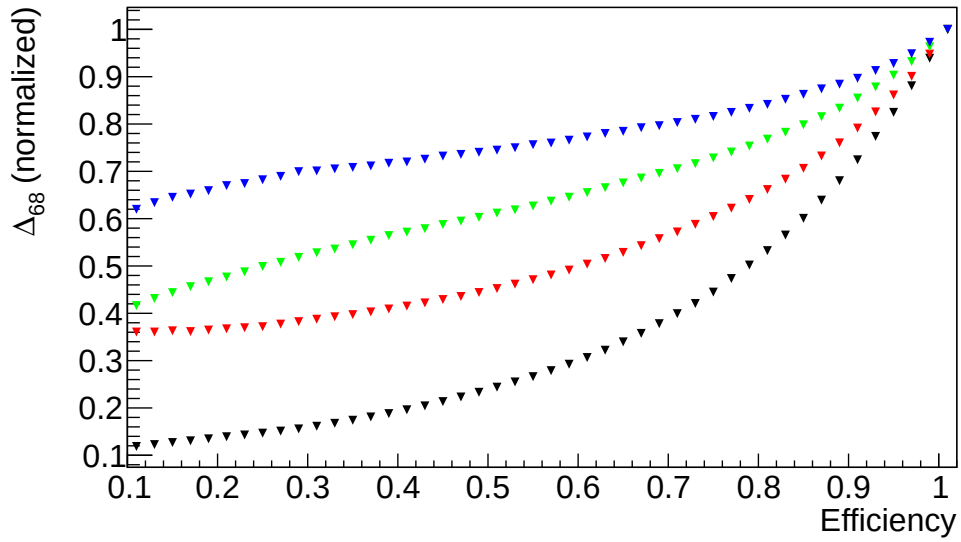


(b)

Figure 4.5. P_E distribution for (a) train and (b) test sample.



(a)



(b)

Figure 4.6. Performance of the CT in different energy ranges (a) on-axis ($\cos\theta > 0.7$) and (b) off-axis ($\cos\theta < 0.7$). Each point represents a value $P_0 \in [0\,1]$. On the x-axis is shown the global efficiency of the cut $P_E > P_0$, on the y-axis the value of Δ_{68} (normalized) computed on the subsample defined by that cut. The lower is the curve, the better are the performance of the CT.

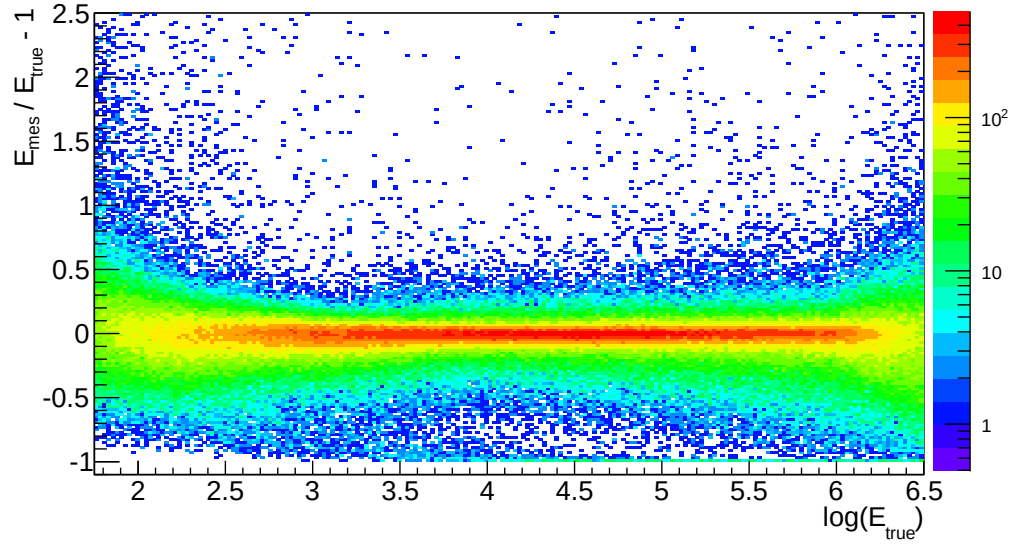
Another way to visualize the effect of a cut on P_E is to compare the energy dispersion before and after a cut removing the worst 25% of events (i.e. the first 25% quantile in P_E.) This is shown in Figure 4.7 and 4.8, where it can be seen that the vast majority of events removed comes from the tails of the distribution.

Two alternative approaches have been explored, in an attempt to improve the classifier performance. In the first, we tried to exploit the intrinsic differences between the two separate reconstruction algorithms currently in use to estimate the photon's energy (see section 3.2). In the vast majority of cases an event is assigned to either one algorithm or another (the choice depending mostly on the energy range), and the two populations share only a small overlap zone, where the energy is computed as the weighted average of the estimates produced by the two algorithms. The weight used is stored in a dedicated *figure of merit*[4], namely W_E, with values spacing from 0 to 1. Since the factors limiting the resolution of the energy measurement are partially different at low and high energy, it seemed plausible that the classification would benefit from undergoing the same separation; so we split the training sample with a cut on W_E into two mutually exclusive subsamples, and trained a different CT for each of them. The final estimator was obtaining averaging the output of these two CTs, weighted with the same value of W_E.

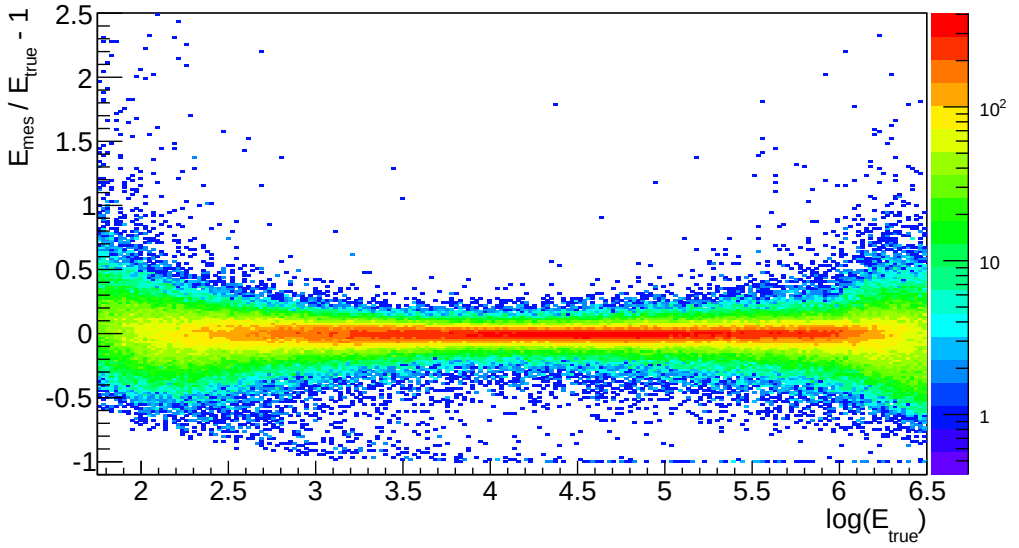
Despite sounding promising, we did not find any significant benefit in this approach, the performance of the resulting classifier being essentially identical to the one presented above. A possible explanation is that, since W_E is comprised in the list of the input variables of the original CT, the algorithm already uses it in an almost optimal way. It's noticeable, in fact, that such variable is not in the top-ranking list reported by TMVA, meaning that the distinction between high-energy and low-energy events is less relevant then we expected, when compared to others of topological or geometrical nature.

The second strategy tested was to change the definition of the "signal" class given in section 4.1, extending it to events in a 95% containment win-

[4]Note that only a single weight is needed to average two numbers, the other being automatically obtained for complementarity.

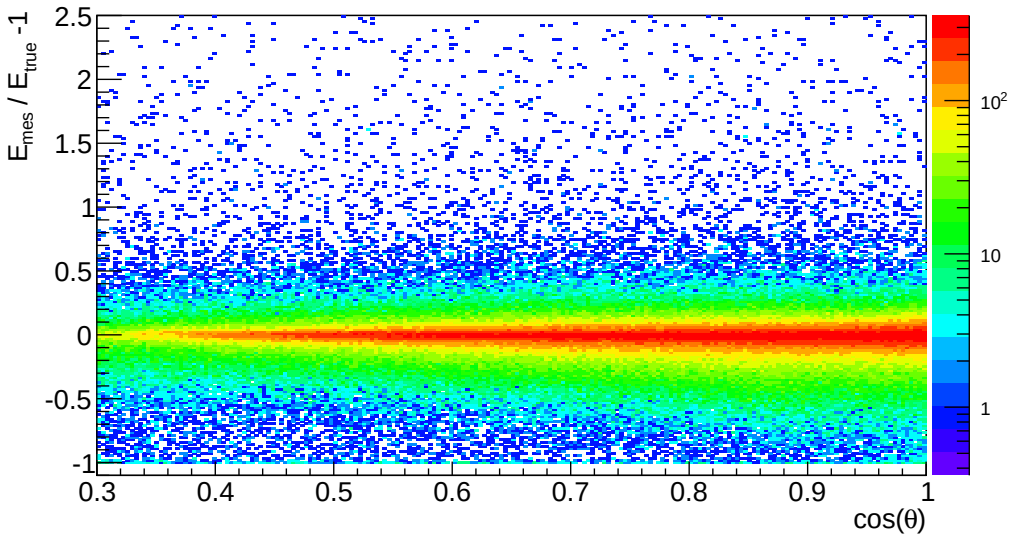


(a)

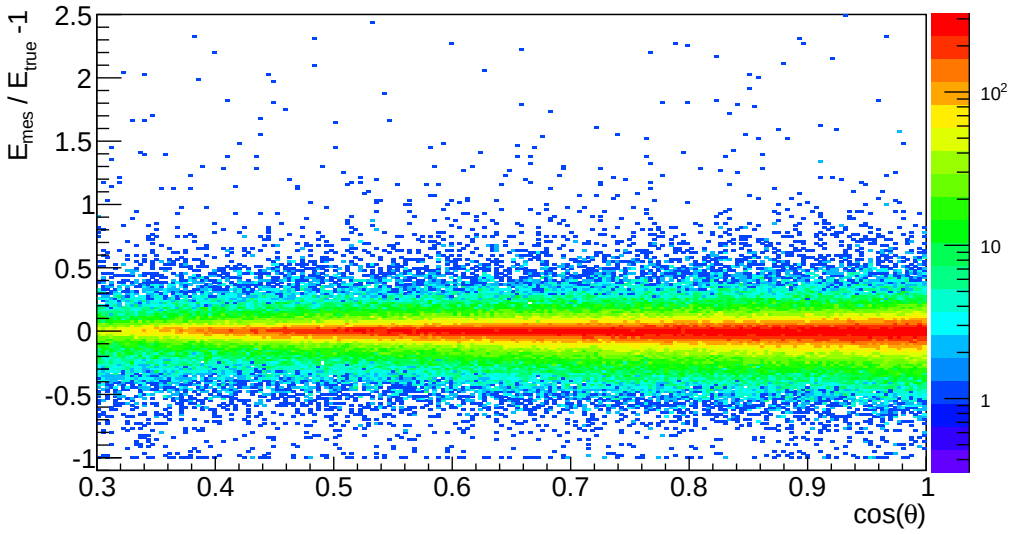


(b)

Figure 4.7. Energy Dispersion as a function of energy before a and after b a cut removing the first 25% quantile in P_E



(a)



(b)

Figure 4.8. Energy Dispersion as a function of inclination angle before a and after b a cut removing the first 25% quantile in P_E

dow. The idea behind this approach was to allow the CT to focus more on the very "bad" events (i.e. those in the far tails of the energy dispersion). However, the classifier produced with this approach proved to be substantially less effective then the initial one (Figure 4.9), so we decided to stick to the original set-up.

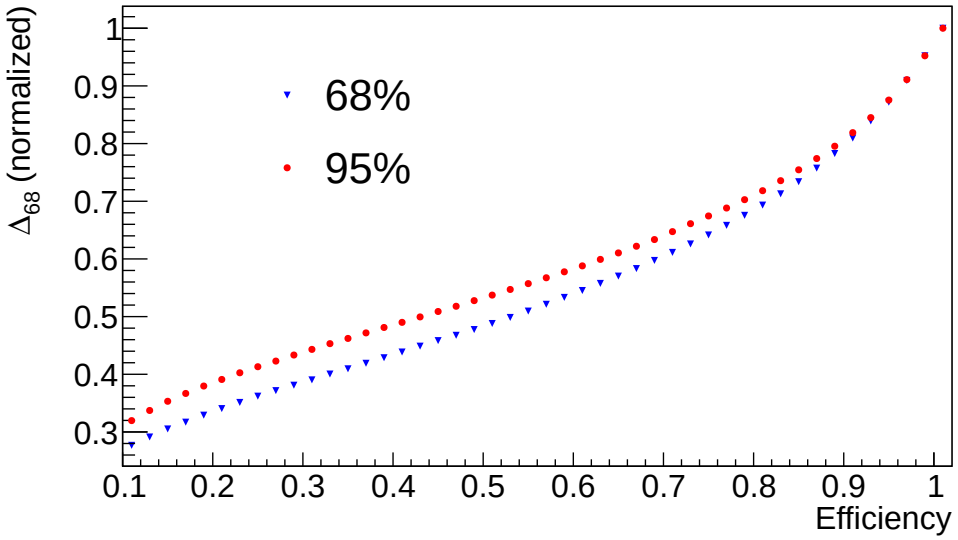


Figure 4.9. Performance plot for the original CT (blu triangles) and for a CT trained with a definition of signal extended to a 95% width window (red circle), integrated across the whole phase space. The former performs best on average.

4.4 Using P_E in the search of spectral lines

The classifier described above is used mainly as part of the standard event selection, in the terms reported in chapter 3, to remove the tails of the energy dispersion. However, a more interesting application of the event-by-event information it provides about the quality of the energy measurement has been studied, in the search for spectral lines.

The search for spectral lines has been the subject of a certain interest over the last years,in connection with indirect Dark Matter (DM) searches (see section 1.2) and the possible detection of a monochromatic γ-ray sig-

nal produced by an annihilation process like $\chi\chi \rightarrow \gamma X$ or by a decay process like $\chi \rightarrow \gamma X$.

Whether these processes could produce a signal observable by the LAT depends on many factors, first of all the mass of the WIMP particle, which must be in the right range of energy. Detections of a line-like feature at an energy of ~ 130 GeV have been reported in the literature [Bri12] and a detailed search for such features has been performed and periodically updated by the Fermi collaboration [Abd10] [Ack12b] [Ack13].

A detailed discussion of these results is far beyond the purposes of this study. Here we focused on a different aspect of the problem: using event-by-event information about the quality of the energy measurement in order to enhance the sensitivity to a hypothetical monochromatic line in a given gamma-ray spectrum. In fact, it's rather intuitive to expect that, if a similar line-like signal is actually present, events with a better energy measurement should cluster with a smaller dispersion around the energy at which the signal occurs. Our aim is to try capturing such behavior including P_E into the analysis.

This strategy has already been tested by the Fermi collaboration in the most recent iteration of the line-like features search ([Ack13]), which used the P7_REP version of the data. Since in Pass 8 the whole event reconstruction and, in particular, the building of P_E, has been deeply revisited, it is important to check the effectiveness of such strategy within this new framework.

When fitting a monochromatic line of energy E_γ we expect the distribution of observed energies, $f_{sig}(E')$, to follow the Energy Dispersion at the line's energy $D(E'; E_\gamma)$ (see 3.4):

$$f_{sig}(E'|E_\gamma) = n_{sig} \int D(E'; E)\ \delta(E - E_\gamma) dE \tag{4.4}$$

$$= n_{sig}\ D(E'; E_\gamma) \tag{4.5}$$

where n_{sig} is the number of signal counts and it's a free parameter of the fit. A parametrization of $D(E'; E_\gamma)$ in the energy range considered with an analytical form of some sort is required in order to perform the fit. In [Ack13]

a sum of Gaussians was used, whose parameters were estimated by fitting the distribution of $D(E';E_\gamma)$ obtained from the *allGamma* simulation (see section 2.6). Such parametrization can be improved if the simulated sample is divided in bins of P_E and the fitting procedure is performed bin per bin. This improves the description of the Energy Dispersion in two ways:

1. Increases the number of parameters of the fit.

2. Reproduces better the different dispersions and biases of the energy measurement for events with different topologies in the detector (information contained in P_E).

We tested this procedure on a simulated dataset, in order to study its effectiveness in the context of Pass 8.

First, as said, we fitted the energy dispersion with a triple Gaussian sum in bins of $\log(E)$ and P_E, in order to appropriately model $D(E';E)$. The distribution of $D(E';E)$ was obtained from the *allGamma* simulation, filtered with the combination of cuts of the Pass 8 SOURCE class. In detail, we divided the sample in 20 equally spaced bins of $\log(E)$, spacing from ~ 17 Me V ~ 1.7 TeV, and in 9 equally spaced bins of P_E, spacing from 0.1 to 1, and performed the fit in each bin.

Then, we used the set of parameters produced by the fits to extrapolate, with a simple interpolation, the functional form $f_{sig}(E,P_E)$ of the observed energy distribution for a line-like signal at any given energy and value of P_E.

As test data sample we used the superposition of a line-like signal and a power-law background distribution. Specifically we generated:

- A monochromatic line with intensity N_{sig} at a fixed energy E_γ.

- A background with a power-law spectrum of index Γ_B, in a window of width $\pm 2\sigma_E$ centered on E_γ and normalized so that the total number of background events was N_{bkg}.

here σ_E is the average energy resolution of the LAT at the energy E_γ.

We fitted this data sample with the sum of the $f_{sig}(E,P_E)$ described above and a background power-low. In the fit we only fixed the energy E_γ

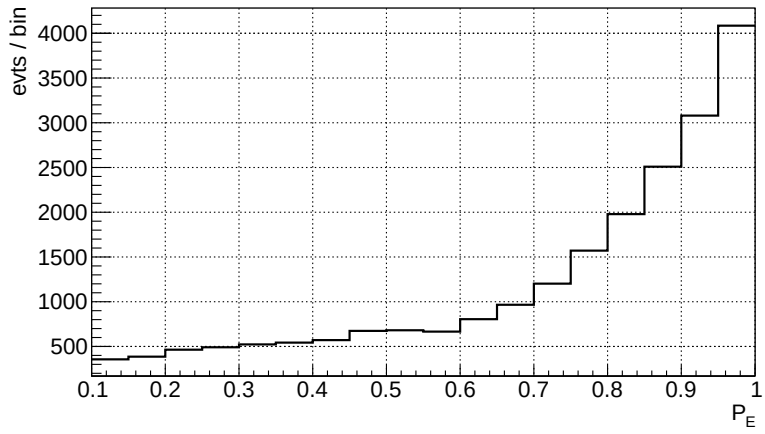


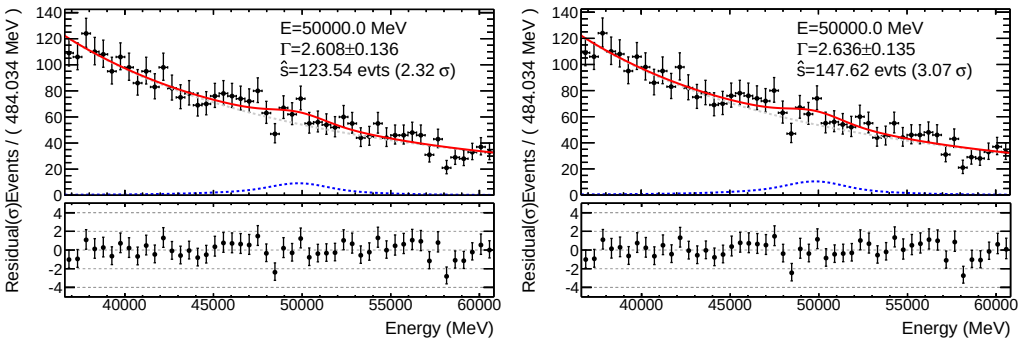
Figure 4.10. Distribution of P_E for the simulated dataset, showing the binning used for the sampling.

to which the signal occurs, leaving N_{sig}, N_{bkg} and Γ_B as free parameters. We did not apply any correction to the background index, since the effect of the energy dispersion on a power-law is small enough that it can be neglected.

In order to perform the fit, the knowledge of the value of P_E for each event is needed. In our simulation we simply took the distribution of P_E for the filtered *allGamma* sample, divided it in 18 uniform bins and pick one bin randomly for each event, with a probability given by the distribution itself (Figure 4.10). We made this choice because we were interested in a generic evaluation of the procedure, so using an isotropic flux was adequate for the task. Note that using a different distribution of P_E, for example matching the observing profile of a specific region of the sky or of a specific celestial γ-ray source (which is what one would do in a real analysis), can in principle lead to slightly different results.

In order to study the effectiveness of the procedure we made 1000 realization of the test dataset, simulating a line signal at 50 GeV. Each time, for comparison, we also repeated the fit with an averaged $f_{sig}(E)$, which did not use the information from P_E. Figure 4.11 shows an example of a single realization of the simulation.

As check test, we also repeated the whole procedure with a purely back-



(a) **(b)**

Figure 4.11. Example of a random genereted dataset, fitted with (a) an average $D(E)$ and (b) a $D(E, P_E)$ binned in P_E

.

ground dataset. The specifics and results of the various simulations are summarized in Table 4.2

In all the tests, the fit was able to get the right spectral index of the background power law. Also, in the check test, it correctly assign a value of zero to the signal counts in most realizations. (Figure 4.12). Finally, the average significance of the fit increase of $\sim 15\%$ when taking into account the information from P_E (Figure 4.13).

This significance gain confirms essentially the result of an analog study made using Pass 7 [Ack13]. This fitting procedure will be used in a search for line-like signals currently in preparation by the *Fermi* collaboration, where the data reprocessed with Pass 8 will be used.

Table 4.2. Summary of the simulations specific (up) and fit results (down).

	Averaged $D(E)$	Binned $D(E, P_E)$	Control Dataset
N_{bkg}	10^5	10^5	10^5
N_{sig}	1250	1250	0
Γ_B	-2.6	-2.6	-2.6
N_{sig}	1253	1255	50
σ_N	194	171	77
Γ_B	-2.564	-2.565	-2.605
σ_Γ	0.013	0.013	0.059
$Sign.$	6.44	7.43	0.36
$\sigma_{Sign.}$	0.99	1.00	0.57

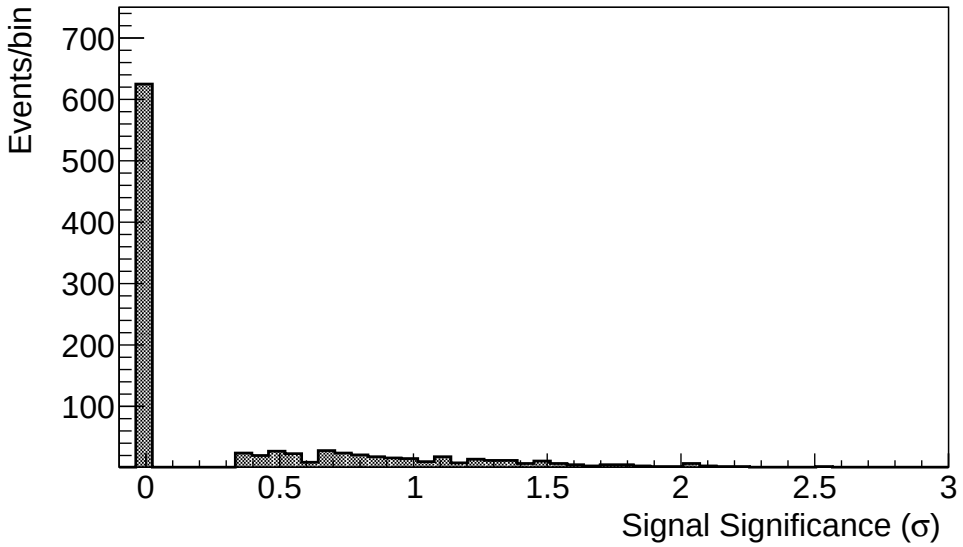


Figure 4.12. Significance of the line fit for the simulated with zero signal event generated (control dataset).

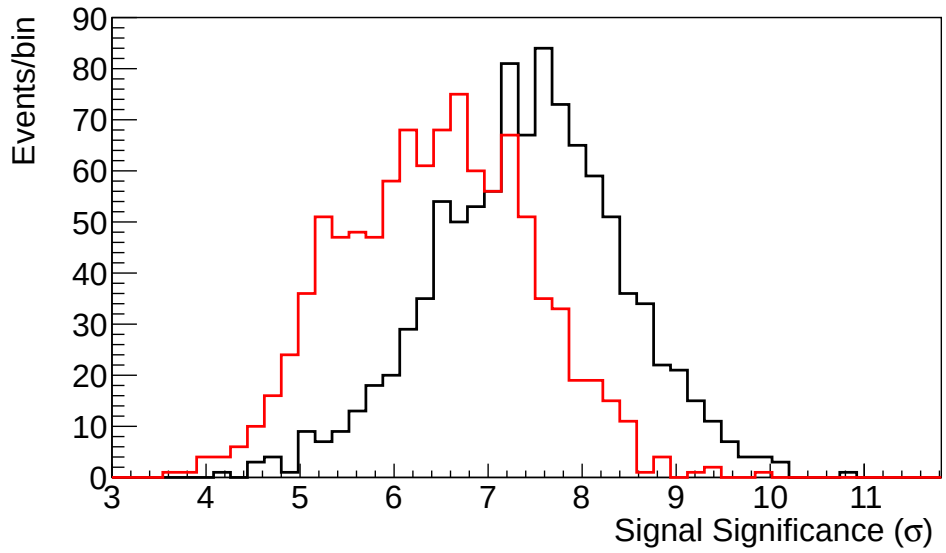


Figure 4.13. Significance of the line fit for the simulated dataset for a binned $D(E, P_E)$ (black) and a averaged $D(E)$ (red).

Chapter 5

High-Energy Electron and Positron Spectrum

5.1 Introduction

High-energy cosmic-ray electrons and positrons (here generically referred to as "electrons") constitute a peculiar component of the cosmic radiation. In fact, unlike protons and heavier nuclei, they rapidly loose energy by synchrotron radiation on galactic magnetic fields and by inverse Compton (IC) scattering on the interstellar radiation field, thus presenting a significantly steeper spectrum at Earth with respect to protons (Figure 5.1).

At low energy (up to a few GeV) the cosmic-ray electron (CRE) spectral shape is heavily influenced by solar modulation; as energy increases, this effect becomes progressively less relevant and the spectrum is expected to be essentially determined by other factors: the slope of the source injection power-law, the diffusion in the turbulent Galactic magnetic fields and the two aforementioned energy-loss mechanisms.

Above a few hundred GeV, the growing contribution of synchrotron emission and IC scattering to the CRE average lifetime implies that the majority of the electron and positron flux observed on Earth must be produced by sources closer than a few hundred pc. Thus, measuring the high-energy CRE spectrum with high accuracy can provide important information to constrain theoretical models of production and propagation of CRs in the

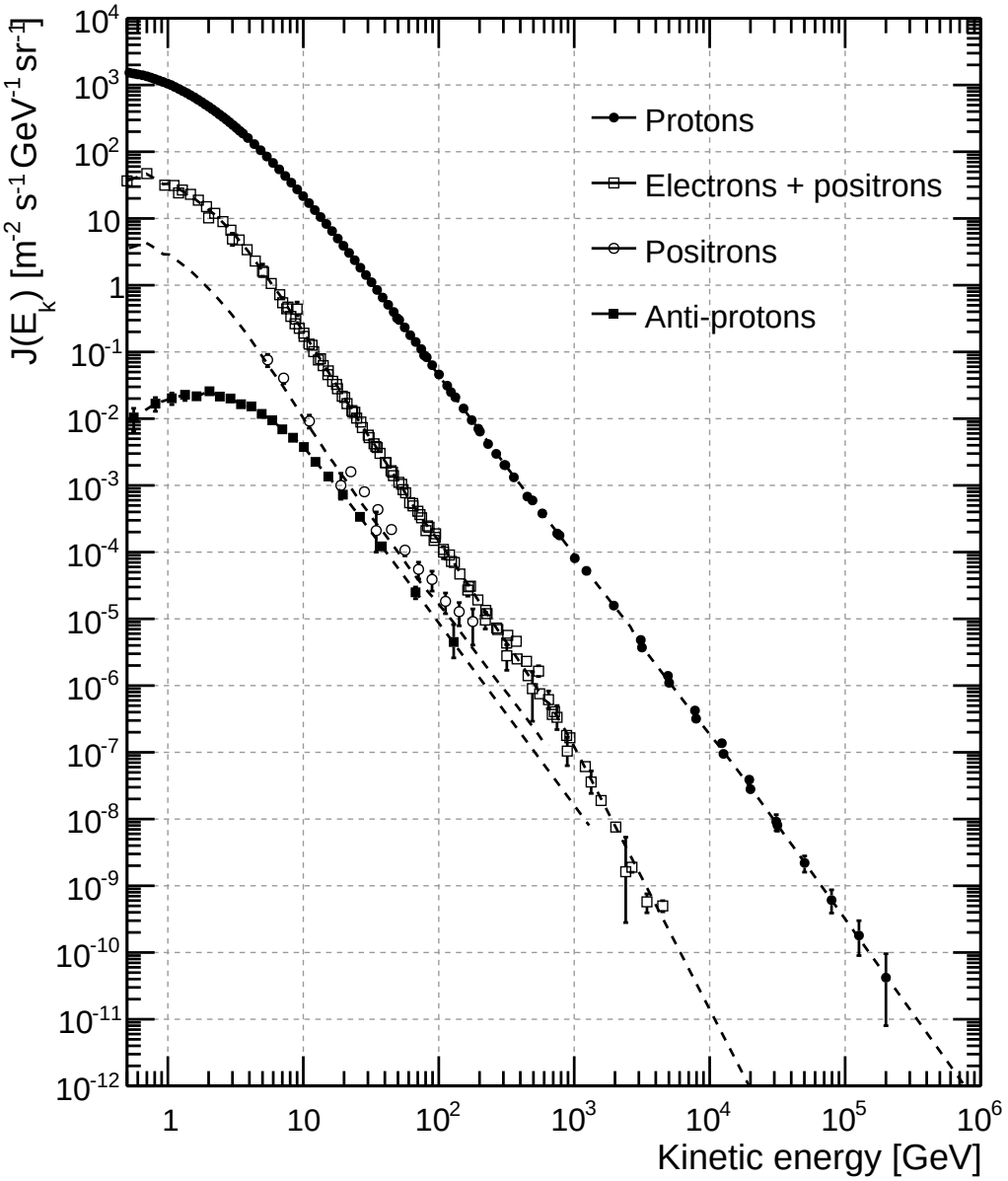


Figure 5.1. Comparison between the differential spectra of various CR charged species. The steeper nature of the electron spectrum with respect to the proton's one is clearly visible.

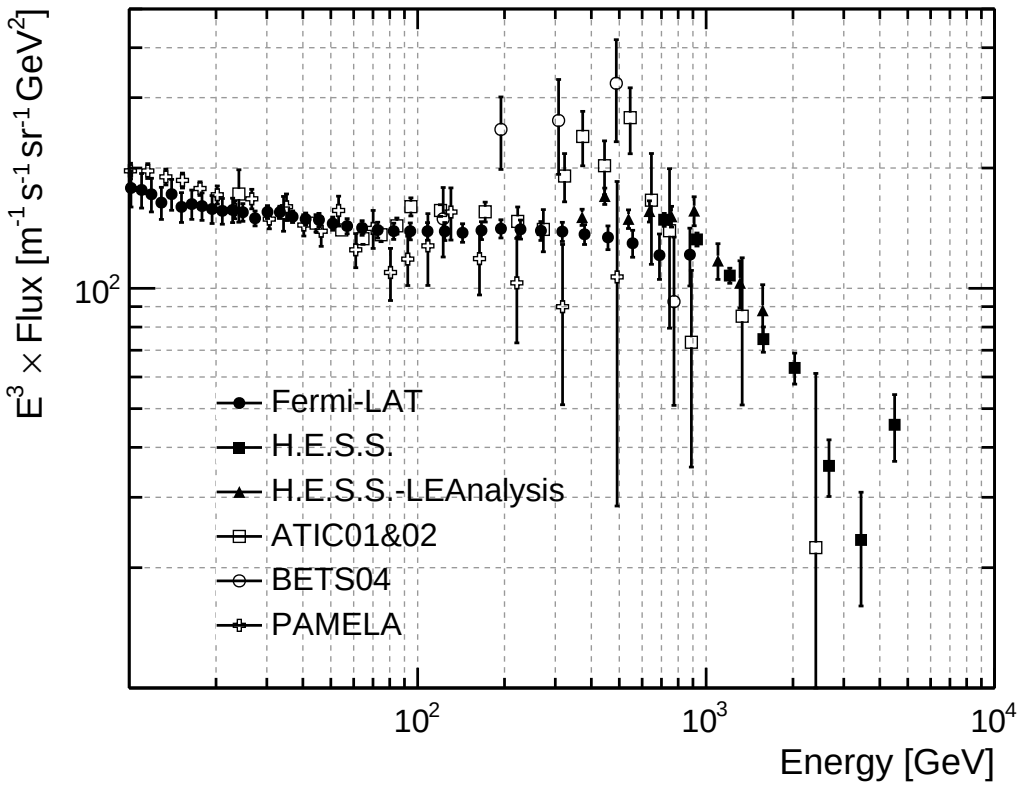


Figure 5.2. Comparison between different measurements of CRE differential spectrum.

nearby or local galactic space, possibly including signatures of new physics.

Since electromagnetic (EM) cascades are common to both electron and photon interactions in matter, the LAT is also by its nature a detector for electrons. A first measurement of the inclusive spectrum of CREs between 20 GeV and 1 TeV, based on the data taken in the first six months of mission, was published by the LAT collaboration in 2009 [Abd09b], followed in 2010 by an update [Ack10], including more statistics and extending the inferior energy limit to 7 GeV.

The development of the new event-level analysis framework, being expected to improve both the LAT acceptance and energy resolution in the high-energy range, represents a promising occasion for an updated measurement of the high-energy CRE spectrum. In this chapter, the study of a preliminary event selection is presented, aimed at discriminating the lep-

tonic component from other charged species in the detector.

5.2 Event selection

The onboard GAMMA filter (see section 2.5) is optimized to identify and reject charged particles, mainly by means of veto signals in the ACD. However, in order to keep the maximum possible efficiency for the rare high-energy photons, the onboard filter is disabled for events releasing at least 20 GeV in the CAL, thus allowing to effectively collect CRE events in the high-energy range. More in detail, since at normale incidence an electron of 50 GeV releases approximately half of its energy in the detector, above this energy the filter becomes fully efficient for CRE study. Below this threshold some information can be recovered from the DIAGNOSTIC filter, which provides a small unbiased sample of all kind of events triggering the instrument.

A measurement of the electron spectrum with the LAT requires three main tasks to be accomplished:

1. Build a selection to isolate CREs events from protons and from other charged components.

2. Estimate and subtract the residual background contamination rate.

3. Estimate the instrument acceptance for such selection and use it to appropriately translate the count spectrum into ana ctual measurement of the CRE intensity.

Contamination from γ rays is easily addressed, since the photon rate at high-energy is relatively low compared to charged species, and since they are strongly discriminated by the absence of signal in the ACD. Thus, the main focus of this study has been on removing the hadronic component.

We chose to base on CTs as our main tool for developing the event selection. Contamination study and optimization of the selection algorithms made use of the *background* Monte Carlo simulation described in

section 2.6. Other two dedicated simulations were used, similar to the *allGamma* but involving pure samples of high-energy electrons (*allElectron*) and protons (*allProton*).

5.2.1 Pre-selection

First we defined a preliminary cut, similar to those described in section 3.3, selecting only events for which at least a minimal amount of information is provided from the TKR and the CAl. We required:

- at least one successfully reconstructed track in the TKR;
- at least 4 RL of intersection between the projection of the track and the CAL;
- at least a minimal estimated quality for the energy and direction measurement.

We also ask events to release a minimum of 20 GeV in the CAL, to avoid any possible alteration due to the effects of the GAMMA filter.

We applied this preliminary selection to a small sample of data ($\sim$few months) and to all the MC simulated datasets. Then we performed a detailed study to individuate, among the variables produced by the various reconstruction algorithms, those with the best separation power between electrons and other charged species. For each variable, we checked also that its distribution in the *background* simulation matches with the one in the flight data (once the two samples have been appropriately scaled for the active orbital time period of collection).

5.2.2 Removing alphas and heavy nuclei

One of the technical difficulties encountered at this stage is that the modeling of inelastic interactions of alpha particles and heavier ions used in the simulation is in strong disagreement with observations. This issue has already been reported by the *Fermi* collaboration, for example in [Ack12a], and can be easily visualized comparing the energy released in the ACD

by an incoming particle respectively in the data and in the MC simulation (fig: 5.3). It must be emphasized the fact that in a plot like fig: 5.3 we are comparing two distributions (the sum of all the MC components and the data) that, were the MC simulation perfect, should be exactly equal, in terms of both shape and normalization.

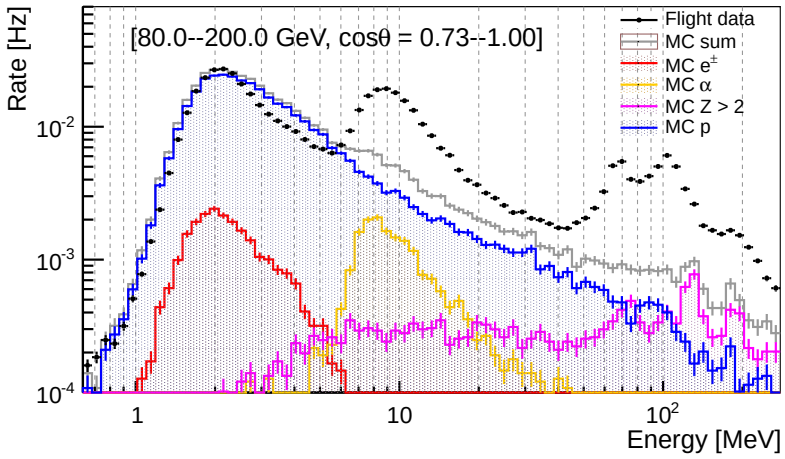


Figure 5.3. Energy released in the ACD in a cone of 15 degrees ahead the best reconstructed track, corrected for the inclination angle between the track and the hit tile. Different charged species in the MC *background* simulation are plotted in different colors: blue for protons, red for electrons/positrons, orange for alpha nuclei and violet for heavier ions. The grey line is the sum of all the MC components, black dots represent data.

In the plot, the distribution of the data clearly shows a series of peaks, corresponding to the average energies released by Minimum Ionizing Particles (MIPs) for increasing values of their charge Z; in the simulation, on the other side, only the first peak appears adequately reproduced. As for alphas and heavier nuclei, even if the positions of their peaks are roughly correct, a great number of them seems to be simply missing in the simulation, indicating that the model largely underestimate their inelastic interactions in the CAL.

Despite not being a issue for γ-ray analysis, since alpha particles are a source of background of much smaller entity compared to protons (and typically easier to reject), the presence of such discrepancy negatively affects our capability to individuate well reconstructed variables in the simulation.

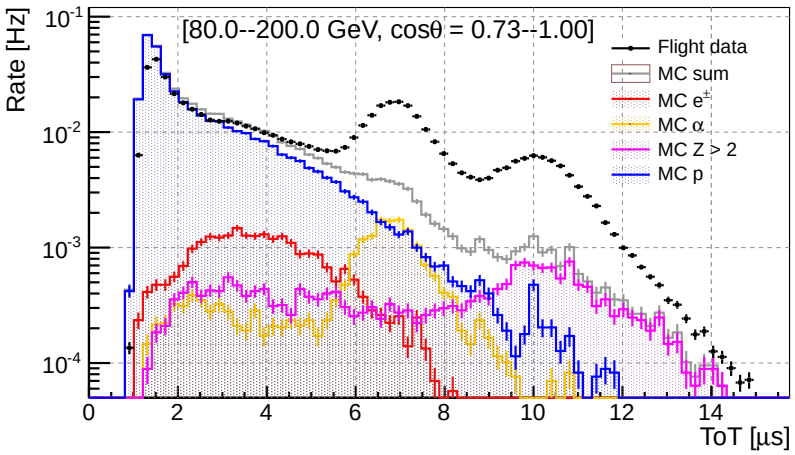


Figure 5.4. Average ToT of the layers along the best reconstructed track, corrected for the incoming particle direction.

Furthermore, it is possible that it can degrade the performance of the CT that will be trained on the the simulated sample. For these reasons, we decided to add a further cut to our pre-selection, in order to remove particles with Z>1 from the sample.

For that purpose, we exploit the dependence of the ionization from the charge of the particle. The LAT provides two independent measurements of this quantity: one, already mentioned, comes from the amount of energy released in the ACD; the other comes from the average ToT (see section 2.1) in the TKR layers (Figure 5.4).

We defined a hand-made cut in the plan formed by these two variables, with the goal of removing a significant fraction of alphas and heavy ions. Note that, at this stage of mere pre-selection, we consider a high efficiency on electrons to have priority over the rejection power. The cut was optimized using the *background* simulation and has an estimated efficiency of $> 95\%$ on the electron sample, with a fraction of particle with Z>1 rejected of $\sim 80\%$.

Figure 5.5 shows the distribution of the aforementioned ToT variable after the cut. The former un-reproduced high-Z peaks have disappeared and the agreement between the data and the MC simulation, though still not perfect, has sensibly improved.

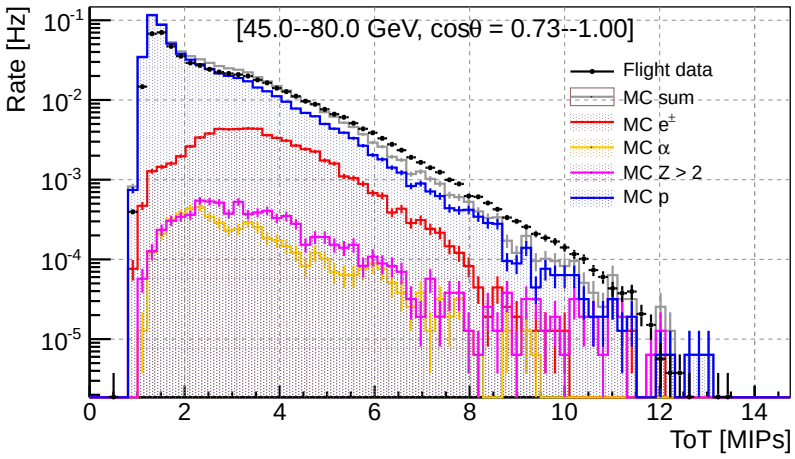


Figure 5.5. Average ToT of the layers along the best reconstructed track, corrected for the incoming particle direction, after the cut used to remove alpha particles and heavier nuclei.

5.2.3 Template fitting

After this pre-selection stage, we performed a detailed study of the agreement between data and MC simulation for several different variables. For that purpose, we introduced a procedure of *template fitting*: given a variable, we scaled the distribution of each component (electrons, protons and other charged species) of the simulated sample with an appropriate normalization factor, so that their sum fits the distribution of the data. The fit is done in bins of energy and incidence angle.

Figure 5.6 shows an example of template fitting in a single bin of energy and incidence angle. The results of the fit indicates an excess of protons in the MC simulation; we consistently observed a similar excess, of approximately 30-50%, in different variables and in different bins of energy and incidence angle, strongly suggesting that the simulation overestimates the number of protons at this stage of the selection.

Since we rely on the simulation to estimate and subtract the residual background contamination after the selection stage (see next section), we are currently studying a way to take into account the systematic error that can be introduced in the final spectrum by this discrepancy.

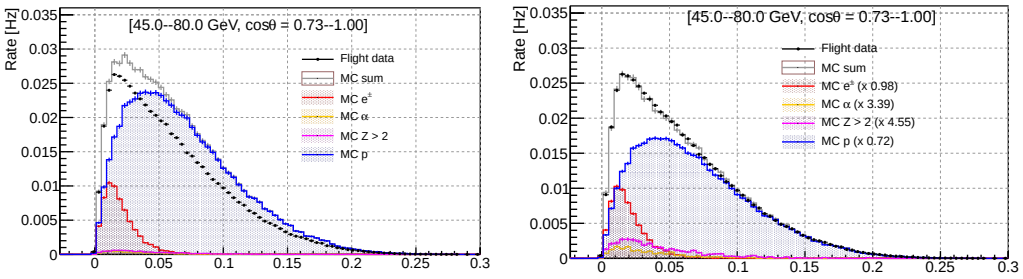


Figure 5.6. Example of template fitting: the picture on the left shows the distribution of a variable (in this case a variable correlated with the asymmetry of the shower in the CAL) in the MC simulation and in the data. On the right there is the same distribution after the template fitting: the single MC components have been scaled so that their sum matches the data. The value of the scale factor for each component is shown in the legend.

5.2.4 Main selection

At this point, most of the remaining hadronic background is composed of protons. As anticipated, we made use of CTs to isolate and remove such component from the data (refer to Chapter 4 for a detailed discussion about this tool). For the selection, we exploit the different interactions of the leptonic and hadronic populations within the detector.

Electron and protons[1], in fact, produce events of rather different topologies in the LAT, as for the latter the contribution of strong interactions allows for a much larger variety of secondary produced particles, at the same time redirecting a significant fraction of energy into nuclear reactions.

The typical dimensions of a shower produced by inelastic interaction of a hadron in the instrument (usually in the CAL) are dictated by the nuclear interaction length λ_{int}. For an electromagnetic (EM) shower the analogous relevant quantity is X_0. In particular, for a CsI crystal $\lambda_{int} = 38.04$ cm and $X_0 = 1.86$ cm.

This difference affects both the longitudinal and transverse extension of

[1]For brevity, we will always refer to "protons" here and in the following, but the discussion applies to the whole hadronic component of CR in the detector (i.e. protons, alpha particles and heavier ions)

the shower. Typically an EM shower is entirely contained in 2 or 3 Molière radii R_M, 3.53 cm in CsI, while an average hadronic shower is larger (Figure 5.7). Furthermore the distribution of released energy in an EM shower tends to be more concentrate along the shower axis (Figure 5.8).

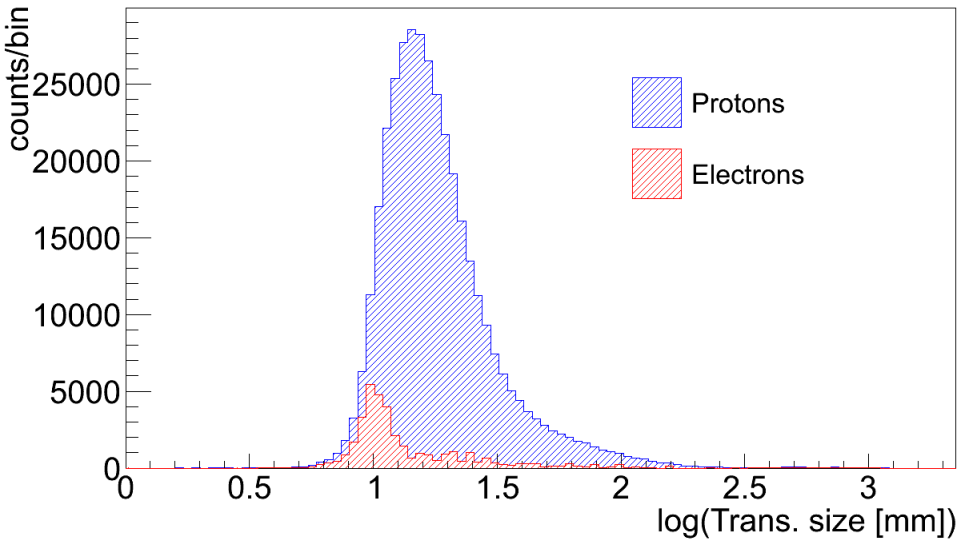


Figure 5.7. Transverse size dimension of the shower in the CAL in the MC *background* simulation. In red are shown electrons, in blue protons.

Another effective discriminant between the two populations is the quality of the fit to the shower profile, measured by its χ^2. Since the fit it is optimized for the EM case, hadronic showers are often poorly modeled Figure 5.9.

In the TKR, an important difference is that an electron can transfer a grater fraction of its momentum compared to a proton in a collision with another electron; thus, on average, it produces much more delta rays while propagating through the layers. A way to measure this difference is by counting the number of hit in a small cylinder around the principal track, which, on average, is higher for electron Figure 5.10 Also, the high number of secondary particles tends to increase the average ToT of the layers along the track (Figure 5.11; remember, from section 2.1 that the ToT is the logical OR of all the channel in half a layer).

We made use of what described above to built a CT with a preliminary

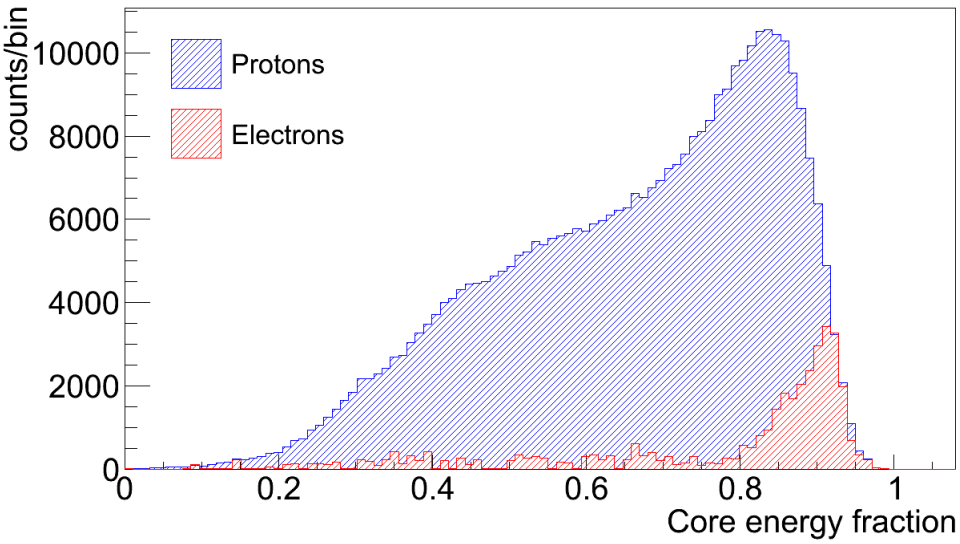


Figure 5.8. Fraction of energy released in a cone of 15 degrees ahead the shower principal axis. In red are shown electrons, in blue protons.

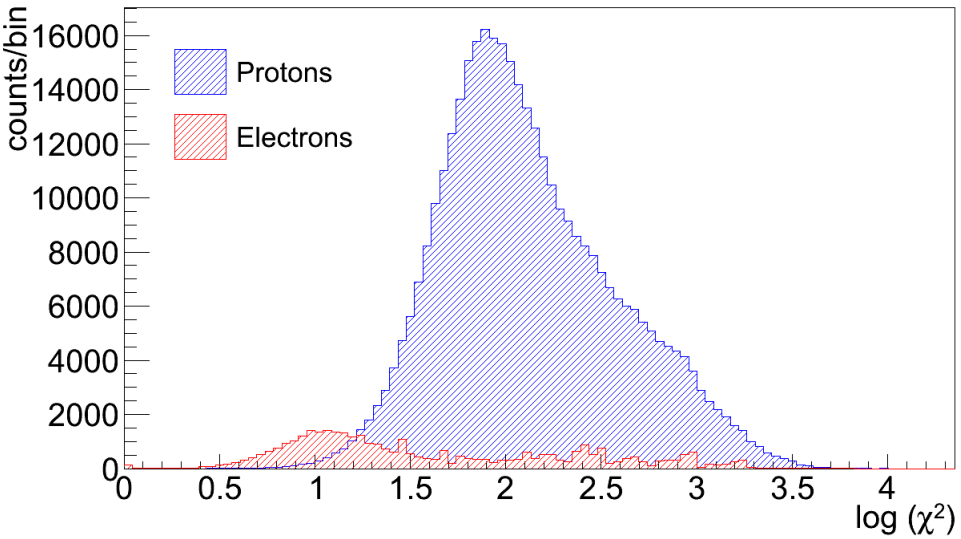


Figure 5.9. χ^2 of the fit to the shower profile. In red are shown electrons, in blue protons.

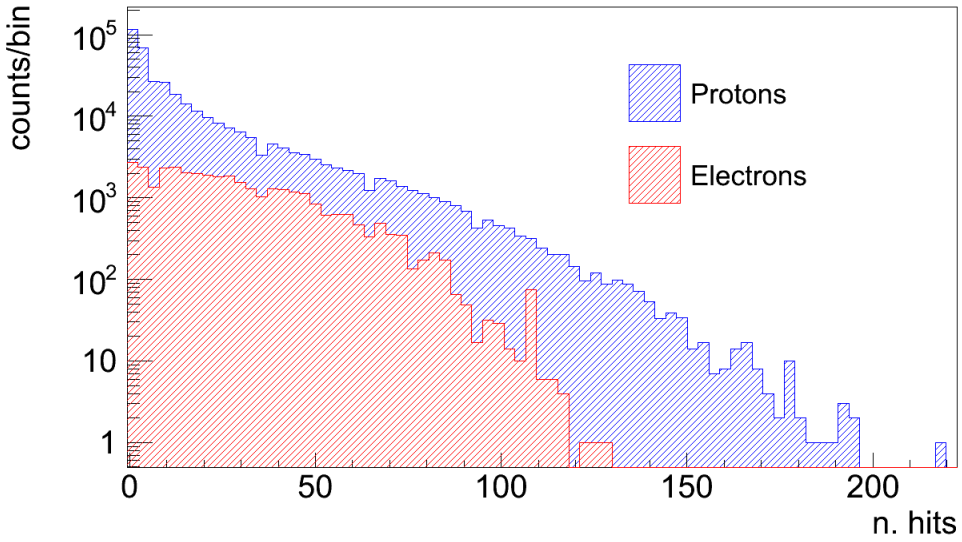


Figure 5.10. Number of clusters within a cylindrical region of radius 10 mm around the best track, excluding the clusters that belong to the track itself. In red are shown electrons, in blue protons.

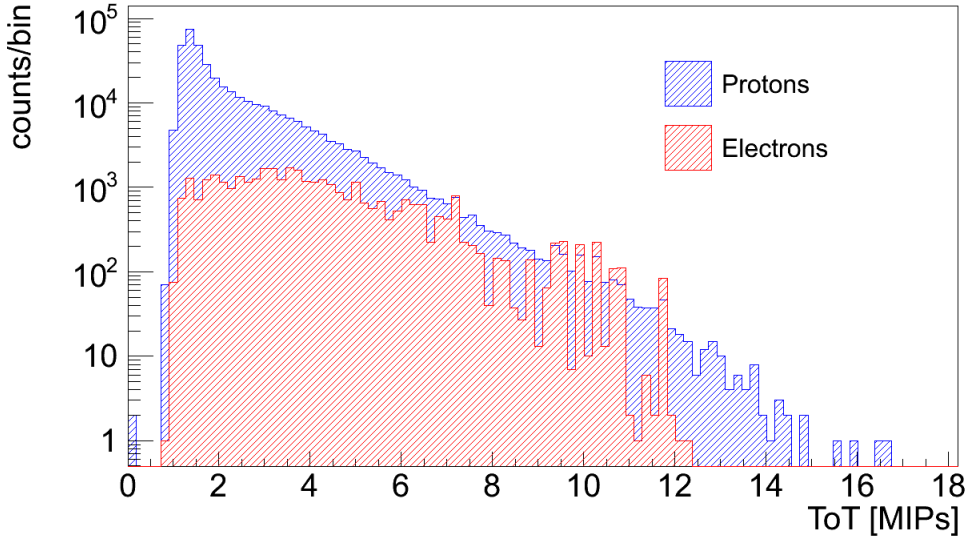


Figure 5.11. Average of the ToT in all the TKR planes. In red are shown electrons, in blue protons.

list of input variables and training options. The same boosting technique described in Chapter 4 has been employed. For each input variable selected, we checked the agreement between its distribution in the data and in the MC simulation. A few candidate variables were identified as being badly reproduced in the simulation and excluded from the analysis.

As a preliminary choice, we extracted our training sample from the *background* simulation, applying the pre-cut described in 5.2. This sample has the advantage that the various charged species are simulated with spectra similar to those present in the data (see section 2.6), but has the drawback of being severely statistic-limited at high energy. Because of that, we could note reliably use it above 500 GeV at the moment. A detailed study of how the specific CR spectra implemented in the training sample can influence the performance of the CT is currently ongoing, as well as tests of alternative simulated data samples.

The CT output is given, for each classified event, in terms of probability that the event is indeed an electron (P_{EL}). We used this variable to perform a energy-dependent cut on the data, given by the expression:

$$\log\left(1 - P_{EL}\right) < 0.4 \cdot \left(5 - \log(E)\right) - 1.8 \tag{5.1}$$

where the energy dependence has been chosen in such a way that the acceptance for this selection, estimated using the *allElectron* simulation, is roughly constant with energy (Figure 5.12). Please note that at the moment this is purely a convenience choice.

After the selection, we use the *background* simulation to estimate the residual contamination of our sample (Figure 5.13).

Since the signal to background ratio decreases with energy (because the electron spectrum is harder then the proton's one), with our choice of a "flat" selection, the background fraction clearly increases at high energy. We subtract bin per bin this background estimated rate from the event rate to get the effective signal rate, then we divide in each bin for the acceptance to get the intensity:

$$I(E) = \frac{R_{evt}(E) - R_{bkg}(E)}{\mathcal{A}(E)} \tag{5.2}$$

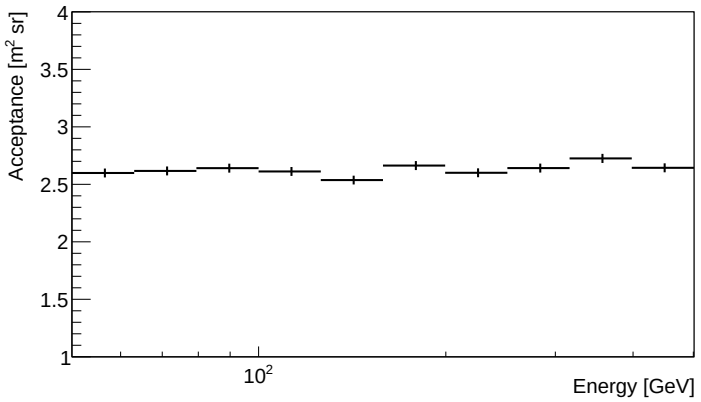


Figure 5.12. Acceptance of the LAT for CREs after the selection applied to remove the contribution of other charged CR species, estimated using the *allElectron* MC simulation.

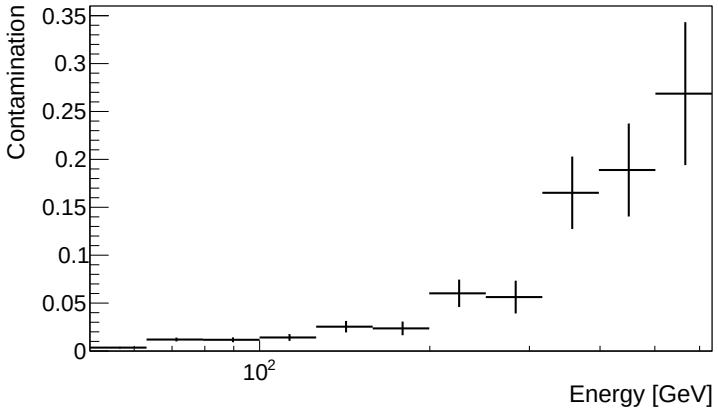


Figure 5.13. Residual background contamination after the selection (i.e. fraction of background events on the total count), estimated using the *background* MC simulation.

where $R_{evt}(E)$ is the count rte in the bin centered at energy E, $R_{bkg}(E)$ the estimated background rate and $\mathcal{A}(E)$ the Acceptance.

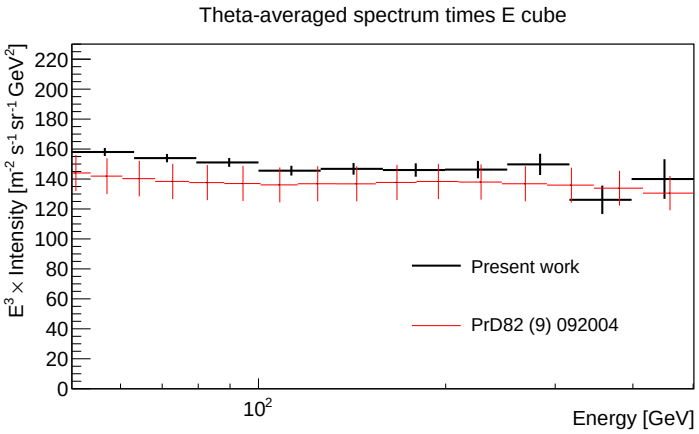


Figure 5.14. Preliminary CRE spectrum from 50 GeV to 500 GeV, multiplied for E^3 for clarity of visualization. A previous measurement published by the *Fermi* collaboration is shown for comparison. Note that the error bars on the black point do not include an estimate of the systematic error.

The Intensity as a function of energy is shown in Figure 5.14. Please note that this spectrum is still preliminary and, among other things, does not include any attempt to estimate the systematic errors. As said we limit ourselves to 500 GeV, because the low statistics currently available in the MC is not enough to continue the analysis at higher energies. Also, we exclude for now those region below 50 GeV, where the effects of the GAMMA onboard filter require a dedicated analysis to be addressed.

Chapter 6

Conclusions

The main body of the work described in this thesis fits into the context of the new event-level analysis being developed for the Fermi Large Area Telescope—and going under the name of Pass 8.

I implemented a multivariate analysis, based on classification trees, aimed at estimating the quality of the energy measurement across the LAT phase space. The output probability from this analysis proved to be effective in removing events with poor energy reconstruction, and has become part of the definition of the standard gamma-ray event classes used by the *Fermi*-LAT collaboration.

By means of Monte Carlo simulations I showed that the information provided by the energy-quality analysis can be readily used to improve the LAT sensitivity in searches for monochromatic gamma-ray lines, such as those predicted by models where DM particles annihilate or decay in a two-body final state with at least one photon. The work described in this thesis will be used by the *Fermi*-LAT collaboration in an update of the previous line-search analyses, to be published at the beginning of next year, based on 6 years of Pass 8 data.

Finally, I actively contributed to the development of a preliminary event selection for the measurement of the cosmic-ray inclusive electron spectrum, the main challenge being the separation between electron/positrons and hadrons, most notably protons. This selection, based again on a classification tree analysis, is characterized in the thesis in terms of acceptance

and residual contamination. While more dedicated, CPU-consuming simulations are needed to cover the entire dynamic range of the LAT, a preliminary CRE spectrum between 50 GeV and and 500 GeV is presented.

Bibliography

[Abd09a] Abdo, A. A. et al., *"The on-orbit calibration of the Fermi Large Area Telescope"* Astroparticle Physics 32, pag. 193-219 (2009)

[Abd09b] Abdo, A. A. et al., *"Measurement of the Cosmic Ray e^+ e^- Spectrum from 20 GeV to 1 TeV with the Fermi Large Area Telescope"*, Phys. Rev. Lett. 102 (18), 181101 (2009).

[Abd09c] Abdo, A. A. et al., *"A limit on the variation of the speed of light arising from quantum gravity effects"*, Nature, Vol. 462, pp. 331-334 (2009)

[Abd10] Abdo, A. A. et al., *"Fermi Large Area Telescope Search for Photon Lines from 30 to 200 GeV and Dark Matter Implications"* Phys. Rev. Lett. 104, 091302 (2010).

[Abd10b] Abdo, A. A. et al., *"Spectrum of the Isotropic Diffuse Gamma-Ray Emission Derived from First-Year Fermi Large Area Telescope Data"*, Phys. Rev. Lett. 104, 101101 (2010).

[Abd13] Abdo, A. A. et al., *"The second Fermi Large Area Telescope catalog of Gamma-ray Pulsars"*, The Astrophysical Journal Supplement Series, Vol. 208, pag. 17 (2013).

[Ack10] Ackermann, M. et al., *"Fermi LAT observations of cosmic-ray electrons from 7 GeV to 1 TeV"*, Phys. Rev. D 82 (9), 092004 (2010).

[Ack11] Ackermann, M. et al., *"Constraining Dark Matter Models from a Combined Analysis of Milky Way Satellites with the Fermi Large Area Telescope"*, Phys. Rev. Lett. 107, 241302 (2011).

[Ack11b] Ackermann, M. et al., *"The Second Catalog of Active Galactic Nuclei Detected by the Fermi Large Area Telescope"*, Astrophysical Journal, Vol. 743, pag. 171 (2011).

[Ack12a] Ackermann, M. et al., "*The Fermi Large Area Telescope on Orbit: Event Classification, Instrument Response Functions, and Calibration*", The Astrophysical Journal Supplement Series, Volume 203, pag. 4 (2012), arXiv1206.1896

[Ack12b] Ackermann, M. et al., *"Fermi LAT search for dark matter in gamma-ray lines and the inclusive photon spectrum"* Phys. Rev. D 86, 022002 (2012).

[Ack12c] Ackermann, M. et al., *"Fermi-LAT Observations of the Diffuse γ-Ray Emission: Implications for Cosmic Rays and the Interstellar Medium"*, Astrophysical Journal, Vol. 750, pag. 3 (2012).

[Ack12d] Ackermann, M. et al., *"Measurement of separate cosmic-ray electron and positron spectra with the Fermi Large Area Telescope"*, Phys. Rev. Lett. 108, 011103 (2012)

[Ack13] Ackermann, M. et al., *"Search for gamma-ray spectral lines with the Fermi Large Area Telescope and dark matter implications"*, Phys. Rev. D 88, 082002 (2013)

[Ack13b] Ackermann, M. et al., *"The First Fermi-LAT Gamma-Ray Burst Catalog"*, The Astrophysical Journal Supplement Series, Vol. 209, pp. 11-101 (2013).

[Ack13c] Ackermann, M. et al., *"The Fermi All-sky Variability Analysis: a list of flaring Gamma-ray Sources and the search for Transients in our Galaxy"*, Astrophysical Journal, Vol. 771, pag. 57 (2013).

[Ack13d] Ackermann, M. et al., *"The First Fermi-LAT Catalog of Sources Above* 10 *GeV"*, arXiv:1306.6772v3 [astro-ph.IM] (2013).

[Ack13e] Ackermann, M. et al., *"Fermi-LAT Observations of the Gamma-ray Burst GRB 130427A"*, Science, Vol. 343, no. 6166, pp. 42-47 (2013).

[Ade13] Ade, P. et al. (Planck Collaboration), *"Planck 2013 results"*, arXiv:1303.5076 (2013).

[Ago03] Agostinelli, S. et al., *"Geant4 - A Simulation Toolkit"*, Nucl. Instrum. Methods Phys. Res. A, 506, 250 (2003)

[All06] Allison, J. et al., *"Geant4 Developments and Applications"*, IEEE Transactions on Nuclear Science 53 No. 1 270-278 (2006).

[Atw07] Atwood, W. B. et al., *"Design and initial tests of the Tracker-converter of the Gamma-ray Large Area Space Telescope"* Astroparticle Physics 28 422-434 (2007)

[Atw09] Atwood, W. B. et al., "*The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission*", The Astrophysical Journal, Volume 697, pag. 1071-1102 (2009), arXiv0902.1089

[Atw13] Atwood, W. B. et al., *"New Fermi-LAT event reconstruction reveals more High-Energy Gamma Rays from Gamma-Ray Bursts"*, The Astrophysical Journal, Vol. 774, pag.76 (2013).

[Bal06] Baldini, L. et al., "*GLAST LAT Full Simulation*", Nuclear Physics B Proceedings Supplements, Volume 150, pag. 62-65 (2006)

[Bis09] Bissaldi, E. et al. *"Ground-based calibration and characterization of the Fermi gamma-ray burst monitor detectors"*, Experimental Astronomy, 24, 47 (2009) and [astro-ph/08122908].

[Bre13] Bregeon, J.; Charles, E.; Wood, M. for the Fermi-LAT collaboration *"Fermi-LAT data reprocessed with updated calibration constants"*, arXiv:1304.5456 [astro-ph.HE] (2013)

[Bre84] Breiman, L.; Friedman, J. H.; Olshen, R. A. and Stone, C. J. *Classification and Regression Trees*, Chapman & Hall (1984)

[Bri12] Bringmann, T. et al., *"Fermi LAT search for internal bremsstrahlung signatures from dark matter annihilation"*, J. Cosmol. Astropart. Phys. 07 (2012) 054.

[Bru12] Bruel, P.,*Gamma rays, electrons and positrons up to* 3 *TeV with the Fermi Gamma-ray Space Telescope*, Journal of Physics: Conference Series 404(1), 012033 (2012)

[Fri01] Friedman, J. H., *Greedy Function Approximation: A Gradient Boosting Machine*, The Annals of Statistics, Vol. 29, No. 5 (Oct., 2001), pp. 1189-1232

[Fru87] Fruhwirth, R., *"Application of Kalman filtering to track and vertex fitting"*, Nucl. Instrum. Meth. A262, pp. 444-450 (1987).

[Gro10] Grove, J. E.; Johnson, W. N., *"The calorimeter of the Fermi Large Area Telescope"*, Proc. SPIE, 7732 (2010)

[Gru14] Gruber, D. et al., *"The Fermi GBM Gamma-Ray Burst Spectral Catalog: Four Years Of Data"*, arXiv:1401.5069v2 [astro-ph.HE] (2014).

[Has09] Hastie, T.; Tibshirani, R.; Friedman, J. H. *(*"10.Boosting and Additive Trees"). The Elements of Statistical Learning (2nd ed.), Springer. pp. 337-384 (2009).

[Hoe09] Hoecker, A. et al., *"TMVA - Toolkit for Multivariate Data Analysis"*, arXiv:physics/0703039v5

[Jun96] Jungman G.; Kamionkowski, M.; Griest, K., *"Supersymmetric Dark Matter"* Phys. Rep. 267, 195 (1996)

[Kal60] Kalaman,R. E., *"A New Approach to Linear Filtering and Prediction Problems"*, Journal of Basic Engineering 82 (1), pp. 35–45 (1960).

[Kie14] von Kienlin, A. et al., *"The 2nd Fermi GBM Gamma-Ray Burst Catalog: The First Four Years"*, arXiv:1401.5080v2 (2014).

[Mar11] Martin, S. P., *"A Supersymmetry Primer"*, arXiv:hep-ph/9709356.

[Mee09] Meegan, C. A. et al., *"The Fermi Gamma-ray Burst Monitor"*, Astrophysical Journal, Vol. 702, pag. 791, (2009) and [astro-ph/09080450].

[Miz04] Mizuno et al., *"Cosmic-Ray Background Flux Model Based on a Gamma-Ray Large Area Space Telescope Balloon Flight Engineering Model"*, The Astrophysical Journal, Volume 614, pag. 1113-1123, (2004)

[Moi07] Moiseev, A. A. et al., *"The Anti-Coincidence Detector for the GLAST Large Area Telescope"*, Astroparticle Physics, 27, 339 (2007)

[Nol12] Nolan, P. L. et al., *"Fermi Large Area Telescope Second Source Catalog"* The Astrophysical Journal Supplement Series, Vol. 199, pag. 31 (46pp), 2012 April.

[Par97] van Paradijs, J. et al., *"Transient optical emission from the error box of the γ-ray burst of 28 February 1997"*, Nature, Vol. 386, nº 6626, pag. 686-689 (1997).

[Sha01] Schapire, R. E., *"The Boosting Approach to Machine Learning, An Overview"*, Nonlinear Estimation and Classification, Springer (2003)

[Str01] Strong, A. W.; Moskalenko, I. V., *"Models for glactic Cosmic-ray propagation"* Adv. Space Rex Vol. 27, No. 4, pp. 717-726 (200l)

[Tho12] Thompson, D.J.; Baldini, L.; Uchiyama, Y., *"Cosmic ray studies with the Fermi Gamma-ray Space Telescope Large Area Telescope"*, Astroparticle Physics 39–40, 22–32 (2012)

Acknowledgements

A conclusione di questo lavoro, desidero ringraziare innanzitutto Luca, per essere stato tutto quello che si potrebbe chiedere di essere ad un relatore: gentile nei modi, prodigo nelle spiegazioni, tempestivo e puntuale nelle correzioni, tollerante per i miei (numerosi) errori, e per avere preso sinceramente a cuore la riuscita della tesi.

Un grazie va anche agli altri ragazzi del gruppo Fermi qui a Pisa, per avermi aiutato e fatto sentire a casa sin dall'inizio: Melissa, Massimiliano, Francesca e soprattutto Carmelo, che mi ha dato una mano in un numero incommensurabile di occasioni.

Per lo studio di sensibilità alla linea devo un ringraziamento particolare ad Andrea, per tutte le spiegazioni e l'aiuto datemi sia via mail che dal vivo. La sua gentilezza e disponibilità sono state impagabili.

Ancora, non potrei non menzionare i ragazzi di Torino, Luca, Raffaella e Michela, sia per aver lavorato con noi dall'inizio alla fine all'analisi sugli elettroni e ancor di più per la simpatia e cordialità con cui lo hanno fatto.

Sarebbe troppo lungo menzionare tutte le altre persone all'interno della collaborazione che hanno contribuito, direttamente o indirettamente, a questo lavoro, con consigli, indicazioni, suggerimenti, critiche . . . per cui mi limito a ringraziarle tutte, indistintamente.

Infine un grazie a tutti quelli che hanno contribuito in qualunque forma ad aiutarmi a portare a termine questo lavoro, materialmente e non, anche solo con quattro chiacchiere davanti a un caffè. C'è un po' di tutti voi in questa tesi (e spero non vi dispiaccia!).